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

Abstract

Much political science research involves analysis of a dependent variable that has boundary restrictions. In econometric textbooks, these studies should apply the truncated regression model, otherwise the OLS estimate is likely to generate out-of-bounds predicted values. However, political scientists seldom use truncated regression and are unaware of this methodological problem. In this article, the author investigates this issue and finds that both the OLS and truncated regression models suffer from boundary violations. To resolve this problem, the author proposes a revised truncated regression model with constrained optimization and successfully eliminates boundary violations. Simulation results found via various settings confirm the superiority of the revised model. Further analysis indicates that hypothesis testing results are quite sensitive if different models are applied. This finding significantly challenges the appropriateness of the current model. Through a replication study, the author demonstrates how the revised model can be applied to political studies and why it is preferable to OLS and the current truncated regression model.

Keywords: Truncated Normal Distribution; Truncated Regression Model; Constraint Optimization; Sequential Quadratic Programming; Boundary Violations

1. INTRODUCTION

Much political science research involves analysis of a dependent variable that has boundary restrictions. In electoral studies, vote share and voter turnout (aggregate-level) have a percentage measure and are naturally bounded between 0 and 1. (Geys, 2006; Holbrook and McClurg, 2005) In public opinion studies, the presidential approval rating is another percentage variable of the same nature. (Kernell, 1978; West, 1991) In public policy, learning achievement measured with standardized test scores like the SAT or achievement measures like grade point average are constrained by a bounded point scale. (Berger and Toma, 1994; Henry and Rubenstein, 2002) In world politics, the polity score that measures the level of democracy ranges from -10 to 10. (Knack, 2004; Rudra, 2005) Examples like these abound, and these dependent variables are all distributed within a certain boundary. In econometric textbooks, the truncated regression model (hereafter the TRM model) would be a more appropriate regression method to apply.[1] (Greene, 2008, 863-869)

The TRM model has been developed over more than three decades.[2] It can be easily

[1] The truncated regression model is usually applied when the dependent variable has boundary restrictions. (Amemiya, 1973, 1984) The boundary restrictions can be singly bounded at a lower or upper limit, or doubly bounded within an interval. (Johnson et al., 1970) When such truncation reflects the essential feature of the distributional assumption, the truncated regression model differs from the censored regression, such as the Tobit model (Tobin, 1958), in two aspects. First, truncated regression does not allow any observation outside the boundary, including dependent and independent variables. Censored regression, on the other hand, does have observations outside the boundary, but the values of the dependent variable are all collapsed into the boundary values. (Greene, 2008, 869) Second, given the different nature of truncation, the probability density function is also different for the two models. For truncated regression, the pdf function is simply the untruncated normal density divided by a probability measure from the lower to upper limits. For censored regression, the pdf function is a mixture of discrete and continuous distributions in which the former captures the censoring mechanism, and the latter remains as the same as the uncensored case. (Breen, 1996, 4)

[2] We can further distinguish the TRM model from the censored regression and the Heckmen model in terms of a data generating process. For truncated regression, it only needs a distributional assumption, but censored regression contains a distributional assumption and a censoring mechanism. The same distinction can be made about the sample-selected model, such as the Heckman model (Heckman, 1979), in which the data is only available when the criterion of another variable is satisfied. (Sigelman and Zeng, 1999, 177). While observed values of the dependent variable in the three models are all distributed as truncated normal, the censored and sample-selected models have different working assumptions for the dependent variable. The censored model assumes an underlying untrucated normal distribution, plus the censoring mechanism that confines the dependent variable within a certain range. Similarly, the Heckman model assumes a bivariate normal distribution of the error terms for the selection and outcome regressions with a correlation coefficient. (Sartori, 2003, 114) The empirical truncation of the outcome dependent variable depends on the selection mechanism that is specified in the Heckman model. Apparently, the censored and sample-selected models do

executed with statistical software, such as Stata (with the command *truncreg*). Previous literature shows that the TRM model has been applied in many disciplines, such as economics, astronomy, and biology (Jewell and Wu, 1988), but its application in the field of political science has been limited. Why do political scientists seldom use this method? What is the cost of not applying this method when the dependent variable is distributed as truncated normal?

As with the use of the logit or probit model for a binary dependent variable, the fundamental reason to use the TRM model is to avoid boundary violations. Given that the distributional assumption has already determined an admissible region of the dependent variable, any solution that generates an out-of-bounds predicted value is ineligible and regarded as a failed estimate. While this problem is common for the binary dependent variable (Aldrich and Nelson, 1984), little attention is paid to the TRM model. If the widespread belief in political science is that the logit or probit model should be applied to a binary dependent variable, the same conclusion should be made about the TRM model for a truncated normal dependent variable.

While we may never know why political scientists seldom use the TRM model,[3] a possible reason stems from the fact that, unlike the logit or probit model, TRM cannot solve the problem of boundary violations. (Orme and Ruud, 2002, 213) Empirical applications show that "The truncated normal model routinely defied convergence and, as often as not, produced nonsense estimates." (Greene, 1999, 158, n60) To validate this concern, we investigated three political science studies and found that they all suffer from boundary violations by either the OLS or TRM model.

not directly assume the dependent variable as univariate truncated normal, and they both add an additional assumption to the data-generating process that might not be true.

[3]Political scientists, on the other hand, do pay more attention to the censored regression and the sample-selected model. Unlike these two models, the analytical purpose of the TRM model is not to recover the information of the underlying untruncated normal distribution, nor to correct the selection bias. Rather, the main task is to derive the best parameter estimates from the eligible parameter space. In econometrics, many efforts have been made in the theoretical study of the maximum likelihood estimator for the TRM model. See Olsen (1978), Orme (1989), Hausman and Wise (1977), Chung and Goldberger (1984), and Greene (1983).

The first case is Timothy Hellwig and David Samuels's 2007 article, "Voting in Open Economies: The Electoral Consequences of Globalization" in *Comparative Political Studies*. We replicate Model I and II in Table 1 (p.292), which explains the incumbent party's vote share and find that the least predicted vote shares are −4.797% and −3.163% by OLS, and −7.938% and −5.874% by TRM. The second case is Thomas Hansford and Brad Gomez's 2010 article, "Estimating the Electoral Effects of Voter Turnout" in *American Political Science Review*. We replicate the model in the column 2 of Table 1 (p.277) that conducts an F-test for excluded instruments. The dependent variable is aggregate-level voter turnout when the incumbent is Republican. The least predicted value is −8.287% by OLS, and −24.277% by TRM. The third case is Daron Acemoglu et al.'s 2008 article, "Income and Democracy" in *American Economic Review*. We replicate the pooled and fixed-effects OLS that explain the level of democracy in Table 2 (p.816). The dependent variable is a normalized Freedom House measure of democracy, ranging from 0 to 1. The least and greatest predicted values for the pooled model are −0.031 and 1.065 by OLS and −0.103 and 1.202 by TRM. For the fixed-effects model, the least and greatest predicted values are −0.052 and 1.077 by OLS, and −0.147 and 4.488 by TRM.

None of the three articles explains why the TRM model is not used, nor explains how to interpret the out-of-bounds predicted values. This scenario indicates that the current version of the TRM model is not widely perceived as the default method among political scientists. Moreover, if the TRM model were used, the problem of boundary violations would have been more significant considering the replication results. Therefore, even though the application of the OLS model is clearly at odds with the distributional assumption, most political scientists are not aware of this methodological problem. And thus far, there is no solution in the field to simultaneously resolve the violations of the distributional assumption and boundary restrictions.

In this article, the author proposes a revised TRM model (hereafter the TRMCO model) within the framework of nonlinear programming and demonstrates how to improve the current model by eliminating boundary violations. By specifying proper boundary restrictions, the estimation of the TRMCO model will be transformed into a constrained optimization problem (COP). (Bertsekas, 1996) While the ideas involved in solving a COP problem are not dramatically different from those that solve an unconstrained optimization problem, few techniques in numerical analysis are applied to the statistical methods familiar to political scientists.[4] In the following sections, the author first explains the problem of boundary violations in the TRM model. Next, the author presents a modified procedure of maximum likelihood estimation with the sequential quadratic programming (SQP) algorithm (Nocedal and Wright, 1999, 529) that solves constrained optimization problems. Third, the author compares the inferential validity of the modified model with the current one through three simulation tests. Finally, an empirical case is presented to illustrate the remarkable difference when the TRMCO model is applied.

