Simulation-Based Two-Step Estimation with Endogenous Regressors

Kamhon Kan
Academia Sinica, Institute of Economics, kan@gate.sinica.edu.tw

Chihwa Kao
Syracuse University, Maxwell School, Center for Policy Research, cdkao@maxwell.syr.edu

Follow this and additional works at: https://surface.syr.edu/cpr

Part of the Econometrics Commons

Recommended Citation
https://surface.syr.edu/cpr/88

This Working Paper is brought to you for free and open access by the Maxwell School of Citizenship and Public Affairs at SURFACE at Syracuse University. It has been accepted for inclusion in Center for Policy Research by an authorized administrator of SURFACE at Syracuse University. For more information, please contact surface@syr.edu.
Center for Policy Research  
Working Paper No. 76  

SIMULATION-BASED TWO-STEP ESTIMATION  
WITH ENDOGENOUS REGRESSORS  

Kamhon Kan and Chihwa Kao*  

December 2005  

$5.00  

Up-to-date information about CPR’s research projects and other activities is available from our World Wide Web site at www-cpr.maxwell.syr.edu. All recent working papers and Policy Briefs can be read and/or printed from there as well.
## CENTER FOR POLICY RESEARCH – Fall 2005

**Timothy Smeeding, Director**  
Professor of Economics & Public Administration

### Associate Directors

<table>
<thead>
<tr>
<th>Margaret Austin</th>
<th>Douglas Holt-Eakin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Associate Director, Budget and Administration</td>
<td>Professor of Economics</td>
</tr>
<tr>
<td></td>
<td>Associate Director, Center for Policy Research</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Douglas Wolf</th>
<th>John Yinger</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professor of Public Administration</td>
<td>Professor of Economics and Public Administration</td>
</tr>
<tr>
<td>Associate Director, Aging Studies Program</td>
<td>Associate Director, Metropolitan Studies Program</td>
</tr>
</tbody>
</table>

### SENIOR RESEARCH ASSOCIATES

<table>
<thead>
<tr>
<th>Badi Baltagi</th>
<th>Jeff Kubik</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pablo Beramendi</td>
<td>Andrew London</td>
</tr>
<tr>
<td>Dan Black</td>
<td>Len Lopoo</td>
</tr>
<tr>
<td>Lloyd Blanchard</td>
<td>Jerry Miner</td>
</tr>
<tr>
<td>William Duncombe</td>
<td>John Moran</td>
</tr>
<tr>
<td>Gary Engelhardt</td>
<td>Jan Ondrich</td>
</tr>
<tr>
<td>Deborah Freund</td>
<td>John Palmer</td>
</tr>
<tr>
<td>Madonna Harrington Meyer</td>
<td>Lori Ploutz-Snyder</td>
</tr>
<tr>
<td>Christine Himes</td>
<td>Grant Reether</td>
</tr>
<tr>
<td>William C. Horrace</td>
<td>Stuart Rosenthal</td>
</tr>
<tr>
<td>Bernard Jump</td>
<td>Ross Rubenstein</td>
</tr>
<tr>
<td>Duke Kao</td>
<td>Margaret Usdansky</td>
</tr>
<tr>
<td>Eric Kingson</td>
<td>Michael Wasylenko</td>
</tr>
<tr>
<td>Thomas Knie snere</td>
<td>Janet Wilmoth</td>
</tr>
<tr>
<td>Tom Krebs</td>
<td>Sociology</td>
</tr>
</tbody>
</table>

### GRADUATE ASSOCIATES

<table>
<thead>
<tr>
<th>Javier Baez</th>
<th>Nadia Greenhalgh-Stanley</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sonali Ballal</td>
<td>Yue Hu</td>
</tr>
<tr>
<td>Dana Balter</td>
<td>Becky Lafrancois</td>
</tr>
<tr>
<td>Jason Bladen</td>
<td>Joseph Marchand</td>
</tr>
<tr>
<td>Jesse Bricker</td>
<td>Larry Miller</td>
</tr>
<tr>
<td>Maria Brown</td>
<td>Wendy Parker</td>
</tr>
<tr>
<td>Yong Chen</td>
<td>Emily Pas</td>
</tr>
<tr>
<td>Ginnie Cronin</td>
<td>Shawn Rohlin</td>
</tr>
<tr>
<td>Ana Dammert</td>
<td>Cynthia Searcy</td>
</tr>
<tr>
<td>Mike Eriksen</td>
<td>Talisha Searcy</td>
</tr>
<tr>
<td>Katie Fitzpatrick</td>
<td>Jeff Thompson</td>
</tr>
<tr>
<td>Alexandre Genest</td>
<td>Wen Wang</td>
</tr>
<tr>
<td>Julie Anna Golobiewski</td>
<td>Yoshikun Yamamoto</td>
</tr>
</tbody>
</table>

### STAFF

<table>
<thead>
<tr>
<th>Kelly Bogart</th>
<th>Kati Foley</th>
</tr>
</thead>
<tbody>
<tr>
<td>Martha Bonney</td>
<td>Administrative Assistant, LIS</td>
</tr>
<tr>
<td>Karen Cimilluca</td>
<td>Kitty Nasto</td>
</tr>
<tr>
<td>Kim Desmond</td>
<td>Administrative Secretary</td>
</tr>
<tr>
<td></td>
<td>Candi Patterson</td>
</tr>
<tr>
<td></td>
<td>Computer Consultant</td>
</tr>
<tr>
<td></td>
<td>Mary Santy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Administrative Secretary</th>
<th>Administrative Secretary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrative Secretary</td>
<td>Administrative Secretary</td>
</tr>
<tr>
<td>Administrative Secretary</td>
<td>Administrative Secretary</td>
</tr>
</tbody>
</table>
Simulation-Based Two-Step Estimation with Endogenous Regressors

Kamhon Kan
The Institute of Economics
Academia Sinica
Taipei, Taiwan
kan@gate.sinica.edu.tw

Chihwa Kao*
Center for Policy Research
Syracuse University
Syracuse, NY 13244-1020
ckhao@maxwell.syr.edu

Fist Draft: January 2005
This Draft: May 24, 2005

1 Introduction

The econometrics of endogeneity is unquestionable one of the most significant contributions in econometrics. The estimation and testing of econometrics models with limited dependent variable (LDV) outcome and discrete/latent endogenous regressors is especially of considerable practical importance. For example, if one wants to estimate the effect of a job training program on later employment by a Probit and include a dummy regressor to denote the treatment status, the dummy regressor may be correlated with the error term in the outcome equation and hence endogenous. This paper considers a control function approach and proposes a simulation-based two-step (STS) estimator for regression models with endogenous latent/discrete regressors. The control function approach treats endogeneity as an omitted variable problem in the same way that the Heckman (1979) two-step estimator corrects for selection bias. It is known that

*We thank seminar participants at Academia Sinica and Syracuse University for helpful comments and suggestions.
the least squares estimator of using the residual from the first stage regression as additional regressor is the same as the two stage least squares (2SLS) estimator. The control function approach has been discussed by Smith and Blundell (1986), Quong and Rivers (1988), Blundell and Smith (1989), Das et al. (2003), Chen and Khan (2003), and Blundell and Powell (2004a, 2004b) to LDV models with continuous endogenous regressors and by Vella (1993, 1998, 1999a, 1999b), Li and Wooldridge (2002) and Christofides et al. (2003) with latent/discrete endogenous regressors.


The remainder of the paper is organized as follows. Section 2 describes the linear model with latent endogenous regressors. The asymptotic properties of the STS estimator are stated in Theorem 1. Section 3 discusses LDV models with latent endogenous regressors. Section 4 presents LDV models with dummy endogenous regressors. Concluding remarks are provided in Section 5. All proofs are given in the Appendix.
2 Linear Model

To motivate the issue we first consider the following equations:

\[ y_{1i} = x_i' \beta_o + \alpha_o y_{2i}^* + \varepsilon_{1i} \]  
(1)

\[ y_{2i}^* = z_i' \delta_o + \varepsilon_{2i} \]  
(2)

with

\[ \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \overset{iid}{\sim} \left( 0, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \right) \]  
(3)

\( i = 1, \ldots, n \), where \( x_i \) \((k \times 1)\) and \( z_i \) \((p \times 1)\) are exogenous regressors such that \( x_i \) is a subset of \( z_i \), \( y_{2i}^* \) is an endogenous latent regressor, \( \beta_o \) and \( \delta_o \) are \( k \times 1 \) and \( p \times 1 \) vectors of parameters respectively. We introduce the subscript “o” to denote the true values of parameters. Rather than observing \( y_{2i}^* \), we observe

\[ y_{2i} = \tau(y_{2i}^*) \]

where \( \tau(\cdot) \) is a nonlinear transformation. The setup represents a class of several different limited dependent variable models. For example, \( \tau(y_{2i}^*) \) could be \( \max(0, y_{2i}^*) \) or \( 1(y_{2i}^* > 0) \), i.e., censored regression or binary regression models, where \( 1(\cdot) \) is an indicator function. We assume \( x_i = (x_i, z_i) \) is independent of \( (\varepsilon_{1i}, \varepsilon_{2i}) \) so that

\[ E[\varepsilon_{2i}|z_i] = 0 \]

and the conditional mean restriction

\[ E[\varepsilon_{1i}|\varepsilon_{2i}, y_{2i}, x_i] = E[\varepsilon_{1i}|\varepsilon_{2i}, y_{2i}] \]  
(4)

are satisfied. We further assume

\[ E[\varepsilon_{1i}|\varepsilon_{2i}] = \rho_o \varepsilon_{2i} \]  
(5)

such that

\[ \rho_o = \frac{E(\varepsilon_{1i}\varepsilon_{2i})}{E(\varepsilon_{1i}^2)E(\varepsilon_{2i}^2)}. \]

Then we can take expectation of (1) and (2) conditional on \( y_{2i} \)

\[ E[y_{1i}|y_{2i}] = x_i' \beta_o + \alpha_o E[y_{2i}^*|y_{2i}] + E[\varepsilon_{1i}|y_{2i}] \]  
(6)

and

\[ E[y_{2i}^*|y_{2i}] = z_i' \delta_o + E[\varepsilon_{2i}|y_{2i}]. \]  
(7)