2. BOUNDARY VIOLATIONS IN THE TRUNCATED REGRESSION MODEL

The TRM model can be described with the following specifications. Suppose we have n *i.i.d.* observations of y_i, which follows the truncated normal distribution $TN\left(\mu_i, \sigma^2; a, b\right)$, where μ_i, σ, a, and b are location parameter, scale parameter, lower limit, and upper limit, respectively. We add m covariates $(x_1, \cdots, x_m)$ in the model to explain μ_i for each observation i by assuming $\mu_i = \boldsymbol{x_i}\boldsymbol{\beta}$. Therefore, the likelihood function is

$$L \equiv \prod_{i=1}^{n} \left\{ \frac{\exp\left(\frac{-(y_i - \boldsymbol{x_i}\boldsymbol{\beta})}{2\sigma^2}\right)}{\int_a^b \exp\left(\frac{-(y - \boldsymbol{x_i}\boldsymbol{\beta})}{2\sigma^2}\right) dy} \right\}.$$

[4] Previous applications of constrained optimization in political science tend to focus on the formal theory instead of numerical analysis. See Moe (1980) and Sorokin (1994). Recently, political scientists started working on numerical problems with constrained optimization. See Sekhon and Mebane (1998) and Mebane and Sekhon (2011)

Using $\Phi(\cdot)$ to replace the cdf function of the normal distribution. We can derive the loglikelihood

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

where $D_i = \sqrt{2\pi}\sigma\left[\Phi\left(\frac{b-\boldsymbol{x_i\beta}}{\sigma}\right) - \Phi\left(\frac{a-\boldsymbol{x_i\beta}}{\sigma}\right)\right]$. With a few manipulations, we can deduce the gradient vector and the Hessian matrix, and apply the generalized Gauss-Newton algorithm to derive maximum likelihood estimates of $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}$. (Hausman and Wise, 1977, 936)

Notice that the above model does not specify any constraints on the dependent variable y_i, regression coefficient $\boldsymbol{\beta}$, and scale parameter σ. However, those parameters do have certain theoretical constraints that need to be specified. Those constraints can be categorized into three types: (i) boundary limits of the dependent variable, (ii) admissible parameter space of the independent variables, and (iii) constraints of the scale parameter. Type I constraints state that the dependent variable should be bounded within the lower limit a and upper limit b. Considering the joint set of the maximum and minimum values for the independent variable x_j (j is the variable indicator), the boundary constraints for $\hat{y}$ should be

$$a \leq \hat{y}^{\min} \leq \hat{y}^{\max} \leq b,$$

where

$$\hat{y}^{\max} = \sum_{j=0}^{m}\left(v_j^+\beta_j x_j^{\max} + v_j^-\beta_j x_j^{\min}\right)$$
$$\hat{y}^{\min} = \sum_{j=0}^{m}\left(v_j^+\beta_j x_j^{\min} + v_j^-\beta_j x_j^{\max}\right),$$

and v_j^+ and v_j^- are indicator variables indicating

$$v_j^+ = \begin{cases} 1 & if\ \beta_j > 0 \\ 0 & otherwise \end{cases}, v_j^- = \begin{cases} 1 & if\ \beta_j < 0 \\ 0 & otherwise \end{cases}.$$

It is not uncommon for the truncated regression model to give a solution that violates boundary constraints of the dependent variable. In such a case, there is no way to give a meaningful interpretation, since the data-generating process forbids its occurrence. For example, it makes no sense to predict a party will get 105% or −5% of votes given that the vote share is bounded within 100% and 0%. Here we use the joint set of maximum or minimum covariate values because sometimes the predicted values of all empirical observations are admissible, but certain combinations of the covariate values can result in boundary violations. For example, all empirical pairs of x_{ji} do not generate out-of-bounds prediction $\hat{y}_i$, but some combinations of $x_j^{\max}$ and $x_j^{\min}$ might give an inadmissible prediction larger than b or smaller than a.[5] Unless we have a reason to rule out the possibility of their joint presence, we should evaluate type I boundary violations by including all possible combinations of the covariate values.[6]

Type II constraints are related to the regression coefficient β, which is best illustrated by the centered model

$$y_i = \boldsymbol{x_i^*}\boldsymbol{\beta},$$

where $\boldsymbol{x}^*$ refers to the matrix of the independent variables $\boldsymbol{x_j}$, $j = 1, \ldots, m$, after being centered at the means $\bar{x}_j$,

$$\boldsymbol{x}^* = \begin{pmatrix} \vdots & \vdots & & \vdots \\ 1 & (x_{1i} - \bar{x}_1) & \cdots & (x_{mi} - \bar{x}_m) \\ \vdots & \vdots & & \vdots \end{pmatrix}.$$

The constant $\hat{\beta}_0$ now represents the mean estimate of y_i when none of the covariates has explanatory power, or represents the baseline predicted value of y_i when all covariates

[5]Imagine our model predicts that *IQ score* and *study hour* are both positively related to *SAT score*, but the data does not have a case in which both variables have the maximum value. Such a case is very likely to exist, and our model should not generate an out-of-bounds predicted value on *SAT score*.

[6]When a set of covariates is composed of regional dummy variables, the joint presence is impossible, and therefore, only the maximum and minimum coefficients of those dummies are specified in the boundary constraints of y_i.

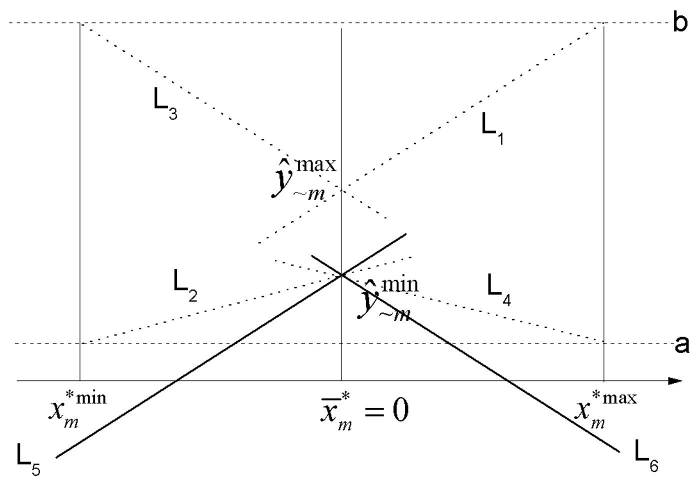

Figure 1: Boundary Constraints of β_m

are held by the means. Apparently, the boundary constraints of $\hat{\beta}_0$ should be within the truncation interval

$$a \leq \hat{\beta}_0 \leq b.$$

To discover the possible range of other $\hat{\beta}_m$, we first hold x_m^* at the mean level $\bar{x}_m^*$ as Figure 1 shows, and then derive the maximum and minimum of the predicted values, $\hat{y}_{\sim m}^{\max}$ and $\hat{y}_{\sim m}^{\min}$, respectively. The notation "$\sim m$" represents the fact that x_m^* has no contribution when it is held at the mean. More precisely, $\hat{y}_{\sim m}^{\max}$ and $\hat{y}_{\sim m}^{\min}$ can be specified

$$\hat{y}_{\sim m}^{\max} = \sum_{j=0, j\neq m} \left(v_j^+ \hat{\beta}_j x_j^{*\max} + v_j^- \hat{\beta}_j x_j^{*\min} \right)$$
$$\hat{y}_{\sim m}^{\min} = \sum_{j=0, j\neq m} \left(v_j^+ \hat{\beta}_j x_j^{*\min} + v_j^- \hat{\beta}_j x_j^{*\max} \right).$$

The upper limit of β_m is the flatter positive slope of the line L_1 or L_2. The lower limit is the flatter negative slope of the line L_3 or L_4. Therefore, the boundary constraints of $\hat{\beta}_m$ can be identified as

$$\max\left(\frac{a-\hat{y}_{\sim m}^{\min}}{x_m^{*\max}}, -\frac{\hat{y}_{\sim m}^{\max}-b}{x_m^{*\min}}\right) \leq \hat{\beta}_m \leq \min\left(\frac{b-\hat{y}_{\sim m}^{\max}}{x_m^{*\max}}, -\frac{\hat{y}_{\sim m}^{\min}-a}{x_m^{*\min}}\right).$$

If $\hat{\beta}_m$ takes the steeper slope, such as L_5 or L_6 shows, it would generate an out-of-bounds predicted value when we vary x_m^* from the mean to the maximum or minimum, holding other variables at the baseline level. In this sense, a type II violation can always be translated into a type I violation.

Different centering methods do not generate different estimates of the beta coefficients, except the constant, which is a linear combination of all other beta coefficients and the centered covariate values.[7] For a truncated regression model with m covariates, there will always be $(2m+2)$ type II boundary constraints, including the constant.

Type III boundary constraints are about the scale parameter σ. While often the constraints are not effective, we can consider adopting the full truncation range as the upper limit and an arbitrary small positive number (κ) as the lower limit[8]

$$\kappa \leq \hat{\sigma} \leq b-a.$$

If σ approaches infinity, y_i will approach the uniform distribution. When the optimization result gives an upper boundary value of $\hat{\sigma}$, it signifies a violation of the distribution assumption and means that y_i does not fit the truncated normal assumption well. For the lower limit constraint, if σ approaches zero or becomes negative, this indicates a negative variance resulting from the non-positive definite Hessian. Many possible explanations can account for this problem, but its occurrence is usually associated with an ill-specified model, and thus regarded as a failed estimate.

In this article, we separate the OLS out-of-bounds violation from the type I violation.

[7]This rule only applies to a strict linear model. If truncated regression is specified with a nonlinear relationship, such as interaction, different centering methods will generate different results.

[8]We set $\kappa = 0.001$ in this paper.

The former happens when the OLS estimate generates an inadmissible predicted value to an empirical observation; the latter is identified when any possible predicted value falls outside the boundary. Apparently, a type I violation is defined with a more rigid standard, and it encompasses the OLS out-of-bounds violation.

3. PARAMETER ESTIMATION WITH CONSTRAINED OPTIMIZATION

The maximum likelihood estimate of the truncated regression model is a nonlinear constrained minimization optimization:[9]

$$
\begin{aligned}
&\textit{Minimize} && -\log L\left(\boldsymbol{\beta}, \sigma | \boldsymbol{x}^{*}, y_i\right) \\
&\textit{Subject to} && g_1 = \beta_0 + \hat{y}_{\sim 0}^{\max} - b \leq 0 \\
& && g_2 = a - \beta_0 - \hat{y}_{\sim 0}^{\min} \leq 0 \\
& && g_3 = \beta_0 - b \leq 0 \\
& && g_4 = -\beta_0 + a \leq 0 \\
& && \vdots \\
& && g_{2m+3} = \beta_m - \min\left(\frac{b - \hat{y}_{\sim m}^{\max}}{x_m^{*\max}}, -\frac{\hat{y}_{\sim m}^{\min} - a}{x_m^{*\min}}\right) \leq 0 \\
& && g_{2m+4} = -\beta_m + \max\left(\frac{a - \hat{y}_{\sim m}^{\min}}{x_m^{*\max}}, -\frac{\hat{y}_{\sim m}^{\max} - b}{x_m^{*\min}}\right) \leq 0 \\
& && g_{2m+5} = \sigma - b + a \leq 0 \\
& && g_{2m+6} = -\sigma + \kappa \leq 0.
\end{aligned}
\tag{1}
$$

[9]In this section, we adopt the centered model for the truncated regression. The covariate matrix is noted with an asterisk as $\boldsymbol{x}^{*}$.

We can specify this problem in matrix terms:

$$\boldsymbol{\gamma} = \begin{pmatrix} \boldsymbol{\beta} \\ \sigma \end{pmatrix}, c_I(\boldsymbol{\gamma}) = \begin{pmatrix} g_1 \\ \vdots \\ g_{2m+6} \end{pmatrix}, (P_I) \begin{cases} \min f(\boldsymbol{\gamma}) \\ c_I(\boldsymbol{\gamma}) \leq 0 \\ \boldsymbol{\gamma} \in \Omega \end{cases},$$

where $\boldsymbol{\gamma}$ refers to the parameter vector being estimated, $c_I(\boldsymbol{\gamma})$ is the vector of inequality constraints, Ω is the feasible parameter space, and P_I represents the general minimization problem with only inequality constraints. The subscript I represents inequality constraints hereafter.

To solve the P_I, we first add Lagrange multipliers to incorporate the inequality constraints and modify the objective function into the Lagrangian

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

If optimality is achieved, a solution should exist for $\boldsymbol{\gamma}_*$ and $\boldsymbol{\lambda}_*$ and satisfy the KKT (Karush-Kuhn-Tucker) conditions. (Kuhn and Tucker, 1951)

$$(KKT) \begin{cases} (a)\, \nabla f(\boldsymbol{\gamma}_*) + A(\boldsymbol{\gamma}_*)^T \boldsymbol{\lambda}_* = 0 & \text{(Stationarity)} \\ (b)\, c_I(\boldsymbol{\gamma}_*) \leq 0 & \text{(Primal feasibility)} \\ (c)\, (\boldsymbol{\lambda}_*)_I \geq 0 & \text{(Dual feasibility)} \\ (d)\, (\boldsymbol{\lambda}_*)_I{}^T c_I(\boldsymbol{\gamma}_*) = 0 & \text{(Complementary slackness)}, \end{cases} \tag{2}$$

where $\nabla f(\boldsymbol{\gamma}_*) = \partial l(\boldsymbol{\gamma}; \boldsymbol{\lambda}) / \partial \boldsymbol{\gamma}$, $A(\boldsymbol{\gamma}) = \partial c_I(\boldsymbol{\gamma}) / \partial \boldsymbol{\gamma}$.

By linearizing (2)[10] and replacing $*$ with k, we can derive a modified system of KKT

[10]For any function $F(x) = 0$, the Newton method generates a sequence of $\{x_k\}$, where $x_{k+1} = x_k + d_k$, to find x_* so that $F(x_*) = 0$. In each iteration, d_k can be derived through the linearization of $F(x)$, in which $F(x_k) + F'(x_k)\, d_k = 0$.

conditions. (Bonnans et al., 2006, 257)

$$\begin{cases} \boldsymbol{L_k d} + \boldsymbol{A_k}^T \boldsymbol{\lambda^{QP}} = -\nabla \boldsymbol{f_k} \\ \left(\boldsymbol{c_k} + \boldsymbol{A_k d}\right)_I = 0 \\ \left(\boldsymbol{\lambda^{QP}}\right)_I \geq 0 \\ \left(\boldsymbol{\lambda^{QP}}\right)_I^T \left(\boldsymbol{c_k} + \boldsymbol{A_k d}\right)_I = 0, \end{cases} \tag{3}$$

where the KKT conditions are satisfied at the kth iteration, $\boldsymbol{\lambda^{QP}} := \boldsymbol{\lambda_k} + \boldsymbol{\mu}$, and $\boldsymbol{L_k} = \left\{\partial^2 l\left(\gamma; \boldsymbol{\lambda}\right)/\partial\gamma\partial\boldsymbol{\gamma}^T\right\}_k$.

The step parameter $\boldsymbol{d}$ is the updated estimate for $\boldsymbol{\gamma}$ at the kth iteration in the numerical analysis of P_I. The Lagrange multiplier $\boldsymbol{\lambda^{QP}}$ is the estimate of $\boldsymbol{\lambda} + \boldsymbol{\mu}$ at the kth iteration. The modified system (3) is, in fact, the optimality system of the osculating quadratic problem (QP). (Bonnans et al., 2006, 218)

$$\begin{cases} \min_d \nabla f\left(\boldsymbol{\gamma_k}\right)^T \boldsymbol{d} + \frac{1}{2}\boldsymbol{d}^T \boldsymbol{L_k d} \\ c_I\left(\boldsymbol{\gamma_k}\right) + A_I\left(\boldsymbol{\gamma_k}\right)\boldsymbol{d} \leq 0 \end{cases} \tag{4}$$

The method described above is the sequential quadratic programming (SQP) algorithm by which we break down a nonlinear constrained optimization problem into a series of osculating quadratic problems. The sequence of the solutions $\left(\boldsymbol{\gamma_k}, \boldsymbol{\lambda^{QP}}\right)$ comes from solving $\boldsymbol{d_k}$ in (4) and $\boldsymbol{\lambda^{QP}}$ in (3) at each iteration, and it will approximate the optimal solution $\left(\boldsymbol{\gamma_*}, \boldsymbol{\lambda_*}\right)$ when the KKT conditions are satisfied. Since the osculating quadratic problem is much easier to solve, the idea behind the SQP algorithm is to break down the complicated nonlinear constrained optimization problem into a series of QP problems and gradually reach the optimal solution.

Parameter estimation is carried out in the Matlab environment by using the built-in quadratic programming solver, *quadprog*. The algorithm is described below: (Bonnans et al., 2006, 257)

1. Set up initial value of $(\boldsymbol{\gamma_0}, \boldsymbol{\lambda_0})$, and compute $c_I(\boldsymbol{\gamma_0})$, $\nabla f(\boldsymbol{\gamma_0})$, and $A_I(\boldsymbol{\gamma_0})$. Set the iteration index $k = 0$.

2. Stop and report $\left(\boldsymbol{\gamma_k}, \boldsymbol{\lambda^{QP}}\right)$ as the optimal solution if the KKT conditions (2) are satisfied.