By the law of iterated expectation we get

\[ E[\varepsilon_{1i}|y_{2i}] = E[E[\varepsilon_{1i}|\varepsilon_{2i}] | y_{2i}] = \rho_o E[\varepsilon_{2i}|y_{2i}]. \]  
(8)

Denote

\[ y_{2i}^* = E[y_{2i}^*|y_{2i}] \]
and
\[ \varepsilon_{2i} = E[\varepsilon_{2i}|y_{2i}] . \]

Plugging equation (8) into equation (6) gives
\[ E[y_{1i}|y_{2i}] = x'_i \beta_o + \alpha_o y_{2i}^* + \rho_o \varepsilon_{2i} \]
or
\[ y_{1i} = x'_i \beta_o + \alpha_o y_{2i}^* + \rho_o \varepsilon_{2i} + [y_{1i} - E[y_{1i}|y_{2i}]] \]
\[ = w'_i \theta_o + u_i \quad (9) \]

where \( u_i = [y_{1i} - E[y_{1i}|y_{2i}]], w_i = (x_i, y_{2i}^*, \varepsilon_{2i})' \) and \( \theta_o = (\beta_o, \alpha_o, \rho_o)' \). In (9) we use \( \varepsilon_{2i} = E[\varepsilon_{2i}|y_{2i}] \) from the first stage regression to control for endogeneity of the regressors. This is the control function approach in the literature\(^1\). The control function approach treats endogeneity as an omitted variable problem, where the inclusion of the first stage error \( \varepsilon_{2i} = E[\varepsilon_{2i}|y_{2i}] \) as a regressor corrects the inconsistency of the second stage regression. Clearly \( y_{2i}^* \) and \( \varepsilon_{2i} \) in (9) are not observable. The idea of this paper is to substitute simulated moment estimates for \( y_{2i}^* = E[y_{2i}^*|y_{2i}] \) and \( \varepsilon_{2i} = E[\varepsilon_{2i}|y_{2i}] \) and derive an estimator (e.g., a least squares estimator) for \( \beta_o, \alpha_o, \) and \( \rho_o \). Let \( \bar{y}_{2i}^* \) and \( \bar{\varepsilon}_{2i} \) be the simulated moment estimates of \( y_{2i}^* = E[y_{2i}^*|y_{2i}] \) and \( \varepsilon_{2i} = E[\varepsilon_{2i}|y_{2i}] \), e.g., \( \bar{y}_{2i}^* \) and \( \bar{\varepsilon}_{2i} \) can be estimated by the simulation-based methods, e.g., GHK simulator (Geweke, 1991; Borsch-Supan and Hajivassiliou, 1993; Keane, 1994). The essence of simulation-based estimation is to replace the population moment by its sample analogue. We replace the expectation,

\[ \varepsilon_{2i} = E[\varepsilon_{2i}|y_{2i}] \]

by its simulated moment estimate\(^2\). For the sample observation, \( z_i \), a simulated moment (simulator) for \( \varepsilon_{2i} = E[\varepsilon_{2i}|y_{2i}] \) is
\[ \frac{1}{R} \sum_{j=1}^{R} \bar{\varepsilon}_{2i}^j \quad (10) \]

where
\[ \bar{\varepsilon}_{2i}^j = h \left( \xi_i^j, z_i, \delta_o \right) , \]
\( \xi_i^1, \ldots, \xi_i^R \) are \( R \) random draws for a random variable \( \xi \) and
\[ E \left( h \left( \xi_i^j, z_i, \delta_o \right) | x \right) = E[\varepsilon_{2i}|y_{2i}] . \]

\(^1\)Suppose \( y_{2i} = y_{2i}^* \). We compute \( \hat{\varepsilon}_{2i} = y_{2i}^* - \hat{z}_i \delta_o \), the residual from the first stage regression. Now consider including \( \hat{\varepsilon}_{2i} \) as an additional regressor in (1) and estimating by least squares. It is easy to show (e.g., Dhrymes 1970; Wooldridge 2002, p. 107-108) that the resulting least squares is the same as two stage least squares (2SLS) estimator.

\(^2\)The simulation-based approach in this paper complements the generalized residual approach in Vella (1993, 1998). However, our simulation-based approach has the advantage that it can be used to the models that \( \varepsilon_{2i} = E[\varepsilon_{2i}|y_{2i}] \) may not be easily calculated.
The superscript $j$ on $\xi$ indicates random draws are independent across different sample observations as in Lee (1995, 1999). However, $\varepsilon_{2i}^j = h\left( \xi_i^j, z_i, \delta_o \right)$ depends on the unknown parameter $\delta_o$. We then replace $\varepsilon_{2i} = h\left( \xi_i^j, z_i, \delta_o \right)$ by $\varepsilon_{2i}^j = h\left( \xi_i^j, z_i, \hat{\delta} \right)$ so that the simulated moments used in this paper is defined as

$$\bar{\varepsilon}_{2i} = \frac{1}{R} \sum_{j=1}^{R} \varepsilon_{2i}^j$$

(11)

where $\hat{\delta}$ is a consistent estimator of $\delta_o$. We define

$$\bar{y}_{2i}^* = z_i \hat{\delta} + \bar{\varepsilon}_{2i}$$

(12)

where $\hat{\delta}$ is a $\sqrt{n}$ consistent estimator. A class of simulators has been introduced by McFadden (1989), Stern (1992), Borsch-Supan and Hajivassiliou (1993), Hajivassiliou et al. (1996), and many others. When $R$ goes to infinity as $n$ goes to infinity, $\bar{\varepsilon}_{2i} = \frac{1}{R} \sum_{j=1}^{R} \varepsilon_{2i}^j$ will be a consistent estimator of $\varepsilon_{2i} = E[\varepsilon_{2i}|y_{2i}]$.

Then we replace $y_{2i}$ and $\varepsilon_{2i}$ by $\bar{y}_{2i}^*$ and $\bar{\varepsilon}_{2i}$ to get

$$y_{1i} = x'_i \beta_o + \alpha_o y_{2i}^* + \rho_o \varepsilon_{2i} + u_i$$

$$= x'_i \beta_o + \alpha_o \bar{y}_{2i}^* + \rho_o \bar{\varepsilon}_{2i} + \alpha_o (y_{2i}^* - \bar{y}_{2i}^*) + \rho_o (\varepsilon_{2i} - \bar{\varepsilon}_{2i}) + u_i$$

$$= x'_i \beta_o + \alpha_o \bar{y}_{2i}^* + \rho_o \bar{\varepsilon}_{2i} + \alpha_o (\bar{\delta} - \delta_o) + \alpha_o (\bar{\varepsilon}_{2i} - \varepsilon_{2i}) + \rho_o (\varepsilon_{2i} - \bar{\varepsilon}_{2i}) + u_i$$

$$= x'_i \beta_o + \alpha_o \bar{y}_{2i}^* + \rho_o \bar{\varepsilon}_{2i} + \alpha_o z_i \left( \bar{\delta} - \delta_o \right) + (\alpha_o - \rho_o)(\bar{\varepsilon}_{2i} - \varepsilon_{2i}) + u_i$$

$$= x'_i \beta_o + \alpha_o \bar{y}_{2i}^* + \rho_o \bar{\varepsilon}_{2i} + \mu_i + u_i$$

$$= x'_i \beta_o + \alpha_o \bar{y}_{2i}^* + \rho_o \bar{\varepsilon}_{2i} + v_i$$

$$= \bar{w}_i \theta_o + v_i$$

where

$$\mu_i = \alpha_o z_i \left( \bar{\delta} - \delta_o \right) + (\alpha_o - \rho_o)(\bar{\varepsilon}_{2i} - \varepsilon_{2i})$$

and

$$v_i = \mu_i + u_i.$$

Thus we estimate

$$y_{1i} = x'_i \beta_o + \alpha_o \bar{y}_{2i}^* + \rho_o \bar{\varepsilon}_{2i} + v_i$$

$$= \bar{w}_i \theta_o + v_i$$

(13)

by least squares, for example, where $\bar{w}_i = (x_i, \bar{y}_{2i}^*, \bar{\varepsilon}_{2i})$ and $\theta_o = \left( \beta'_o, \alpha_o, \rho_o \right)'$, where $v_i$ is the error term. Note (13) is a regression model with generated regressors, $\bar{y}_{2i}^*$ and $\bar{\varepsilon}_{2i}$. See Pagan (1984, 1986) for a survey on the issues of generated regressors in econometrics.
2.1 Asymptotic Properties

We now impose a set of regularity conditions:

**Assumption 1:**

1. The sample observations $x_i = (x_i, z_i)$, $i = 1, ..., n$ are i.i.d. with a compact support $X$.
2. The parameter space $\Theta$ is a compact convex subset of a $k + p$ dimensional Euclidean space and the true parameter vector $\theta_o$ is in the interior of $\Theta$.
3. The function $\varepsilon_{2i} = E[\varepsilon_{2i}|y_{2i}]$ is continuous on $\Theta \times X$.
4. The function $\varepsilon_{2i} = E[\varepsilon_{2i}|y_{2i}]$ is differentiable in $\theta$ up to the second order. Those derivatives are continuous on $\Theta \times X$.

**Assumption 2:**

1. The random draws of $\xi$ are from a common distribution and are independent of $x_i = (x_i, z_i)$ and $\theta$.
2. The function $h(\xi, x, \theta)$ is a continuous unbiased estimator of $\varepsilon_{2i} = E[\varepsilon_{2i}|y_{2i}]$ conditional on $x_i$.
3. $h(\xi, z, \delta)$ is twice differentiable in $\theta$. Those derivatives are continuous on $\Theta \times X$.
4. The absolute values of $h(\xi, z, \delta)$ and its first and second derivatives with respect to $\theta$ are dominated by square integrable function of $\xi$ uniformly in $x$ and $\theta$.
5. The first six order moments of $h(\xi, z, \delta)$ and the first four order moments of $\frac{\partial h(\xi, z, \delta)}{\partial \theta}$ exist and are bounded functions on $\Theta$.

**Assumption 3:** The number of random draws $R$ for each individual $i$ goes to infinity as $n$ goes to infinity.

**Assumption 4:** $x_i = (x_i, z_i)$ is independent of $(\varepsilon_{1i}, \varepsilon_{2i})$

**Assumption 5:** $Ew_i'w_i < \infty$, $Ew_iw_i = Q > 0$, $Ew_iw_iu_i^2 = \Omega$, $Ew_i'z_i = Q_1$.

**Assumption 6:** $\sqrt{n} \left( \hat{\delta} - \delta_o \right) \xrightarrow{d} N(0, V_1)$.

**Remarks:**

1. Assumptions 1-3 are similar the ones in Lee (1995).
3. In Assumption 6, $\delta_o$, in fact, represents all the parameters in the first stage regression, for example, $\delta_o$ may include variance or threshold parameters for limited dependent variable models. $\hat{\delta}$ could be MLE, GMM or any $\sqrt{n}$ consistent semi-parametric estimator (e.g., Powell, 1984, 1986).
Let $C_{nR} = \min \{ \sqrt{n}, \sqrt{R} \}$. The following lemma describes the asymptotic properties of the simulated residual, $\tilde{e}_{2i}$.

**Lemma 1** Under Assumptions 1-6,

(a) 
\[
\frac{1}{n} \sum_{i=1}^{n} (\tilde{e}_{2i} - \varepsilon_{2i})^2 = O_p \left( \frac{1}{C_{nR}^2} \right),
\]

(b) 
\[
\frac{1}{n} \sum_{i=1}^{n} x_i (\tilde{e}_{2i} - \varepsilon_{2i}) = O_p \left( \frac{1}{C_{nR}} \right),
\]

(c) 
\[
\frac{1}{n} \sum_{i=1}^{n} y_{2i}^* (\tilde{e}_{2i} - \varepsilon_{2i}) = O_p \left( \frac{1}{C_{nR}} \right),
\]

(d) 
\[
\frac{1}{n} \sum_{i=1}^{n} \varepsilon_{2i} (\tilde{e}_{2i} - \varepsilon_{2i}) = O_p \left( \frac{1}{C_{nR}} \right).
\]

The following lemma summarizes the asymptotic properties of the simulated latent variable, $\tilde{y}_{2i}^*$.

**Lemma 2** Under Assumptions 1-6,

(a) 
\[
\frac{1}{n} \sum_{i=1}^{n} (\tilde{y}_{2i}^* - y_{2i}^*)^2 = O_p \left( \frac{1}{C_{nR}^2} \right),
\]

(b) 
\[
\frac{1}{n} \sum_{i=1}^{n} x_i (\tilde{y}_{2i}^* - y_{2i}^*) = O_p \left( \frac{1}{C_{nR}} \right),
\]

(c) 
\[
\frac{1}{n} \sum_{i=1}^{n} y_{2i}^* (\tilde{y}_{2i}^* - y_{2i}^*) = O_p \left( \frac{1}{C_{nR}} \right),
\]

(d) 
\[
\frac{1}{n} \sum_{i=1}^{n} (\tilde{y}_{2i}^* - y_{2i}^*) \varepsilon_{2i} = O_p \left( \frac{1}{C_{nR}} \right),
\]

(e) 
\[
\frac{1}{n} \sum_{i=1}^{n} (\tilde{y}_{2i}^* - y_{2i}^*) (\tilde{e}_{2i} - \varepsilon_{2i}) = O_p \left( \frac{1}{C_{nR}^2} \right).
\]
Remarks:

1. The simulation error, $\frac{1}{n} \sum_{i=1}^{n} (\varepsilon_{2i} - \varepsilon_{2i})^2$ will generally be $O_p \left( \frac{1}{R} \right)$ if $\tilde{e}_{2i}$ does not depend on the first stage estimator, $\hat{\delta}$, as we have shown in equation (30). However, for most cases, the simulation error actually is

$$\frac{1}{n} \sum_{i=1}^{n} (\varepsilon_{2i} - \varepsilon_{2i})^2 = O_p \left( \frac{1}{R} \right) + O_p \left( \frac{1}{n} \right) = O_p \left( \frac{1}{C^2 R n R} \right)$$

since $\tilde{e}_{2i}$ does depend on $(\hat{\delta} - \delta_o)$. In a broad sense, Lemma 1(a) is similar to Theorem 1 of Bai and Ng (2002).

2. The Lemma 2(a) establishes that the sample average of the squared deviation between the simulated latent variable, $e_{y_{2i}}^*$, and true latent variable, $y_{2i}^*$, and vanishes as $(n, R) \rightarrow \infty$. The rate of convergence is determined by $C^2 n R$. Of course, Lemma 1(a) has a similar interpretation.

Our STS estimator can be written as

$$\hat{\theta} = \begin{pmatrix} \hat{\beta} \\ \hat{\alpha} \\ \hat{\rho} \end{pmatrix} = \left( \sum_{i=1}^{n} \hat{w}_i \hat{w}_i' \right)^{-1} \sum_{i=1}^{n} \hat{w}_i y_i.$$

Hence it follows that

$$\left( \hat{\theta} - \theta_o \right) = \left( \sum_{i=1}^{n} \hat{w}_i \hat{w}_i' \right)^{-1} \sum_{i=1}^{n} \hat{w}_i \left\{ u_i + \alpha_o z_i' (\hat{\delta} - \delta_o) + (\alpha_o - \rho_o) (\varepsilon_{2i} - \varepsilon_{2i}) \right\}$$

$$= \left( \sum_{i=1}^{n} \hat{w}_i \hat{w}_i' \right)^{-1} \sum_{i=1}^{n} \hat{w}_i u_i$$

$$+ \left( \sum_{i=1}^{n} \hat{w}_i \hat{w}_i' \right)^{-1} \sum_{i=1}^{n} \hat{w}_i \alpha_o z_i' (\hat{\delta} - \delta_o)$$

$$+ \left( \sum_{i=1}^{n} \hat{w}_i \hat{w}_i' \right)^{-1} \sum_{i=1}^{n} \hat{w}_i (\alpha_o - \rho_o) (\varepsilon_{2i} - \varepsilon_{2i}).$$

Lemma 3 Under Assumptions 1-6, we have

(a)

$$\frac{1}{n} \sum_{i=1}^{n} \hat{w}_i \hat{w}_i' = \frac{1}{n} \sum_{i=1}^{n} w w_i' + O_p \left( \frac{1}{C n R} \right).$$
(b) \[ \frac{1}{n} \sum_{i=1}^{n} ||\hat{w}_i - w_i||^2 = O_p \left( \frac{1}{C^2 n R} \right) \]

(c) \[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{w}_i u_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i u_i + O_p \left( \frac{\sqrt{n}}{C n R} \right) = O_p(1) + O_p \left( \frac{\sqrt{n}}{C n R} \right) , \]

(d) \[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{w}_i \alpha_o z_i \left( \hat{\delta} - \delta_o \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i \alpha_o z_i \left( \hat{\delta} - \delta_o \right) + O_p \left( \frac{1}{C n R} \right) = O_p(1) + O_p \left( \frac{1}{C n R} \right) , \]

(e) \[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{w}_i (\alpha_o - \rho_o) (\bar{\varepsilon}_{2i} - \varepsilon_{2i}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i (\alpha_o - \rho_o) (\bar{\varepsilon}_{2i} - \varepsilon_{2i}) + O_p \left( \frac{\sqrt{n}}{C^2 n R} \right) \]

\[ = O_p \left( \frac{\sqrt{n}}{C n R} \right) + O_p \left( \frac{\sqrt{n}}{C^2 n R} \right) . \]

**Theorem 1:** Under Assumptions 1-6 and \( \frac{n}{R} \to 0 \) as \( (n, R) \to \infty \), we have:

\[ \sqrt{n} \left( \hat{\theta} - \theta_o \right) = Q^{-1} \left\{ S_n + L_n + Q_n \right\} \]

where

\[ S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i u_i , \]

\[ L_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i \alpha_o z_i \left( \hat{\delta} - \delta_o \right) , \]

and

\[ Q_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i (\alpha_o - \rho_o) (\bar{\varepsilon}_{2i} - \varepsilon_{2i}) . \]

Furthermore, \( S_n = O_p(1) \), \( L_n = O_p(1) \), and \( Q_n = O_p \left( \frac{\sqrt{n}}{C n R} \right) = O_p(1) \) since \( R > n \).

Remarks:
1. The term $L_n$ involves the estimation error, $\left( \hat{\delta} - \delta_o \right)$, from the first stage regression. The term $Q_n$ involves the errors of the simulated moment, $\bar{\epsilon}_{2i} - \epsilon_{2i}$. We see from Theorem 1 that the bias due to the simulation error may dominate the rest of terms unless $R$ increases faster than the sample size $n$. This is also observed by Lee (1995) though in a different context.