3. Solve the QP problem (4) by computing the Hessian matrix of the Lagrangian $L(\boldsymbol{\gamma_k}, \boldsymbol{\lambda_k})$ and derive d_k with Matlab function *quadprog*.

4. Solve the Lagrange multiplier $\boldsymbol{\lambda^{QP}}$ by the first equation in (3),
$\boldsymbol{\lambda^{QP}} = \left(\boldsymbol{A_k}^T\right)^{-1}(-\nabla \boldsymbol{f_k} - \boldsymbol{L_k d})$.

5. Set the new solution of the $k+1$ iteration as $\boldsymbol{\gamma_{k+1}} = \boldsymbol{\gamma_k} + \boldsymbol{d_k}$ and $\boldsymbol{\lambda_{k+1}} = \boldsymbol{\lambda^{QP}}$.

6. Compute $c_I(\boldsymbol{\gamma_{k+1}})$, $\nabla f(\boldsymbol{\gamma_{k+1}})$, and $A_I(\boldsymbol{\gamma_{k+1}})$. Go back to the step 2 and set $k = k+1$.

We can directly compute the gradient vector $\nabla f(\boldsymbol{\gamma})$ and the Lagrangian Hessian matrix $L(\boldsymbol{\gamma_k}, \boldsymbol{\lambda_k})$. The initial value is set to the parameter estimates of the truncated regression model without boundary constraints by the Stata *truncreg* command.[11] (Cong, 2000)

Hypothesis testing can be carried out by computing the empirical variance-covariance matrix

$$Var(\hat{\boldsymbol{\gamma}}) = \left(-\frac{\partial^2(\ln L)}{\partial\hat{\boldsymbol{\gamma}}\partial\hat{\boldsymbol{\gamma}}^T}\right)^{-1},$$

where $\hat{\boldsymbol{\gamma}} = \left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}\right)^T$. When the *i.i.d.* assumption is violated, such as with cluster samples, the robust standard error can be generated from

$$Var\left(\hat{\boldsymbol{\beta}}\right)_{robust} = \left(-\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\sum_{t=1}^{T_i}\boldsymbol{x_{it}^*}\boldsymbol{x_{it}^{*\,T}}\right)^{-1}\left[\sum_{i=1}^{n}\left(\sum_{t=1}^{T_i}\frac{1}{\hat{\sigma}^2}\boldsymbol{x_{it}^*}e_{it}\right)\left(\sum_{t=1}^{T_i}\frac{1}{\hat{\sigma}^2}e_{it}\boldsymbol{x_{it}^{*\,T}}\right)\right]\left(-\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\sum_{t=1}^{T_i}\boldsymbol{x_{it}^*}\boldsymbol{x_{it}^{*\,T}}\right)^{-1},$$

[11]Regarding numerical issues and technical model information, please consult with the supplementary materials.

where T refers to a temporal or spatial unit, n is the overall sample size, and T_i is the sample size for the ith unit. (Greene, 2008, 515)

To evaluate the sensitivity of hypothesis testing when different models are applied, the author compares the mean absolute deviation of the four parameter estimates and their corresponding t statistics, a measure that indicates the variability of the regression result and the significance level of parameter estimates per trial.

4. SIMULATIONS

To understand how the TRM model performs with or without boundary constraints, three simulation tests are carried out to evaluate the validity of parameter estimation. For each simulation, two independent variables (x_1 and x_2) are included in the regression model. The dependent variable was randomly generated as truncated normal following $y_i \sim TN\,(\mu, \sigma; 0, 1)$, where $\mu \sim U\,(0.05, 0.95)$ and $\sigma \sim U\,(\min{(1-\mu, \mu)}/5\ , \max{(1-\mu, \mu)})$. The independent variables are generated by varying explanatory power, degree of nonlinearity, and degree of multicollinearity. The sampling scheme can be specified as:

$$\begin{array}{rl} \text{Simulation I} & x_1 = U,\ x_2 = U \cdot w_1 + y\,(1-w_1) \\ \text{Simulation II} & x_1 = U,\ x_2 = U \cdot w_1 + logit(y)\,(1-w_1) \\ \text{Simulation III} & x_1 = U \cdot w_2 + x_2\,(1-w_2)\,,\ x_2 = U \cdot w_1 + y\,(1-w_1)\,, \end{array}$$

where U refers to a continuous uniform random variable following $U\,(0,1)$, $logit(y) = \ln{(y/1-y\ )}$, $w_i \in (0,1)$, and the sampling process of three simulations are independent. Except for x_2 in Simulation II, x_j is bounded within $(0,1)$.[12]

For the first simulation, the linear relationship between x_2 and y is assumed and the explanatory power is completely decided by w_1 given the independence of x_1 and y. When

[12]The logit transformation in the deterministic part of x_2 enlarges its range from -8.66 to 13.79 in Simulation II.

w_1 approaches 0, the random part of x_2 is zero, and the deterministic part of x_2 makes r-squared approach 1. When w_1 approaches 1, x_2 is entirely composed of the random part and r-squared approaches 0. For the second simulation, a nonlinear logistic relationship is set to the relationship of x_2 and y, while x_1 is independent from y. When w_1 approaches 0, x_2 only contains the deterministic part, and it causes strong ceiling and floor effects, which seriously violate boundary restrictions when a linear regression is applied. On the other hand, when w_1 approaches 1, the floor and ceiling effects vanish, and x_2 and y becomes independent. For the third simulation, x_2 is first drawn with varying degree of explanatory power (decided by w_1), and then x_1 is drawn with a certain ratio (decided by w_2) of the deterministic part x_2 and the random part U. When w_2 approaches 0, x_1 is perfectly collinear with x_2. When w_2 approaches 1, x_1 and x_2 are completely independent of each other.

Regarding the numerical setups, the maximum iteration is 100, and the tolerance value is set to 10^{-4}. The criteria to evaluate the validity of parameter estimates are admissible solution and greater loglikelihood. Admissible solution refers to the non-violation of boundary restrictions from g_1 to g_{2m+6}, where $m = 2$. A greater loglikelihood value also indicates a better solution if admissibility is satisfied. Any inadmissible solution is regarded as a failed estimate, and the comparison of the loglikelihood value is only carried out when the TRM and TRMCO models generate admissible estimates.

Both the TRM and TRMCO adopt the centering specification in the regression model. The initial value for the TRMCO model is the parameter estimate of the TRM model. According to the theory of nonlinear programming, when the TRMCO solution satisfies the KKT conditions, it generates the best lower bound of optimality for the Lagrange dual function given the property of weak duality. (Boyd and Vandenberghe, 2004, 225) Therefore, the TRMCO solution could theoretically be inferior to the TRM solution, since the latter only needs to satisfy boundary restrictions, which is only one of the four KKT conditions. Whether the loglikelihood value is larger for the TRM or TRMCO solution is an empirical question. When the TRMCO model reaches the maximum number of iterations, the iteration that generates the largest loglikelihood and satisfies the boundary restrictions is reported, except in the first iteration.[13]

Table 1 presents the results of three simulation tests. The TRM model generates admissible solutions by 35.5%, 11.4%, and 30.6% in the three simulations, while the solutions of the TRMCO model are all 100% admissible. The lower percentage of the second simulation for the TRM model is associated with the serious boundary violation by the designed sampling scheme. Apparently, the TRMCO model is more reliable and not subject to the problem of nonlinearity or multicollearity.

When we consider which method performs better in parameter estimation, the TRMCO model has a better estimate (admissible solution + greater loglikelihood) for 100% in all three simulations. Again, this result indicates the superiority of the TRMCO model. Breaking down the cases by the OLS out-of-bounds violation, we find that the TRM model cannot correct boundary violations at all (0%), but the TRMCO model is capable of doing so (100%). Regarding those cases that do not have the OLS out-of-bounds violation, the TRM model only has admissible solutions in 53.8%, 60.6%, and 44.9% for the three simulations, while TRMCO has no instances of boundary violations.

Another key factor in the sampling scheme is the degree of explanatory power. We break down the cases by different r-squared measures into six categories: [0,0.1], (0.1,0.3], (0.3,0.5], (0.5,0.7], (0.7,0.9], (0.9,1]. The result consistently shows that the TRMCO model performs in an unconditionally superior fashion (100% admissible solution), regardless of r-squared, nonlinearity, and multicollinearity. With regard to the TRM model, the performance quickly worsens if r-squared is larger than 0.1. The successful rate for admissible solutions when r-squared is smaller than 0.1 is 72.8%, 70.4%, and 58.9% for the three simulations, respectively. It is reduced to less than 40% when r-squared is between 0.1 and 0.3 and, continues dropping

[13] According to the weak duality theorem, violation of the complementary slackness condition (the fourth KKT conditions) sometimes occurs and leads to nonconvergence because of the duality gap. (Murty, 2010, 260-261) When the estimation process reaches the maximum iterations, the best eligible TRMCO estimate is considered a reasonable result to report if the boundary restrictions are satisfied. Otherwise, the result is marked as an ineligible solution. The same criterion is applied to the TRM estimate under the Stata environment.

to 20% when r-squared is between 0.3 and 0.5, and finally falls below 15% when r-squared is above 0.5.