2. Note that from Theorem 1 that asymptotic normality may not hold. However, asymptotic normality may hold if $\delta_o$ is known before we simulate the simulated moment. Let’s explain this point in details. Suppose that the simulated moment $\bar{\epsilon}_{2i}$ does not depend on $\alpha \delta$. $Q_n$ is $O_p(\sqrt{n R})$ as shown in (30). It implies that $Q_n = O_p\left( \frac{\sqrt{n}}{\sqrt{R}} \right) = o_p(1)$ as $\frac{\sqrt{n}}{R} \rightarrow 0$ if $\bar{\epsilon}_{2i}$ does not depend on $\hat{\delta}$. Therefore the limiting distribution of $\sqrt{n} \left( \hat{\theta} - \theta_o \right)$ is determined by $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i u_i$ and $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i \alpha_o \bar{z}_i \left( \hat{\delta} - \delta_o \right)$. Hence it follows that

$$\sqrt{n} \left( \hat{\theta} - \theta_o \right) = Q^{-1} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i u_i + \frac{1}{n} \sum_{i=1}^{n} w_i \alpha_o \bar{z}_i \sqrt{n} \left( \hat{\delta} - \delta_o \right) \right] + o_p(1)$$

Also

$$\left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i u_i \right] \xrightarrow{d} N \left( 0, \left( \begin{array}{cc} \Omega & Q_2 \\ Q_2 & V_1 \end{array} \right) \right)$$

by a central limit theorem and Assumption 5 where

$$\text{Cov} \left[ \sqrt{n} \left( \hat{\delta} - \delta_o \right), \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i u_i \right] = Q_2.$$ 

Hence,

$$\sqrt{n} \left( \hat{\theta} - \theta_o \right) \xrightarrow{d} N \left( 0, V_2 \right)$$

where

$$V_2 = Q^{-1} \left[ \Omega + \alpha_o^2 Q_1 V_1 Q_1' + Q_1 V_1 Q_2' + Q_2 V_1 Q_2' \right] Q^{-1}$$

and

$$\Sigma = \left[ \Omega + \alpha_o^2 Q_1 V_1 Q_1' + Q_1 V_1 Q_2' + Q_2 V_1 Q_2' \right].$$

3. Now $R$ increases at a rate slower than $n$, i.e.,

$$\lim_{n \rightarrow \infty} \frac{n}{R} = \infty.$$
It follows that
\[
\sqrt{R} (\hat{\theta} - \theta_o) = Q^{-1} \left[ \frac{1}{O_p} \left( \frac{1}{\sqrt{n}} \right) \right] + o_p(1)
\]
\[
= Q^{-1} \left[ O_p \left( \frac{\sqrt{R}}{C_{nR}} \right) + O_p \left( \frac{\sqrt{R}}{\sqrt{n}} \right) + O_p \left( \frac{\sqrt{R}}{C_{nR}} \right) \right] + o_p(1)
\]
\[
= Q^{-1} \left[ O_p (1) + O_p \left( \frac{\sqrt{R}}{\sqrt{n}} \right) + O_p (1) \right] + o_p(1)
\]
\[
= Q^{-1} \left[ O_p (1) + O_p \left( \frac{\sqrt{R}}{\sqrt{n}} \right) + O_p (1) \right] + o_p(1)
\]
\[
= O_p (1)
\]
since
\[
\lim \frac{R}{n} \to 0.
\]

4. When \( R \) increases slower than \( n \), the limiting distribution of
\[
\sqrt{n} (\hat{\theta} - \theta_o) = \sqrt{\frac{n}{R}} \sqrt{R} (\hat{\theta} - \theta_o)
\]
diverges. Only when \( R \) increases faster than \( n \) is the limiting distribution of \( \sqrt{n} (\hat{\theta} - \theta_o) \) properly behaved.

5. The iid assumption for \( (\varepsilon_{1i}, \varepsilon_{2i}) \) seems to be restrictive. In fact, the results of Lemmas 1-3 and Theorem 1 still hold for the heterokedastic error terms. If the homoskedastic error terms hold, then
\[
\Omega = E w_i w_i' u_i^2 = \sigma^2_i Q.
\]

6. Note \( y_{2i}^* \) is endogenous in (1) if and only if \( E (\varepsilon_{1i}\varepsilon_{2i}) \neq 0 \). We could use the results of Theorem 1 to test
\[
H_0 : E (\varepsilon_{1i}\varepsilon_{2i}) = 0.
\]
Therefore testing \( H_0 : E (\varepsilon_{1i}\varepsilon_{2i}) = 0 \) is equivalent to testing
\[
H_0 : \rho_o = 0
\]
in (9) by a \( t \)-statistic if \( R \) increases faster than \( n \).
3 Limited Dependent Variable Models

In this section we extend our results to the situation in which the second stage regression is a LDV model. The LDV model with latent endogenous regressor has been discussed extensively in the literature, e.g., Heckman (1978), Amemiya (1978, 1979), Lee (1978, 1979), Nelson and Olson (1978, 1979), Newey (1987), and Vella (1993, 1998). In this section, we develop a STS estimator for the LDV model where there is a latent endogenous regressor. The proposed STS estimator is easily implemented and provides a test of exogeneity. We consider

$$y^*_1 = x'_i \beta_o + \alpha_o y^*_2 + \varepsilon_{1i}$$  \hspace{1cm} (14)

$$y^*_2 = z'_i \delta_o + \varepsilon_{2i}$$  \hspace{1cm} (15)

where $y^*_1$ and $y^*_2$ are both latent variables with

$$\begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix} \sim iid \begin{pmatrix} 0, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \end{pmatrix}.$$  \hspace{1cm} (16)

Rather than observing $y^*_1$ and $y^*_2$, we observe

$$y_{1i} = \tau_1(y^*_{1i})$$

and

$$y_{2i} = \tau_2(y^*_{2i})$$

respectively. The setup includes several different limited dependent variable models. For example, $\tau_1(y^*_{1i})$ could be $\max(0, y^*_{1i})$ and $\tau_2(y^*_{2i})$ could be $1(y^*_{2i} > 0)$, then the model is a system of censored regression and binary regression models.

Note $1(\cdot)$ is an indicator function.

Note that under the assumption that

$$E[\varepsilon_{1i}|y_{2i}] = \rho_o \varepsilon_{2i}$$  \hspace{1cm} (17)

as in (5) we get

$$E[\varepsilon_{1i}|y_{2i}] = \rho_o E[\varepsilon_{2i}|y_{2i}]$$  \hspace{1cm} (18)

by the law of iterated expectation. Again we take the expectation of (14) and (15) conditional on $y_{2i}$ to get

$$E[y^*_{1i}|y_{2i}] = x'_i \beta_o + \alpha_o E[y^*_{2i}|y_{2i}] + E[\varepsilon_{1i}|y_{2i}]$$  \hspace{1cm} (19)

Plugging equation (18) into equation (19) gives

$$E[y^*_{1i}|y_{2i}] = x'_i \beta_o + \alpha_o E[y^*_{2i}|y_{2i}] + \rho_o E[\varepsilon_{2i}|y_{2i}]$$

or

$$y^*_{1i} = x'_i \beta_o + \alpha_o E[y^*_{2i}|y_{2i}] + \rho_o E[\varepsilon_{2i}|y_{2i}] + u_i$$

$$= x'_i \beta_o + \alpha_o y^*_{2i} + \rho_o \varepsilon_{2i} + u_i$$

$$= w'_i \theta_o + u_i$$  \hspace{1cm} (20)
Assumption 7:

Suppose that the STS estimator, GMM, could estimate \( g \) when the MLE may not be easily obtained computationally) since \( v_i \) is the sum of \( u_i \) and and as in (13). Note that the distribution of \( y \) where \( y = \frac{y_i - E[y_i]}{\sigma^2_i} \) and \( \sigma^2_i \) is the unknown variance. The second equality follows by expanding \( \alpha_o, \alpha_o, \) and \( \rho_o \) by a MLE if we knew the distribution of \( u_i \) or a GMM. With \( y_i \) and \( \bar{y}_i \), unobservable, we can use \( \bar{y}_i \) to approximate \( y_i \) and \( \bar{y}_i \) where \( \bar{y}_i \) and \( \bar{y}_i \) are given in (10) and (12). Thus, we have

\[
y_i = \hat{w}_i \theta_o + v_i
\]

as in (13). Note that the distribution of \( v_i \) may be difficult to obtain \(^3\) (hence the MLE may not be easily obtained computationally) since \( v_i \) is the sum of \( u_i \) and

\[
\mu_i = \alpha_o z_i (\hat{\delta} - \delta_o) + (\alpha_o - \rho_o) (\bar{y}_i - \bar{y}_2).
\]

Suppose that the STS estimator, \( \hat{\theta} \), solves the following equation (e.g., by a GMM)

\[
\frac{1}{n} \sum_{i=1}^{n} g(\hat{w}_i, \theta) = 0
\]

(21)

where \( g \) is a vector of functions with the same dimension as \( \theta_o \). We also assume that the first stage estimator \( \hat{\delta} \) is \( \sqrt{n} \) consistent. Expanding the left-hand side of (21) around \( \theta_o \) and solving gives

\[
\sqrt{n} \left( \hat{\theta} - \theta_o \right) = - \left[ \frac{1}{n} \sum_{i=1}^{n} \nabla_\theta g(\hat{w}_i, \theta) \right]^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(\hat{w}_i, \theta_o)
\]

\[
= - \left[ \frac{1}{n} \sum_{i=1}^{n} \nabla_\theta g(\hat{w}_i, \theta) \right]^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(\hat{w}_i, \theta_o) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_w g(\hat{w}_i, \theta_o) (\hat{w}_i - w_i) \right\}
\]

(22)

where \( \bar{\theta} \) and \( \bar{w}_i \) are mean values. The second equality follows by expanding \( g(\hat{w}_i, \theta_o) \) around \( w_i \). We need the following assumptions.

**Assumption 7:**

1. \( \frac{1}{n} \sum_{i=1}^{n} \nabla_w g(\hat{w}_i, \theta_o) \nabla_w g(\hat{w}_i, \theta_o)^T \overset{p}{\to} G_w, \)

2. \( \frac{1}{n} \sum_{i=1}^{n} \nabla_\theta g(\hat{w}_i, \theta_o) \overset{p}{\to} Q, \)

3. \( \frac{1}{n} \sum_{i=1}^{n} \nabla_{\theta w} g(\hat{w}_i, \theta_o) \overset{p}{\to} G_{\theta w}, \)

4. \( \frac{1}{n} \sum_{i=1}^{n} \nabla_{\theta y} g(\hat{w}_i, \theta_o) \overset{p}{\to} Q_1, \)

\(^3\)The generalized residual approach of Vella (1993, 1998) also has this difficulty of computing MLE.
(5) \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(w_i, \theta_o) \xrightarrow{P} N(0, \Omega), \) where
\[
\Omega = E \left[ g(w_i, \theta_o) g(w_i, \theta_o)' \right].
\]

**Theorem 2:** Under Assumptions 1-7, and \( \frac{n}{R} \to 0 \) as \((n, R) \to \infty\), we have:
\[
\sqrt{n} \left( \hat{\theta} - \theta_o \right) = Q^{-1} \{ S_n + L_n + Q_{1n} + Q_{2n} \}
\]
where
\[
S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(w_i, \theta_o),
\]
\[
L_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_y g(w_i, \theta_o) \zeta_i \left( \hat{\delta} - \delta \right),
\]
\[
Q_{1n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_y g(w_i, \theta_o) (\bar{z}_{2i} - \varepsilon_{2i}),
\]
and
\[
Q_{2n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{\varepsilon} g(w_i, \theta_o) (\bar{z}_{2i} - \varepsilon_{2i}).
\]
Furthermore, \( S_n = O_p(1), L_n = O_p(1), Q_{1n} = O_p \left( \frac{\sqrt{n}}{\sqrt{R}} \right), Q_{2n} = O_p \left( \frac{\sqrt{n}}{\sqrt{R}} \right) = O_p(1) \) since \( R > n \).

**Remarks:**

1. Suppose that the simulated moment does not depend on the first stage estimation. It follows that
\[
\sqrt{n} \left( \hat{\theta} - \theta_o \right) = -Q^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(w_i, \theta_o) + \frac{1}{n} \sum_{i=1}^{n} \nabla_y g(w_i, \theta_o) \zeta_i \sqrt{n} \left( \hat{\delta} - \delta \right) \right\} + o_p(1)
\]
d\( \to N(0, \mathbf{V}_2) \)

since
\[
\left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(w_i, \theta_o), \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(w_i, \theta_o) \right] \xrightarrow{d} N \left( 0, \begin{pmatrix} \Omega & Q_2 \\ Q_2' & V_1 \end{pmatrix} \right)
\]
where
\[
Cov \left[ \sqrt{n} \left( \hat{\delta} - \delta \right), \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(w_i, \theta_o) \right] = Q_2,
\]
\[
\mathbf{V}_2 = Q^{-1} \left[ \Omega + Q_1 V_1 Q_1' + Q_1 V_1 Q_2' + Q_2 V_1 Q_1' \right] Q^{-1},
\]
and
\[
\Sigma = \Omega + Q_1 V_1 Q_1' + Q_1 V_1 Q_2' + Q_2 V_1 Q_1'.
\]
4 Dummy Endogenous Regressor

The econometrics models with dummy endogenous regressor commonly arise in, e.g., in the program evaluation literature, dummy endogenous regressor captures the causal relationship between a binary regressor (say treatment status) and an outcome variable. The dummy endogenous regressor in essence is to allow for the possibility of joint determination of outcomes and treatment status or omitted variables related to both treatment status and outcomes. For example, Heckman (1978) imposes the joint normality assumption and develops the MLE for the model. Because the computation of the MLE could be nontrivial, one may want to use two step approach similar in Section 3. This procedure does not produce consistent estimator as Wooldridge (2002, p. 478) points it out, however. Blundell and Powell (2004a) considers a semiparametric estimation in a single index binary response model with continuous regressor. Note that their control function approach can not be used in, for example, a binary model with a dummy endogenous regressor. In this section, we propose a STS estimator to control for possible endogeneity bias as in Sections 3-4.

Let \( y_{1i} \) be the outcome variable of interest and \( y_{2i} \) be the dummy endogenous regressor. We consider

\[
y_{1i}^* = x_{0i}' \beta_o + \alpha_o y_{2i} + \varepsilon_{1i} \tag{23}
\]

\[
y_{2i}^* = z_{i}' \delta_o + \varepsilon_{2i} \tag{24}
\]

where

\[
y_{1i} = \tau_1 (y_{1i}^*)
\]

and

\[
y_{2i} = 1 \text{ if } y_{2i}^* > 0; \quad y_{2i} = 1 \text{ otherwise.}
\]

Again we assume

\[
E[\varepsilon_{1i}|\varepsilon_{2i}] = \rho_o \varepsilon_{2i}
\]

and

\[
E[\varepsilon_{1i}|y_{2i}] = \rho_o E[\varepsilon_{2i}|y_{2i}]
\]

such that

\[
\rho_o = \frac{E(\varepsilon_{1i}\varepsilon_{2i})}{E(\varepsilon_{1i}^2)E(\varepsilon_{2i}^2)}.
\]

We take expectation of (23) and (24) conditional on \( y_{2i} \) to get

\[
E[y_{1i}^*|y_{2i}] = x_{i}' \beta_o + \alpha_o E[y_{2i}|y_{2i}] + E[\varepsilon_{1i}|y_{2i}] \tag{25}
\]

and

\[
E[y_{2i}^*|y_{2i}] = z_{i}' \delta_o + E[\varepsilon_{2i}|y_{2i}] . \tag{26}
\]

Then rewrite (25) as

\[
y_{1i}^* = x_{i}' \beta_o + \alpha_o y_{2i} + \rho_o E[\varepsilon_{2i}|y_{2i}] + u_i
\]

\[
= x_{i}' \beta_o + \alpha_o y_{2i} + \rho_o \varepsilon_{2i} + u_i
\]

\[
= w_{i}' \theta_o + u_i \tag{27}
\]
where \( \varepsilon_{2i} = E[\varepsilon_{2i} | y_{2i}] \), \( u_i = y_{1i}^* - E[y_{1i}^* | y_{2i}] \), \( w_i = (x_i', y_{2i}, \varepsilon_{2i})' \), and \( \theta_o = (\beta_o, \alpha_o, \rho_o) \). Note here we abuse notation to use \( w_i \) here since \( w_i \) in this section is defined differently from the previous sections. While \( \varepsilon_{2i} = E[\varepsilon_{2i} | y_{2i}] \) cannot be observed, we can estimate \( \varepsilon_{2i} = E[\varepsilon_{2i} | y_{2i}] \) by \( \varepsilon_{2i} \) as in (10). Thus

\[
y_{1i}^* = x_i' \beta_o + \alpha_o y_{2i} + \rho_o (\varepsilon_{2i} - \bar{\varepsilon}_{2i}) + u_i
\]

where \( \bar{\varepsilon}_{2i} = (x_i', y_{2i}, \bar{\varepsilon}_{2i}) \) and \( \nu_i = \rho (\varepsilon_{2i} - \bar{\varepsilon}_{2i}) + u_i \). Again assume (28) can be estimated, say, by a GMM. Then the STS estimator, \( \hat{\theta} \), satisfies the following equation

\[
\frac{1}{n} \sum_{i=1}^{n} g(\bar{\varepsilon}_{2i}, \theta) = 0
\]

as in (21). We expand (29) as in (22) to get

\[
\sqrt{n} (\hat{\theta} - \theta_o) = - \left[ \frac{1}{n} \sum_{i=1}^{n} \nabla g(\bar{\varepsilon}_{2i}, \theta) \right]^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(\bar{\varepsilon}_{2i}, \theta_o) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla g(\bar{\varepsilon}_{2i}, \theta_o) (\bar{\varepsilon}_{2i} - w_i) \right\}.
\]

**Theorem 3:** Under Assumptions 1-7 and \( n R \rightarrow 0 \) as \( (n, R) \rightarrow \infty \), we have:

\[
\sqrt{n} (\hat{\theta} - \theta_o) = -Q^{-1} [S_n + Q_{2n}] + o_p(1)
\]

where \( S_n \) and \( Q_{2n} \) are defined in Theorem 2.

## 5 Conclusions

This paper introduces a STS estimation procedure for regression models with endogenous latent/discrete regressors. The procedure simulated residuals from the reduced form as an additional regressor in the outcome model to control the endogeneity. The paper makes two contributions. First, we develop the asymptotic theory and rate of convergence for the STS estimator. The STS estimator behaves badly, i.e., \( \sqrt{n} (\hat{\theta} - \theta_o) \) diverges, unless the number of simulated random variables, \( R \), goes to infinity with a rate faster than the sample size, \( n \), i.e., \( \frac{R}{n} \rightarrow 0 \) as \( (n, R) \rightarrow \infty \). Second, the proposed STS estimator allows endogenous regressors to be latent or discrete.

**Appendix**
A  Proof of Lemma 1

Proof. The proof of part (a) is similar to the Proposition A.3 in Lee (1995). Recall \( \bar{\varepsilon}_{2i} = \frac{1}{R} \sum_{j=1}^{R} \varepsilon_{2i}^{j} \). Let

\[
\bar{\varepsilon}_{2i} = \frac{1}{R} \sum_{j=1}^{R} \varepsilon_{2i}^{j} (\delta_o)
\]

and

\[
q_i = \left( \varepsilon_{2i}^{j} - \varepsilon_{2i} \right)^{2}
\]

\[
= \left( \frac{1}{R} \sum_{j=1}^{R} \left( \varepsilon_{2i}^{j} (\delta_o) - \varepsilon_{2i} \right) \right)^{2}.
\]