Regardless of the superiority of the TRMCO model, it is informative to understand the difference in the solutions from both methods. Table 2 crosstabulates the mean absolute deviation of the estimates ($\hat{\beta}_0$,$\hat{\beta}_1$,$\hat{\beta}_2$,$\hat{\sigma}$) for the TRM model, by using the TRMCO solution as the default answer. The result indicates very little difference between all parameter estimates when the solution of the TRM model has no boundary violations. The mean absolute deviation of the parameter estimates is at most 0.001 and is almost negligible. On the other hand, when the TRM model fails to generate an admissible solution, the margin of difference is significantly larger, particularly when the multicollinearity is high, except for the scale parameter σ.

The mean absolute deviation of the t statistic can provide further information regarding the sensibility of the hypothesis testing results. As Tables 2 shows, t_{β_0}, t_{β_2}, and t_σ have milder variability when the solution of the TRM model is admissible. However, the variability significantly increases when the TRM model has boundary violations, regardless of different sampling schemes. This result indicates that hypothesis testing can be greatly affected if the TRM fails to find an admissible solution. Note that t_{β_1} does have little variability in all three simulations since x_1 and y in the sampling scheme are supposed to be independent (Simulation I and II) or highly collinear to x_2 (Simulation III). Given that little explanatory power is designed to x_1, the variability is marginal.

Some cases of the TRM model suffer from the problem of negative variance. All of these cases occur in Simulation II in which the strong nonlinear logit relationship is assumed between x_2 and y. This result indicates that negative variance becomes a serious problem for the TRM model when the boundary violation is extreme.[14] As Table 2 shows, the TRM model generates negative variance in 33.4% (296/886) of the cases when it fails to find an admissible solution. This reflects the weakness of the TRM model in handling data for which

[14]When ceiling or floor effects are very strong, negative variance happens and this indicates strong violation of the linear specification of the regression model. (Kolenikov and Bollen, 2012, 6)

the linear assumption has been significantly violated.

Although the difference of the TRMCO and TRM models is limited when the latter solution has no boundary violations, the weakness of handling various data properties, such as varying explanatory power, nonlinearity, and multicollinearity, as well as the problem of negative variance, all makes the TRM model highly unreliable. Thus, the overall findings of the three simulation tests consistently suggest that the TRMCO model is superior to the TRM model.

5. A REPLICATION STUDY

To demonstrate how TRMCO can be applied to political studies and what difference TRMCO makes in the current method, the author presents a replication study of the aforementioned 2007 Hellwig and Samuels article. The reason for choosing this work for replication is threefold. First, the dependent variable is the vote share of the incumbent party, which perfectly fits the distributional assumption of the truncated normal distribution and is naturally bounded between 0 and 1. Second, the dataset is available for public access and anyone can replicate the analysis. Third, the original analysis adopts the OLS model without any complication, and this makes the replication study easier.

There are two regression models in Hellwig and Samuels's article. Both models apply incumbent party's vote-share as the dependent variable, and use 13 independent variables in each regression with minor changes. The estimation is conducted with the OLS method, and robust standard errors are reported for hypothesis testing (Stata). In Model I, the major explanatory variables include *Previous Vote* (the incumbent party's vote share in the previous election), *Economy* (annual percentage change in real per capital GDP), and *Trade Openness* (the sum of the country's exports and imports as a proportion of its GDP). The control variables are *Presidential Election* (a dummy variable), *Re-election* (a dummy variable based on whether the incumbent president was running for re-election), *Effective Number of Parties*, *Income* (thousand of constant U.S. dollars), and four regional dummy variables *Africa*, *Asia*, *Central and Eastern Europe*, and *Latin America and the Caribbean* (using the advanced industrial democracies as the baseline category). Two remaining variables are interaction terms *Economy*×*Trade Openness* and *Economy*×*Presidential Election*. Model II replaces the explanatory variable *Trade Openness* with *Capital Flows* (gross private capital flows as a share of GDP), and thus, the interaction term *Economy*×*Trade Openness* is replaced with *Economy*×*Capital Flows*. The rest of the model specifications remain the same. The sample size is 426, including all democracies (+6 or better in Polity IV's ranking of democratic quality) in the world from 1975 to 2002.

Two adjustments are made to the TRMCO model. First, instead of centering by the means, the author adopts a model specification that fixes all independent variables at the minimum level for handling dummy variables.[15] Centering dummy variables makes no sense in interpretation, especially for those unbalanced dummies that could cause numerical problems in parameter estimation.[16] Therefore, the constant estimate of the OLS model is different from what was originally reported. Second, given that the truncated normal distribution is a more plausible distributional assumption, the criteria for model comparison should be based on the log pseudolikelihood function of the TRMCO model, as well as admissibility of the parameter estimates.[17]

Due to the spatial dependence of the sample, robust standard error is applied for hypothesis testing. For each model, the author compares three methods (OLS, TRM, and TRMCO), with three criteria: first, which methods produces the greatest values of the log pseudolikelihood function; second, whether the predicted value of vote-share is admissible; third, whether the parameter estimate complies to the boundary constraints. Only the

[15] We do not fix the original interaction variables to the minimum. Instead, we fix all the non-interaction variables at the minimum value first, and then compute the crossproducts to generate two interaction terms.

[16] Consider that the regional dummy variable *Africa* only has 5.6% of the cases in the overall sample. If it is centered to the mean, the centered dummy only has the value of either −.056 or .944, of which the former is very small as a denominator and would sometimes lead to numerical problems in the estimation process.

[17] Since the data are clustered samples and violate the *i.i.d.* assumption, the product of all the likelihood function, regardless of cluster dependence, is the so-called "pseudo-likelihood" function. (Strauss and Ikeda, 1990)

answers satisfied with the latter two criteria are eligible for log pseudolikelihood comparison.

Regarding the numerical estimation of the TRMCO model, we use TRM's solution as the initial value. The number of maximum iterations and the tolerance value remain the same as 100 and 10^{-4}. In addition, since all independent variables are fixed at minimum, the boundary constraint for $\hat{\beta}_m$ is modified correspondingly

$$\frac{a - \hat{y}_{\sim m}^{\min}}{x_m^{\max} - x_m^{\min}} \leq \hat{\beta}_m \leq \frac{b - \hat{y}_{\sim m}^{\max}}{x_m^{\max} - x_m^{\min}}, \tag{5}$$

where

$$\hat{y}_{\sim m}^{\max} = \begin{cases} \hat{\beta}_0 + \sum\limits_{j \neq m} v_j^+ \hat{\beta}_j \left(x_j - x_j^{\min}\right)^{\max} + Max\left(\beta_{d1}, \cdots, \beta_{d2}\right) & \text{(independent vars)} \\ \hat{\beta}_0 + \sum v_j^+ \hat{\beta}_j \left(x_j - x_j^{\min}\right)^{\max} & \text{(regional dummies)}, \end{cases}$$

$$\hat{y}_{\sim m}^{\min} = \begin{cases} \hat{\beta}_0 + \sum\limits_{j \neq m} v_j^- \hat{\beta}_j \left(x_j - x_j^{\min}\right)^{\max} + Min\left(\beta_{d1}, \cdots, \beta_{d2}\right) & \text{(independent vars)} \\ \hat{\beta}_0 + \sum v_j^- \hat{\beta}_j \left(x_j - x_j^{\min}\right)^{\max} & \text{(regional dummies)}, \end{cases} \tag{6}$$

in which β_{dj} represents a dummy variable, and we drop the terms $\left(x_j - x_j^{\min}\right)^{\min}$ since they are all zero.[18] As the appendix makes evident, we can easily prove that (5) applies to both pure independent and interaction variables.