Lemma A in Serfling (1980, p. 304) implies that

\[
E(q_i)^2 = E \left( \frac{1}{R} \sum_{j=1}^{R} \left( \varepsilon_{2i}^{j} (\delta_o) - \varepsilon_{2i} \right) \right)^4
\]

\[
\leq \frac{c}{R^4} \sum_{j=1}^{R} E \left( \varepsilon_{2i}^{j} (\delta_o) - \varepsilon_{2i} \right)^4
\]

\[
= \frac{c}{R^2} \sum_{j=1}^{R} E \left( \varepsilon_{2i}^{j} (\delta_o) - \varepsilon_{2i} \right)^4
\]

\[
= \frac{c}{R} E \left( \varepsilon_{2i}^{j} (\delta_o) - \varepsilon_{2i} \right)^4.
\]

where \( c \) is a constant. By the Markov inequality and the inequality of absolute moments,

\[
P \left( \frac{R}{n} \sum_{i=1}^{n} |q_i| \geq \epsilon \right)
\leq \frac{R}{\epsilon} E \left[ |q_i| \right]
\leq \frac{R}{\epsilon} E^{1/2} \left[ |q_i|^2 \right]
\leq \frac{1}{\epsilon} c^{1/2} E \left( \varepsilon_{2i}^{j} (\delta_o) - \varepsilon_{2i} \right)^4^{1/2}
= \frac{1}{\epsilon} O_p \left( 1 \right).
\]

since

\[
E \left( \varepsilon_{2i}^{j} (\delta_o) - \varepsilon_{2i} \right)^4 = O \left( 1 \right)
\]
from Assumption 2. Because $\epsilon$ is arbitrary,

$$\frac{R}{n} \sum_{i=1}^{n} |q_i| = O_p(1).$$

It follows that

$$\frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{R} \sum_{j=1}^{R} (\delta_{2i}(\delta_0) - \delta_{2i}) \right)^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left( \bar{\delta}_{2i} - \delta_{2i} \right)^2$$

$$= O_p \left( \frac{1}{R} \right). \quad (30)$$

Now use $C_r'$ inequality to get

$$\frac{1}{n} \sum_{i=1}^{n} \left( \bar{\delta}_{2i} - \delta_{2i} \right)^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left( \bar{\delta}_{2i} - \delta_{2i} + \bar{\delta}_{2i} - \bar{\delta}_{2i} \right)^2$$

$$\leq 2 \left[ \frac{1}{n} \sum_{i=1}^{n} \left( \bar{\delta}_{2i} - \delta_{2i} \right)^2 + \frac{1}{n} \sum_{i=1}^{n} \left( \bar{\delta}_{2i} - \bar{\delta}_{2i} \right)^2 \right].$$

By a mean value theorem

$$\bar{\delta}_{2i} - \delta_{2i} = \nabla \bar{\delta}_{2i} \left( \delta - \delta_0 \right)$$

with $\nabla$ being the gradient operator and $\delta$ lies between $\bar{\delta}$ and $\delta_0$. Then

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \left( \bar{\delta}_{2i} - \delta_{2i} \right)^2 \right\| = \left\| \frac{1}{n} \sum_{i=1}^{n} \left[ \nabla \bar{\delta}_{2i} \left( \delta - \delta_0 \right) \right]^2 \right\|$$

$$\leq \left\| \delta - \delta_0 \right\|^2 \left( \frac{1}{n} \sum_{i=1}^{n} \left\| \nabla \bar{\delta}_{2i} \left( \delta \right) \right\|^4 \right)^{1/2}$$

$$= O_p \left( \frac{1}{n} \right) = O_p(1)$$

Here we use the results that

$$\hat{\theta} - \theta_0 = O_p \left( \frac{1}{\sqrt{n}} \right)$$
the consistency $\delta$ and
\[
\frac{1}{n} \sum_{i=1}^{n} \left\| \nabla \tilde{e}_{2i}(\tilde{\delta}) \right\|^4 = O_p(1).
\]
It follows that
\[
\frac{1}{n} \sum_{i=1}^{n} (\tilde{e}_{2i} - e_{2i})^2
= O_p\left(\frac{1}{R}\right) + O_p\left(\frac{1}{n}\right)
= O_p\left(\frac{1}{C_nR}\right)
\]
This proves part (a). Consider (b). By the Cauchy-Schwarz inequality
\[
\left\| \frac{1}{n} \sum_{i=1}^{n} x_i (\tilde{e}_{2i} - e_{2i}) \right\|
\leq \left( \frac{1}{n} \sum_{i=1}^{n} \| x_i \|^2 \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} (\tilde{e}_{2i} - e_{2i})^2 \right)^{1/2}
= O_p(1) O_p\left(\frac{1}{C_nR}\right)
= O_p\left(\frac{1}{C_nR}\right)
\]
since
\[
\frac{1}{n} \sum_{i=1}^{n} \| x_i \|^2 = O_p(1).
\]
Consider (c).
\[
\left\| \frac{1}{n} \sum_{i=1}^{n} y_{2i}^* (\tilde{e}_{2i} - e_{2i}) \right\|
\leq \left( \frac{1}{n} \sum_{i=1}^{n} \| y_{2i}^* \|^2 \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} (\tilde{e}_{2i} - e_{2i})^2 \right)^{1/2}
= O_p(1) O_p\left(\frac{1}{C_nR}\right)
= O_p\left(\frac{1}{C_nR}\right)
\]
since
\[
\frac{1}{n} \sum_{i=1}^{n} \| y_{2i}^* \|^2 = O_p(1).
\]
Consider (d).

\[
\left\| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{2i} (\bar{e}_{2i} - \varepsilon_{2i}) \right\| \leq \left( \frac{1}{n} \sum_{i=1}^{n} \| \varepsilon_{2i} \| \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} (\bar{e}_{2i} - \varepsilon_{2i})^2 \right)^{1/2}
\]

\[
= O_p (1) O_p \left( \frac{1}{\sqrt{n R}} \right) = O_p \left( \frac{1}{\sqrt{n R}} \right)
\]
since

\[
\frac{1}{n} \sum_{i=1}^{n} \| \varepsilon_{2i} \|^2 = O_p (1).
\]

\[\square\]

**B Proof of Lemma 2**

**Proof.** Note

\[
\bar{y}_{2i}^* - y_{2i}^* = z_i' \left( \hat{\delta} - \delta_o \right) + (\bar{e}_{2i} - \varepsilon_{2i}).
\]

By the \(C^*\) inequality

\[
\left\| \bar{y}_{2i}^* - y_{2i}^* \right\|^2 = \left\| z_i' \left( \hat{\delta} - \delta_o \right) + (\bar{e}_{2i} - \varepsilon_{2i}) \right\|^2
\]

\[
\leq 2 \left( \left\| z_i' \left( \hat{\delta} - \delta_o \right) \right\|^2 + \| \bar{e}_{2i} - \varepsilon_{2i} \|^2 \right)
\]

\[
= 2 (a_i + b_i)
\]

where

\[a_i = \left\| z_i' \left( \hat{\delta} - \delta_o \right) \right\|^2\]

and

\[b_i = \| \bar{e}_{2i} - \varepsilon_{2i} \|^2 .\]

It follows that

\[
\frac{1}{n} \sum_{i=1}^{n} \left\| \bar{y}_{2i}^* - y_{2i}^* \right\|^2 \leq 2 \left[ \frac{1}{n} \sum_{i=1}^{n} (a_i + b_i) \right].
\]

Now

\[
\left\| z_i' \left( \hat{\delta} - \delta_o \right) \right\|^2 \leq \|z_i\|^2 \left\| \left( \hat{\delta} - \delta_o \right) \right\|^2 .
\]

\[20\]
Thus
\[
\frac{1}{n} \sum_{i=1}^{n} \left\| z_i (\hat{\delta} - \delta_o) \right\|^2 \leq \left\| (\hat{\delta} - \delta_o) \right\|^2 \frac{1}{n} \sum_{i=1}^{n} \| z_i \|^2
\]
\[
= \left[ O_p \left( \frac{1}{\sqrt{n}} \right) \right]^2 O_p (1)
\]
\[
= O_p \left( \frac{1}{n} \right)
\]
because
\[
(\hat{\delta} - \delta_o) = O_p \left( \frac{1}{\sqrt{n}} \right)
\]
and
\[
\frac{1}{n} \sum_{i=1}^{n} \| z_i \|^2 = O_p (1)
\]
by Assumptions 4 and 5. For (b), we have that
\[
\frac{1}{n} \sum_{i=1}^{n} b_i = \frac{1}{n} \sum_{i=1}^{n} (\tilde{e}_{2i} - \varepsilon_{2i})^2
\]
\[
= O_p \left( \frac{1}{C_{nR}} \right)
\]
from Lemma 1. Combining these results, we have
\[
\frac{1}{n} \sum_{i=1}^{n} (a_i + b_i) = O_p \left( \frac{1}{n} \right) + O_p \left( \frac{1}{C_{nR}} \right)
\]
\[
= O_p \left( \frac{1}{C_{nR}} \right)
\]
This proves part (a). By the Cauchy-Schwarz inequality:
\[
\left\| \frac{1}{n} \sum_{i=1}^{n} x_i (\tilde{y}_{2i} - y_{2i}) \right\| \leq \left( \frac{1}{n} \sum_{i=1}^{n} \| x_i \|^2 \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} (\tilde{y}_{2i} - y_{2i})^2 \right)^{1/2}
\]
which is
\[
O_p (1) O_p \left( \frac{1}{C_{nR}} \right) = O_p \left( \frac{1}{C_{nR}} \right)
\]
by part (a) and Assumption 4. This proves part (b). Consider part (c). By the Cauchy-Schwarz inequality
\[
\left\| \frac{1}{n} \sum_{i=1}^{n} y_{2i}^* (\tilde{y}_{2i} - y_{2i}) \right\| \leq \left( \frac{1}{n} \sum_{i=1}^{n} \| y_{2i}^* \|^2 \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} (\tilde{y}_{2i} - y_{2i})^2 \right)^{1/2}
\]
\[
= O_p (1) O_p \left( \frac{1}{C_{nR}} \right) = O_p \left( \frac{1}{C_{nR}} \right)
\]
Consider (d).

\[
\left\| \frac{1}{n} \sum_{i=1}^{n} (\tilde{y}_{2i} - y_{2i}^*) \varepsilon_{2i} \right\| \leq \left( \frac{1}{n} \sum_{i=1}^{n} \| (\tilde{y}_{2i} - y_{2i}^*) \| ^2 \right) ^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{2i}^2 \right) ^{1/2} = O_p \left( \frac{1}{C_nR} \right) O_p (1) = O_p \left( \frac{1}{C_nR} \right).
\]

Consider (e).

\[
\left\| \frac{1}{n} \sum_{i=1}^{n} (\tilde{y}_{2i}^* - y_{2i}^*) (\tilde{\varepsilon}_{2i} - \varepsilon_{2i}) \right\| \leq \left( \frac{1}{n} \sum_{i=1}^{n} \| (\tilde{y}_{2i}^* - y_{2i}^*) \| ^2 \right) ^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} (\tilde{\varepsilon}_{2i} - \varepsilon_{2i})^2 \right) ^{1/2} = O_p \left( \frac{1}{C_nR} \right) O_p \left( \frac{1}{C_nR} \right) = O_p \left( \frac{1}{C_nR} \right).
\]

\[\blacksquare\]