Table 3 shows replication results of Model I by three methods. TRMCO is significantly different from TRM and OLS. For the TRMCO model, *Economy*,*Trade Openness*, and *Income* are no more significant, while they are positively significant in the other two models. None of the regional dummies is significant in the TRMCO model, but the OLS and TRM models have one and two significant results, respectively. Besides, the negative relationship of the interaction term *Economy*×*Trade Openness* does not hold in the TRMCO model either. Among the three methods, the OLS model has the greatest number of significant

[18]For the interaction term, the indicator variables v_j^+ and v_j^- are not independently decided by its own beta coefficient. Rather, they are decided by the signs of the composing variable's beta coefficients. For instance, if $x_7^* = x_5^* \times x_6^*$, then $v_7^+ = 1$ when $sign(\beta_5) \times sign(\beta_6) > 0$, and $v_7^- = 1$ when $sign(\beta_5) \times sign(\beta_6) < 0$.

results (9), larger than TRM (8) and TRMCO (4). This result indicates a great variability of causal analysis when different methods are applied.

In terms of model performance, the TRM suffers the problem of inadmissible predicted values $\hat{y}^{\min} = -7.938$ and 13 boundary violations. The OLS model also has an inadmissible predicted value $\hat{y}^{\min} = -4.797$ and 13 boundary violations. The only eligible solution is generated from the TRMCO model, which has a slightly lower log psedolikelihood value, but the solution is admissible and no boundary violation occurs. Apparently, the TRMCO model has the best performance among the three.

Similar findings are concluded in the replication results of Model II in Table 4. Again, *Economy* is not significant in the TRMCO model, and the adjusted constant is significantly larger than zero but not in the OLS or TRM model. For *Presidential Election* and *Economy*$\times$*Presidential Election*, the TRMCO model shows a significant negative and positive relationship, but the two findings do not appear in the OLS and TRM models. For the rest of the parameter estimates findings, while the significance tests show the same result, the beta estimates are somewhat different. In sum, the TRMCO model has the greatest number of significant results (8), larger than OLS (6) and TRM (6), and apparently, the regression results also show a great variability when different methods are applied.

The OLS and TRM models still suffer the out-of-bounds predicted values in Model II, and they have 12 and 13 boundary violations, respectively. In contrast, the TRMCO model does not have the above problems and performs better. Based on the above results, we can conclude that the TRMCO model is a superior method to the current models in use. This conclusion casts doubt on the inferential validity of the current methods, such as the OLS or TRM model when the dependent variable fits the truncated normal assumption better.

At last, to understand what those boundary violations are about, we translate the type II boundary violations into type I violations as Table 5 shows. The translation proceeds as follows: First, we hold the baseline profile at the combination that generates the greatest or least predicted value while fixing the predictor in interest at its minimum. Second, by varying the predictor from its minimum to maximum, we add its contribution, which results in the greatest and least predicted value. Third, we evaluate whether the parameter estimate has a type-II violation by checking its admissibility. As Table 5 shows, TRM has 13 boundary violations in both models, while TRMCO's predicted values are all admissible. Most of the boundary violations for the TRM solution are lower-bound violations, except for *Capital Flow* in Model II. This result nicely illustrates why we need to adopt the TRMCO model as a replacement for the TRM and OLS models.

6. CONCLUSIONS

The current truncated regression model suffers significantly from boundary violations. Under no circumstances can an ineligible solution achieve inferential validity. This article demonstrates that this problem is widespread in the OLS and TRM models when a regression analysis is applied to a truncated normal dependent variable. To resolve this problem, the author proposes a modified truncated regression model (TRMCO) by incorporating the techniques of constrained optimization and successfully eliminates boundary violations and generates admissible and interpretable results. The major contribution of this article is twofold. First, the application of the non-linear programming method SQP successfully solves the boundary violation problems in the parameter estimation process of maximum likelihood. Second, this article provides simulation evidence and a replication study to demonstrate the superiority of TRMCO over the existing model.

The findings in this article have profound implications for statistical theory as well as empirical application. From a theoretical perspective, the plausibility of the distributional assumption for the dependent variable is critical to inferential validity. When boundary limits exist for a normal random variable, the failure to specify boundary constraints would lead to an invalid statistical inference. Unlike TRMCO, the current model does not solve the problem of boundary violations. Nor does the existing literature include relevant discussions regarding how boundary violations affect the validity and interpretability of the regression result. This article proves that the boundary violations can be fixed, and hence, no compromise should be made to accept those ineligible results.

From the empirical perspective, this article demonstrates how to work directly with a truncated normal distribution by maximum likelihood under the framework of constrained optimization. This involves the setup of boundary constraints with the specification of the centered or fixed model and the application of sequential quadratic programming algorithm. Together, those efforts engender a new regression method that is exempt from boundary violations.

Many studies in political science analyze truncated normal dependent variables. In addition to party's vote share, any variable that uses the percentage measure is likely subject to boundary restrictions, such as voter turnout or politician's approval rating. However, some variables do have boundary restrictions, but these restrictions are largely neglected, since the untruncated normal distribution works fairly well. These variables include test scores and effective number of parties. Still, other variables, such as media exposure or formal education, have implicit boundary restrictions, but researchers are often unaware of their existence. Given the situations discussed above, it is strongly recommended that researchers compare the results of their original model with the TRMCO model and check the robustness of regression outcomes. Otherwise, inferential validity could be seriously compromised if boundary violations actually occur.

APPENDIX

This following appendix contains three proofs of the boundary constraints for pure independent variables, interaction variables, and dummy variables. The purpose is to illustrate that all of the boundary constraints can be generalized by (5).

Notation

- $x_j^\dagger$: A column vector of the covariate matrix after being fixed at the minimum.

- β_0: Baseline predicted vote share when all covariates are fixed at the minimum, except the dummies which are all set to 0.

- β_1 to β_4: Beta coefficient estimates for the pure independent variables, *Previous Votes*, *Re-election*, *Effective Number of Parties*, and *Income*, respectively.

- β_5 to β_9: Beta coefficient estimates associated with the interaction variables, *Economy*, *Trade Openness*, *Economy*$\times$*Trade Openness*, *Presidential Election*, and *Economy*$\times$*Presidential Election*, respectively.

- β_{10} to β_{13}: Beta coefficient estimates for the regional dummy variables, *Africa*, *Asia*, *Central and Eastern Europe*, and *Latin America and the Caribbean*, respectively.

Case I: Pure Independent Variables

Given the overall predicted vote share

$$\hat{y}^{\max} = \hat{\beta}_0 + \sum_{i=1}^{4} v_i^+ \hat{\beta}_i x_i^{\dagger\max} + \sum_{j=5}^{9} v_j^+ \hat{\beta}_j x_j^{\dagger\max} + \max\left(\hat{\beta}_{10}, \hat{\beta}_{11}, \hat{\beta}_{12}, \hat{\beta}_{13}\right)$$
$$\hat{y}^{\min} = \hat{\beta}_0 + \sum_{i=1}^{4} v_i^- \hat{\beta}_i x_i^{\dagger\max} + \sum_{j=5}^{9} v_j^- \hat{\beta}_j x_j^{\dagger\max} + \min\left(\hat{\beta}_{10}, \hat{\beta}_{11}, \hat{\beta}_{12}, \hat{\beta}_{13}\right),$$

we know

$$\hat{y}_{\sim j}^{\max} = \hat{y}^{\max} - v_j^+ \hat{\beta}_j x_j^{\dagger\max}$$
$$\hat{y}_{\sim j}^{\min} = \hat{y}^{\min} - v_j^- \hat{\beta}_j x_j^{\dagger\max},$$

where $j = \{1, 2, 3, 4\}$.

Since $a \leq \hat{y}^{\min} \leq \hat{y}^{\max} \leq b$,

$$y_{\sim j}^{\max} + v_j^{+}\beta_j x_j^{\dagger\max} \leq b$$

$$y_{\sim j}^{\min} + v_j^{-}\beta_j x_j^{\dagger\max} \geq a$$

Therefore,

$$\frac{a - \hat{y}_{\sim j}^{\min}}{x_j^{\dagger\max}} \leq \beta_j \leq \frac{b - \hat{y}_{\sim j}^{\max}}{x_j^{\dagger\max}}.$$

□

Case II: Interaction Variables

To simplify the proof, we only present the upper bound constraint. The same proof can easily be applied to the lower bound constraint. In the following proof, we first deal with β_5, beta coefficient of *Economy*, one of the composition variables for *Economy*×*Trade Openness*(β_7) and *Economy*×*Presidential Election*(β_9).