C Proof of Lemma 3

Proof. Note

\[
\sum_{i=1}^{n} \hat{w}_i \hat{w}_i' = \sum_{i=1}^{n} \begin{bmatrix} x_i x_i' & x_i \tilde{y}_{2i} & x_i \tilde{\varepsilon}_{2i} \\ \tilde{y}_{2i} x_i' & \tilde{y}_{2i} \tilde{y}_{2i}' & \tilde{y}_{2i} \tilde{\varepsilon}_{2i} \\ \tilde{\varepsilon}_{2i} x_i' & \tilde{\varepsilon}_{2i} \tilde{y}_{2i}' & \tilde{\varepsilon}_{2i} \tilde{\varepsilon}_{2i}' \end{bmatrix}.
\]

Hence using the results of Lemmas 1 and 2 we have

\[
\frac{1}{n} \sum_{i=1}^{n} x_i \tilde{y}_{2i} = \frac{1}{n} \sum_{i=1}^{n} [x_i \tilde{y}_{2i} + x_i (\tilde{y}_{2i} - y_{2i})] = \frac{1}{n} \sum_{i=1}^{n} x_i \tilde{y}_{2i} + \frac{1}{n} \sum_{i=1}^{n} x_i (\tilde{y}_{2i} - y_{2i}) = \frac{1}{n} \sum_{i=1}^{n} x_i \tilde{y}_{2i} + O_p \left( \frac{1}{C_nR} \right),
\]

\[
\frac{1}{n} \sum_{i=1}^{n} x_i \tilde{\varepsilon}_{2i} = \frac{1}{n} \sum_{i=1}^{n} x_i \varepsilon_{2i} + \frac{1}{n} \sum_{i=1}^{n} x_i (\tilde{\varepsilon}_{2i} - \varepsilon_{2i}) = \frac{1}{n} \sum_{i=1}^{n} x_i \varepsilon_{2i} + O_p \left( \frac{1}{C_nR} \right),
\]

22
\[ \frac{1}{n} \sum_{i=1}^{n} \bar{y}_{2i}^2 = \frac{1}{n} \sum_{i=1}^{n} (y_{2i}^* + \bar{y}_{2i}^2 - y_{2i})^2 \]
\[ = \frac{1}{n} \sum_{i=1}^{n} y_{2i}^2 - 2 \frac{1}{n} \sum_{i=1}^{n} y_{2i} (\bar{y}_{2i} - y_{2i}) + \frac{1}{n} \sum_{i=1}^{n} (\bar{y}_{2i} - y_{2i})^2 \]
\[ = \frac{1}{n} \sum_{i=1}^{n} y_{2i}^2 + O_p \left( \frac{1}{CnR} \right) + O_p \left( \frac{1}{C^2nR} \right) \]
\[ = \frac{1}{n} \sum_{i=1}^{n} y_{2i}^2 + O_p \left( \frac{1}{CnR} \right) , \]

\[ \frac{1}{n} \sum_{i=1}^{n} \bar{y}_{2i} \bar{\varepsilon}_{2i} = \frac{1}{n} \sum_{i=1}^{n} (y_{2i}^* + \bar{y}_{2i} - y_{2i}) (\varepsilon_{2i} + \bar{\varepsilon}_{2i} - \varepsilon_{2i}) \]
\[ = \frac{1}{n} \sum_{i=1}^{n} y_{2i} \varepsilon_{2i} + \frac{1}{n} \sum_{i=1}^{n} y_{2i}^* (\bar{\varepsilon}_{2i} - \varepsilon_{2i}) \]
\[ + \frac{1}{n} \sum_{i=1}^{n} (\bar{y}_{2i} - y_{2i}) \varepsilon_{2i} + \frac{1}{n} \sum_{i=1}^{n} (\bar{y}_{2i} - y_{2i}^*) (\bar{\varepsilon}_{2i} - \varepsilon_{2i}) \]
\[ = \frac{1}{n} \sum_{i=1}^{n} y_{2i} \varepsilon_{2i} + O_p \left( \frac{1}{CnR} \right) + O_p \left( \frac{1}{CnR} \right) + O_p \left( \frac{1}{C^2nR} \right) \]
\[ = \frac{1}{n} \sum_{i=1}^{n} y_{2i} \varepsilon_{2i} + O_p \left( \frac{1}{CnR} \right) , \]

and

\[ \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{2i}^2 = \frac{1}{n} \sum_{i=1}^{n} (\varepsilon_{2i} + \bar{\varepsilon}_{2i} - \varepsilon_{2i})^2 \]
\[ = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{2i}^2 + 2 \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{2i} (\bar{\varepsilon}_{2i} - \varepsilon_{2i}) + \frac{1}{n} \sum_{i=1}^{n} (\bar{\varepsilon}_{2i} - \varepsilon_{2i})^2 \]
\[ = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{2i}^2 + O_p \left( \frac{1}{CnR} \right) + O_p \left( \frac{1}{C^2nR} \right) \]
\[ = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_{2i}^2 + O_p \left( \frac{1}{CnR} \right) . \]

Thus
\[ \frac{1}{n} \sum_{i=1}^{n} \hat{w}_i \hat{w}_i' = \frac{1}{n} \sum_{i=1}^{n} w w_i' + O_p \left( \frac{1}{CnR} \right) . \]

This proves (a). Consider (b). It is easy to see that
\[ \hat{w}_i - w_i = \begin{pmatrix} \frac{x_i - \bar{x}}{\bar{y}_{2i} - \bar{y}_{2i}^2} \\ \frac{\varepsilon_{2i} - \bar{\varepsilon}_{2i}}{\bar{y}_{2i} - \bar{y}_{2i}^2} \end{pmatrix} \]

23
and
\[
\frac{1}{n} \sum_{i=1}^{n} \| \hat{w}_i - w_i \|^2 = \frac{1}{n} \sum_{i=1}^{n} (\hat{w}_{2i} - y_{2i})^2 + \frac{1}{n} \sum_{i=1}^{n} (\hat{w}_{2i} - \hat{\varepsilon}_{2i})^2
\]
\[
= O_p \left( \frac{1}{C^2 nR} \right) + O_p \left( \frac{1}{\varepsilon^2 nR} \right)
\]
\[
= O_p \left( \frac{1}{C^2 nR} \right).
\]

Consider (c). Note
\[
\frac{1}{n} \sum_{i=1}^{n} \hat{w}_i u_i = \frac{1}{n} \sum_{i=1}^{n} w_i u_i + \frac{1}{n} \sum_{i=1}^{n} (\hat{w}_i - w_i) u_i.
\]
For the second term, given that the two variables \( \hat{w}_i - w_i \) and \( u_i \) are uncorrelated, one can use the Cauchy-Schwarz inequality to get a sharper bound by applying the correction factor \( O_p \left( T^{-1/2} \right) \) as shown in Trapani (2004). Therefore
\[
\left\| \frac{1}{n} \sum_{i=1}^{n} (\hat{w}_i - w_i) u_i \right\| \leq \sqrt{\frac{1}{n} \sum_{i=1}^{n} \| \hat{w}_i - w_i \|^2} \sqrt{\frac{1}{n} \sum_{i=1}^{n} u_i^2 O_p \left( \frac{1}{\sqrt{n}} \right)}
\]
\[
= O_p \left( \frac{1}{C^2 nR} \right) O_p (1) O_p \left( \frac{1}{\sqrt{n}} \right)
\]
\[
= O_p \left( \frac{1}{C^2 nR\sqrt{n}} \right)
\]
by part (b) and
\[
\frac{1}{n} \sum_{i=1}^{n} u_i^2 = O_p (1).
\]

Assume the law of large number holds here so that
\[
\frac{1}{n} \sum_{i=1}^{n} x_i y_i \overset{p}{\to} E [xy]
\]
which is \( O_p \) if \( x_i \) and \( y_i \) are uncorrelated and \( O_p \) if \( x_i \) and \( y_i \) are correlated. Hence
\[
\left\| \frac{1}{n} \sum_{i=1}^{n} x_i y_i \right\| \leq \left( \frac{1}{n} \sum_{i=1}^{n} \| x_i \|^2 \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} \| y_i \|^2 \right)^{1/2}
\]
\[
= O_p (1) O_p (1)
\]
if we assume \( \frac{1}{n} \sum_{i=1}^{n} \| x_i \|^2 < \infty \) and \( \frac{1}{n} \sum_{i=1}^{n} \| y_i \|^2 < \infty \). However we know \( \frac{1}{n} \sum_{i=1}^{n} x_i y_i = O_p \) only when \( E [x_i y_i] \neq 0 \). This means a correction term should be added in order to get a sharper bound when \( E [x_i y_i] = 0 \) and the correction term is \( O_p \left( \frac{1}{\sqrt{n}} \right) \) as shown in Trapani (2004).
Hence,
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{w}_i u_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i u_i + O_p \left( \frac{1}{\sqrt{n}} \right) + O_p \left( 1 \right)
\]
\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i u_i + O_p \left( \frac{1}{\sqrt{n}} \right). 
\]

Consider (d).
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{w}_i \alpha_o z_i' \left( \hat{\delta} - \delta_o \right) = \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} (w_i + \hat{w}_i - w_i) \alpha_o z_i' \left( \hat{\delta} - \delta_o \right) \right)
\]
\[
= \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} w_i \alpha_o z_i' \left( \hat{\delta} - \delta_o \right) + \frac{1}{n} \sum_{i=1}^{n} (\hat{w}_i - w_i) \alpha_o z_i' \left( \hat{\delta} - \delta_o \right) \right)
\]
\[
= \sqrt{n} \left( I + II \right). 
\]

We consider each term in turn.