Given

$$\hat{y}_{\sim 5}^{\max} = \hat{y}^{\max} - v_5^{+}\hat{\beta}_5 x_5^{\dagger\max} - v_7^{+}\hat{\beta}_7 x_7^{\dagger\max} - v_9^{+}\hat{\beta}_9 x_9^{\dagger\max},$$

we know

$$\hat{y}^{\max} - \hat{y}_{\sim 5}^{\max} = v_5^{+}\beta_5 x_5^{\dagger\max} + v_7^{+}\beta_7 x_7^{\dagger\max} + v_9^{+}\beta_9 x_9^{\dagger\max} \leq b - \hat{y}_{\sim 5}^{\max}.$$

Therefore,

$$\beta_5 \leq \frac{b - \hat{y}_{\sim 5}^{\max} - v_7^{+}\beta_7 x_7^{\dagger\max} - v_9^{+}\beta_9 x_9^{\dagger\max}}{x_5^{\dagger\max}},$$

and (5) can be generalized to describe this boundary constraint if we change the definition of $\hat{y}_{\sim 5}^{\max}$ as (6) states.

For the resultant variables, such as *Economy*$\times$*Trade Openness*,

$$\hat{y}_{\sim 7}^{\max} = \begin{cases} \hat{y}^{\max} - v_5^+ \hat{\beta}_5 x_5^{\dagger\max} & \\ \qquad - v_7^+ \hat{\beta}_7 x_7^{\dagger\max} - v_9^+ \hat{\beta}_9 x_9^{\dagger\max} & \text{if} \quad v_6^+ \beta_6 x_6^{\dagger\max} \geq v_5^+ \beta_5 x_5^{\dagger\max} + v_9^+ \beta_9 x_9^{\dagger\max} \\ \hat{y}^{\max} - v_6^+ \hat{\beta}_6 x_6^{\dagger\max} - v_7^+ \hat{\beta}_7 x_7^{\dagger\max} & \text{otherwise} \end{cases}$$

If $v_6^+ \hat{\beta}_6 x_6^{\dagger\max} \geq v_5^+ \hat{\beta}_5 x_5^{\dagger\max} + v_9^+ \hat{\beta}_9 x_9^{\dagger\max}$,

$$\begin{aligned} b - \hat{y}_{\sim 7}^{\max} &\geq v_5^+ \beta_5 x_5^{\dagger\max} + v_7^+ \beta_7 x_7^{\dagger\max} + v_9^+ \beta_9 x_9^{\dagger\max} \\ \hat{\beta}_7 &\leq \frac{b - \hat{y}_{\sim 7}^{\max} - v_5^+ \hat{\beta}_5 x_5^{\dagger\max} - v_9^+ \hat{\beta}_9 x_9^{\dagger\max}}{x_7^{\dagger\max}}; \end{aligned}$$

otherwise,

$$\begin{aligned} b - \hat{y}_{\sim 7}^{\max} &\geq v_6^+ \hat{\beta}_6 x_6^{\dagger\max} + v_7^+ \hat{\beta}_7 x_7^{\dagger\max} \\ \hat{\beta}_7 &\leq \frac{b - \hat{y}_{\sim 7}^{\max} - v_6^+ \hat{\beta}_6 x_6^{\dagger\max}}{x_7^{\dagger\max}}. \end{aligned}$$

Both cases can be generalized by (5) and (6). $\square$

Case III: Regional Dummy Variables

For a regional dummy variable $x_j^{\dagger}$, where $j = \{10, 11, 12, 13\}$,

$$\hat{y}_{\sim 10}^{\max} = \hat{y}^{\max} - \max\left(\hat{\beta}_{10}, \hat{\beta}_{11}, \hat{\beta}_{12}, \hat{\beta}_{13}\right).$$

Thus, we can derive

$$\hat{\beta}_j \leq b - \hat{y}_{\sim 10}^{\max},$$

and this relationship can be generalized by (5) and (6) since $x_{\sim j}^{\dagger\max} = 1$. $\square$

REFERENCES

Aldrich, J. H. and F. D. Nelson (1984). *Linear Probability, Logit, and Probit Models*. Thousand Oaks, C.A.: Sage Publications.

Amemiya, T. (1973). Regression analysis when the dependent variable is truncated normal. *Econometrica 41*(6), 997–1016.

Amemiya, T. (1984). Tobit models: A survey. *Journal of Econometrics 24*(1-2), 3–61.

Berger, M. C. and E. F. Toma (1994). Variation in state education policies and effects on student performance. *Journal of Policy Analysis and Management 13*(3), 477–91.

Bertsekas, D. P. (1996). *Constrained Optimization and Lagrange Multiplier Methods*. Belmont, M.A.: Athena Scientific.

Bonnans, F. J., J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal (2006). *Numerical Optimization: Theoretical and Practical Aspects*. New York: Springer.

Boyd, S. and L. Vandenberghe (2004). *Convex Optimization*. Cambridge: Cambridge University Press.

Breen, R. (1996). *Regression Models: Censored, Sample Selected, or Truncated Data*. Thousand Oaks, C.A.: Sage Publications.

Chung, C. F. and A. S. Goldberger (1984). Proportional projections in limited dependent variable models. *Econometrica 52*(2), 531–34.

Cong, R. (2000). Truncated regression. *Stata Technical Bulletin 9*(52), 47–52.

Geys, B. (2006). Explaining voter turnout: A review of aggregate-level research. *Electoral Studies 25*(4), 637–63.

Greene, W. (1983). Estimation of limited dependent variable models by ordinary least squares and the method of moments. *Journal of Econometrics 21*(2), 195–212.

Greene, W. (1999). Frontier production functions. In M. H. Pesaran and P. Schmidt (Eds.), *Handbook of Applied Econometrics Volume II: Microeconomics* (2 ed.). Oxford: Blackwell Publisher.

Greene, W. (2008). *Econometric Analysis* (6 ed.). Upper Saddle River, N.J.: Pearson Prentice Hall.

Hausman, J. A. and D. A. Wise (1977). Proportional projections in limited dependent variable models. *Econometrica 45*(4), 919–38.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica 45*(4), 153–61.

Henry, G. T. and R. Rubenstein (2002). Paying for grades: Impact of merit-based financial aid on educational quality. *Journal of Policy Analysis and Management 21*(1), 93–109.

Holbrook, T. M. and S. D. McClurg (2005). The mobilization of core supporters: Campaigns, turnout, and electoral composition in united states presidential elections. *American Journal of Political Science 49*(4), 689–703.

Jewell, K.-L. T. N. P. and C. F. J. Wu (1988). A nonparametric approach to the truncated regression problem. *Journal of the American Statistical Association 83*(403), 785–92.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1970). *Distributions in Statistics: Continuous Univariate Distribution-2*. New York: John Wiley & Sons.

Kernell, S. (1978). Explaining presidential popularity. how ad hoc theorizing, misplaced emphasis, and insufficient care in measuring one's variables refuted common sense and led conventional wisdom down the path of anomalies. *American Political Science Review 72*(2), 506–22.

Knack, S. (2004). Does foreign aid promote democracy? *International Studies Quarterly 48*(1), 251–66.

Kolenikov, S. and K. Bollen (2012). Testing negative error variances: Is a heywood case a symptom of misspecification? *Sociological Methods and Research (forthcoming).*

Kuhn, H. W. and A. W. Tucker (1951). Nonlinear programming. In *Proceedings of 2nd Berkeley Symposium*. Berkeley, C.A.: University of California Press.

Mebane, W. R. and J. S. Sekhon (2011). Genetic optimization using derivatives: The rgenoud package for r. *Journal of Statistical Software 42*(11), 1–26.

Moe, T. M. (1980). A calculus of group membership. *American Journal of Political Science 24*(4), 593–632.

Murty, K. G. (2010). *Optimization for Decision Making: Linear and Quadratic Models*. New York: Springer.

Nocedal, J. and S. Wright (1999). *Numerical Optimization*. New York: Springer.

Olsen, R. J. (1978). Note on the uniqueness of the maximum likelihood estimator for the tobit model. *Econometrica 46*(5), 1211–15.

Orme, C. D. (1989). On the uniqueness of the maximum likelihood estimator in truncated regression models. *Econometric Reviews 8*(2), 217–22.

Orme, C. D. and P. A. Ruud (2002). On the uniqueness of the maximum likelihood estimator. *Economics Letters 75*(2), 209–21.

Rudra, N. (2005). Globalization and the strengthening of democracy in the developing world. *American Journal of Political Science 49*(4), 704–30.

Sartori, A. E. (2003). An estimator for some binary-outcome selection models without exclusion restriction. *Political Analysis 11*(2), 111–38.

Sekhon, J. S. and W. R. Mebane (1998). Genetic optimization using derivatives: Theory and application to nonlinear models. *Political Analysis 7*(1), 187–210.

Sigelman, L. and L. Zeng (1999). Analyzing censored and sample-selected data with tobit and heckit models. *Political Analysis 8*(2), 167–82.

Sorokin, G. L. (1994). Arms, alliances, and security tradeoffs in enduring rivalries. *International Studies Quarterly 38*(3), 421–46.

Strauss, D. and M. Ikeda (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association 85*(409), 204–12.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica 26*(1), 24–36.

West, D. M. (1991). Television and presidential popularity in america. *British Journal of Political Science 21*(2), 199–214.

	Simulation I		Simulation II		Simulation III	
	TRM	TRMCO	TRM	TRMCO	TRM	TRMCO
Better Estimate	0%	100%	0%	100%	0%	100%
	(0/1000)	(1000/1000)	(0/1000)	(1000/1000)	(0/1000)	(1000/1000)
Admissible Solution	35.5%	100%	11.4%	100%	30.6%	100%
	(355/1000)	(1000/1000)	(114/1000)	(1000/1000)	(306/1000)	(1000/1000)
OLS Predicted Value	Admissible Solution					
Out-of-bounds Violation	0%	100%	0%	100%	0%	100%
	(0/340)	(340/340)	(0/812)	(812/812)	(0/318)	(318/318)
No Violation	53.8%	100%	60.6%	100%	44.9%	100%
	(355/660)	(660/660)	(114/188)	(188/188)	(306/682)	(682/682)
R-Squared	Admissible Solution					
$[0, 0.1]$	72.8%	100%	70.4%	100%	58.9%	100%
	(217/298)	(298/298)	(57/81)	(81/81)	(186/316)	(316/316)
$(0.1, 0.3]$	35.7%	100%	32.8%	100%	34.3%	100%
	(61/171)	(171/171)	(22/67)	(67/67)	(61/178)	(178/178)
$(0.3, 0.5]$	22.4%	100%	21.9%	100%	17.0%	100%
	(22/98)	(98/98)	(14/64)	(64/64)	(17/100)	(100/100)
$(0.5, 0.7]$	13.8%	100%	6.7%	100%	8.9%	100%
	(12/87)	(87/87)	(7/105)	(105/105)	(8/90)	(90/90)
$(0.7, 0.9]$	10.8%	100%	1.4%	100%	6.7%	100%
	(15/139)	(139/139)	(6/425)	(425/425)	(8/119)	(119/119)
$(0.9, 1]$	13.5%	100%	3.1%	100%	13.2%	100%
	(28/207)	(207/207)	(8/258)	(258/258)	(26/197)	(197/197)

Table 1: Simulation Results for the TRM and TRMCO model

	Simulation I		Simulation II		Simulation III	
	Admissible	Inadmissible	Admissible	Inadmissible	Admissible	Inadmissible
			Mean Absolute Deviation			
$\hat{\beta}_0$	.0003	.4015	.0002	.0346	.0003	.1154
$\hat{\beta}_1$	.0006	.8814	.0004	.0208	.0010	1.7016
$\hat{\beta}_2$	.0008	.4865	.0007	.2068	.0011	1.5170
$\hat{\sigma}$	.0004	.0570	.0002	.0529	.0002	.0561
			Mean Absolute Deviation			
t_{β_0}	5.455	29.013	.287	58.091	5.043	120.165
t_{β_1}	.016	.856	.009	.724	.012	.812
t_{β_2}	2.208	18.900	.037	42.525	.799	37.991
t_{σ}	.075	4.740	.036	26.280	.059	4.472
N	355	645	114	886(590)[a]	306	694

[a] There are 296 cases of failed TRM estimates on σ, and thus, only 590 valid cases for σ and t_{σ}. The error message indicates Stata has a problem computing the numerical derivative in the maximum likelihood estimation. However, this problem does not appear in the TRMCO method, and thus, the number of valid cases is still 886.

Table 2: Comparison of the Parameter Estimates

	OLS		TRM		TRMCO	
Independent Variables	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Previous Votes	.478**	.079	.470**	.079	.464**	.084
Economy	.827**	.237	.873**	.251	-.000	.303
Trade Openness	15.108*	5.965	15.427*	6.051	-.000	5.524
Economy×Trade Openness	-.710*	.313	-.710*	.322	-.000	.292
Presidential Election	-6.025	5.929	-3.704	5.639	-.000	5.251
Economy×Presidential Election	.261	.313	.165	.303	.094	.279
Re-election	6.151**	1.928	5.927**	1.860	4.162*	1.994
Effective number of parties	-2.959**	.484	-3.425**	.544	-1.940**	.505
Income	.172**	.043	.190**	.045	.110	.065
Africa	3.372	3.151	3.199	3.172	4.267	3.501
Asia	2.679*	1.190	2.425	1.248	.837	1.910
Central and Eastern Europe	-3.579	1.935	-3.679	2.101	-.446	2.467
Latin America and the Caribbean	2.957*	1.466	3.253*	1.432	.701	1.887
Adjusted Constant (fixed at x_j^{min})	2.916	6.054	2.756	6.245	18.986**	6.833
$\hat{\sigma}$	9.100		8.947		9.434	
Log pseudolikelihood	-1523.422		-1522.447		-1551.544	
Empirical $\hat{y}^{\max}$	64.545		64.865		63.550	
Empirical $\hat{y}^{\min}$	-4.797		-7.938		3.402	
Boundary Violations	13		13		0	
N	424		424		424	

Table 3: Replication of Model I, Hellwig and Samuels (2007: 292)

	OLS		TRM		TRMCO	
Independent Variables	Coefficient	*SE*	Coefficient	*SE*	Coefficient	*SE*
Previous Votes	.495**	.081	.488**	.081	.501**	.080
Economy	.486*	.219	.571**	.215	.000	.281
Capital Flows	17.885	25.966	27.465	24.925	.000	33.010
Economy×Capital Flows	-.887	1.172	-1.316	1.124	.000	1.548
Presidential Election	-5.944	5.468	-4.569	5.354	-12.344*	5.449
Economy×Presidential Election	.268	.286	.223	.282	.669*	.285
Re-election	5.149**	1.799	4.896**	1.717	4.302*	1.765
Effective number of parties	-2.952**	.502	-3.359**	.577	-2.551**	.478
Income	.177**	.051	.180**	.051	.151**	.054
Africa	7.310**	2.755	6.969*	2.752	7.218*	2.808
Asia	2.143	1.246	2.064	1.292	1.845	1.325
Central and Eastern Europe	-3.514	2.115	-4.081	2.239	-2.164	2.125
Latin America and the Caribbean	2.774	1.497	2.862	1.491	1.735	1.530
Adjusted Constant (fixed at x_j^{min})	9.604	6.033	8.776	6.155	18.244**	6.764
$\hat{\sigma}$	8.981		8.807		9.086	
Log pseudolikelihood	-1479.400		-1479.049		-1486.769	
Empirical $\hat{y}^{\max}$	67.146		67.298		67.410	
Empirical $\hat{y}^{\min}$	-3.164		-5.874		1.510	
Boundary Violations	12		13		0	
N	413		413		413	

Table 4: Replication of Model II, Hellwig and Samuels (2007: 292)

	Model I				Model II			
Predicted Value by Varying	TRM		TRMCO		TRM		TRMCO	
(based on $\hat{y}^{\max}_{\sim m}$ and $\hat{y}^{\min}_{\sim m}$)	max	min	max	min	max	min	max	min
All Variables	85.9	-32.9*	79.5	0	81.3	-22.8*	81.7	0
Previous Votes	85.9	12.2	79.5	44.5	81.3	21.1	81.7	45.1
Economy	85.9	-7.7*	79.5	0	81.3	-6.3*	81.7	0
Trade Openness(Capital Flows)	85.9	-4.5*	79.5	0	135.4*	81.3	81.7	0
Economy×Trade Openness(Capital Flows)	85.9	-66.1*	79.5	0	81.3	-188.8*	81.7	0
Presidential Election	82.2	-32.9*	79.5	0	76.7	-22.8*	69.4	0
Economy×Presidential Election	90.4	-32.9*	79.5	2.6	87.4	-22.8*	100	0
Re-election	85.9	-26.9*	79.5	4.2	81.3	-17.9*	81.7	4.3
Effective Number of Parties	53.2	-32.9*	60.9	0	52.3	-22.8*	59.7	0
Income	85.9	-24.2*	79.5	5.0	81.3	-14.6*	81.7	6.9
Africa	85.9	-26.0*	79.5	4.7	81.3	-11.7*	81.7	9.4
Asia	85.1	-26.8*	76.1	1.3	76.4	-16.6*	76.3	4.0
Central and Eastern Europe	79.0	-32.9*	74.8	0	70.3	-22.8*	72.3	0
Latin America and the Caribbean	85.9	-25.9*	75.9	1.1	77.2	-15.8*	76.2	3.9

*boundary violation

Table 5: Translation of Boundary Violations into Predicted Values For Model I and II