\[
||I|| \leq \|\alpha_o\| \|\hat{\delta} - \delta_o\| \sqrt{\frac{1}{n} \sum_{i=1}^{n} \|w_i z_i'\|^2}
\]
\[
= O_p \left( \frac{1}{\sqrt{n}} \right) O_p \left( 1 \right)
\]
\[
= O_p \left( \frac{1}{\sqrt{n}} \right) 
\]

since
\[
\frac{1}{n} \sum_{i=1}^{n} \|w_i z_i'\|^2 = O_p \left( 1 \right). 
\]

\[
||II|| \leq \|\alpha_o\| \|\hat{\delta} - \delta_o\| \sqrt{\frac{1}{n} \sum_{i=1}^{n} \|\hat{w}_i - w_i\|^2} \sqrt{\frac{1}{n} \sum_{i=1}^{n} \|z_i\|^2}
\]
\[
= O_p \left( \frac{1}{\sqrt{n}} \right) O_p \left( \frac{1}{C_{nR}} \right) O_p \left( 1 \right)
\]
\[
= O_p \left( \frac{1}{C_{nR} \sqrt{n}} \right). 
\]

since
\[
\hat{\delta} - \delta_o = O_p \left( \frac{1}{\sqrt{n}} \right) 
\]
and

\[
\frac{1}{n} \sum_{i=1}^{n} \| z_i \|^2 = O_p(1)
\]

by Assumptions 4 and 5. Hence

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{w}_i \alpha_0 \hat{z}'_i (\bar{\delta} - \delta_o) = \sqrt{n} \left( O_p \left( \frac{1}{\sqrt{n}} \right) + O_p \left( \frac{1}{C_{nR} \sqrt{n}} \right) \right) = O_p(1) + O_p \left( \frac{1}{C_{nR}} \right).
\]

Consider (e). Note

\[
\frac{1}{n} \sum_{i=1}^{n} \hat{w}_i (\alpha_o - \rho_o) (\bar{\varepsilon}_{2i} - \varepsilon_{2i}) = \frac{1}{n} \sum_{i=1}^{n} (w_i + \hat{w}_i - w_i) (\alpha_o - \rho_o) (\bar{\varepsilon}_{2i} - \varepsilon_{2i}) = \frac{1}{n} \sum_{i=1}^{n} w_i (\alpha_o - \rho_o) (\bar{\varepsilon}_{2i} - \varepsilon_{2i}) + \frac{1}{n} \sum_{i=1}^{n} (\hat{w}_i - w_i) (\alpha_o - \rho_o) (\bar{\varepsilon}_{2i} - \varepsilon_{2i})
\]

Now

\[
\left\| \frac{1}{n} \sum_{i=1}^{n} w_i (\alpha_o - \rho_o) (\bar{\varepsilon}_{2i} - \varepsilon_{2i}) \right\| \leq \| \alpha_o - \rho_o \| \left( \frac{1}{n} \sum_{i=1}^{n} \| w_i \|^2 \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} (\bar{\varepsilon}_{2i} - \varepsilon_{2i})^2 \right)^{1/2} = \| \alpha_o - \rho_o \| \times I \times II.
\]

We begin with I. Note

\[
\frac{1}{n} \sum_{i=1}^{n} \| w_i \|^2 = O_p(1)
\]

and

\[
\frac{1}{n} \sum_{i=1}^{n} (\bar{\varepsilon}_{2i} - \varepsilon_{2i})^2 = O_p \left( \frac{1}{C_{nR}^2} \right).
\]

Next

\[
\left\| \frac{1}{n} \sum_{i=1}^{n} (\hat{w}_i - w_i) (\alpha_o - \rho_o) (\bar{\varepsilon}_{2i} - \varepsilon_{2i}) \right\| \leq \| \alpha_o - \rho_o \| \left( \frac{1}{n} \sum_{i=1}^{n} \| \hat{w}_i - w_i \|^2 \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} \| \bar{\varepsilon}_{2i} - \varepsilon_{2i} \|^2 \right)^{1/2} = O_p \left( \frac{1}{C_{nR}^2} \right)
\]

26
since
\[
\frac{1}{n} \sum_{i=1}^{n} \| \hat{w}_i - w_i \|^2 = O_p \left( \frac{1}{C_n R} \right)
\]
and
\[
\frac{1}{n} \sum_{i=1}^{n} \| \bar{\varepsilon}_i - \varepsilon_i \|^2 = O_p \left( \frac{1}{C_n R} \right)
\]

Hence
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{w}_i \left( \alpha_o - \rho_o \right) (\bar{\varepsilon}_i - \varepsilon_i) = \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} w_i \left( \alpha_o - \rho_o \right) (\bar{\varepsilon}_i - \varepsilon_i) + O_p \left( \frac{1}{C_n R} \right) \right)
\]
\[
= O_p \left( \sqrt{n} \frac{1}{C_n R} \right) + O_p \left( \frac{\sqrt{n}}{C_n R} \right).
\]

\[\boxed{\text{D Proof of Theorem 1}}\]

\[\text{Proof.} \text{ Lemma 4 implies that }\]
\[
\sqrt{n} \left( \hat{\theta} - \theta_o \right) = \left( \frac{1}{n} \sum_{i=1}^{n} \hat{w}_i w_i' \right)^{-1} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{w}_i u_i + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{w}_i \alpha_o z_i' \left( \hat{\delta} - \delta_o \right) \right] + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i (\alpha_o - \rho_o) (\bar{\varepsilon}_2 - \varepsilon_2) + O_p \left( \frac{1}{C_n R} \right)
\]
\[
= \left[ \frac{1}{n} \sum_{i=1}^{n} w_i w_i' + O_p \left( \frac{1}{C_n R} \right) \right]^{-1} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i u_i + O_p \left( \frac{1}{C_n R} \right) \right] + O_p \left( \frac{n}{C_n R} \right) + O_p \left( \frac{\sqrt{n}}{C_n R} \right) + O_p \left( \frac{n}{C_n R^2} \right)
\]
\[
= \left[ \frac{1}{n} \sum_{i=1}^{n} w_i w_i' + O_p \left( \frac{1}{C_n R} \right) \right]^{-1} \left\{ O_p \left( \frac{1}{C_n R} \right) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i (\alpha_o - \rho_o) (\bar{\varepsilon}_2 - \varepsilon_2) \right\} + O_p \left( \frac{\sqrt{n}}{C_n R} \right) + O_p \left( \frac{n}{C_n R^2} \right).
\]
E Proof of Theorem 2

Proof. Note by a Taylor expansion

\[ \sqrt{n} (\hat{\theta} - \theta_o) \]

\[ = - \left[ \frac{1}{n} \sum_{i=1}^{n} \nabla g (\hat{\theta}, w_i) \right]^{-1} \]

\[ \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g (w_i, \theta_o) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_w g (\bar{w}, \theta_o) (\hat{w}_i - w_i) \right\}. \]

By the Cauchy-Schwarz inequality

\[ \left\| \frac{1}{n} \sum_{i=1}^{n} \nabla_w g (\bar{w}, \theta_o) (\hat{w}_i - w_i) \right\| \]

\[ \leq \left( \frac{1}{n} \sum_{i=1}^{n} \| \nabla_w g (\bar{w}, \theta_o) \|^2 \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} \| \hat{w}_i - w_i \|^2 \right)^{1/2} \]

\[ = O_p (1) O_p \left( \frac{1}{C_n R} \right) \]

\[ = O_p \left( \frac{1}{C_n R} \right) \]

since

\[ \frac{1}{n} \sum_{i=1}^{n} \| \nabla_w g (\bar{w}, \theta_o) \|^2 = O_p (1) \]

by Assumption 6 and

\[ \frac{1}{n} \sum_{i=1}^{n} \| \hat{w}_i - w_i \|^2 = O_p \left( \frac{1}{C_n^2 R} \right) \]

from Lemma 3 (b). Expanding \( \nabla g (\hat{w}, \theta_o) \) around \( w \) gives

\[ \frac{1}{n} \sum_{i=1}^{n} \nabla g (\hat{w}, \theta_o) = \frac{1}{n} \sum_{i=1}^{n} \nabla g (w_i, \theta_o) + \frac{1}{n} \sum_{i=1}^{n} \nabla_w g (\bar{w}, \theta_o) (\hat{w}_i - w_i) \]

\[ = I + II. \]

Consider II.

\[ \left\| \frac{1}{n} \sum_{i=1}^{n} \nabla_{\theta w} g (\bar{w}, \theta_o) (\hat{w}_i - w_i) \right\| \]

\[ \leq \left( \frac{1}{n} \sum_{i=1}^{n} \| \nabla_{\theta w} g (\bar{w}, \theta_o) \|^2 \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} \| \hat{w}_i - w_i \|^2 \right)^{1/2} \]

\[ = O_p (1) O_p \left( \frac{1}{C_n R} \right) \]

\[ = O_p \left( \frac{1}{C_n R} \right) \]

\[ = o_p (1) \]
since
\[
\frac{1}{n} \sum_{i=1}^{n} \| \nabla_{\theta Y_i} (w_i, \theta) \|^2 = O_p(1)
\]

by Assumption 6 and the consistency of \( \overline{m} \). Because \( \overline{\theta} \) is a consistent estimator of \( \theta_0 \), it follows that

\[
\frac{1}{n} \sum_{i=1}^{n} \nabla_{\theta Y_i} (\overline{w}_i, \overline{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \nabla_{\theta Y_i} (w_i, \theta_0) + o_p(1)
\]

\[
= Q + o_p(1).
\]

It follows that

\[
\sqrt{n}(\hat{\theta} - \theta_0) = -Q^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(w_i, \theta_0) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{w Y_i} (\overline{w}_i, \theta_0) (\overline{w}_i - w_i) \right\}
\]

\[
= -Q^{-1} \left\{ O_p(1) + \sqrt{n} O_p \left( \frac{1}{C_R} \right) \right\} + o_p(1).
\]

This is because \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(w_i, \theta_0) = O_p(1) \) by a central limit theorem. Now suppose \( \frac{2}{n} \to 0 \) as \( (n, R) \to \infty \), we have

\[
\sqrt{n}(\hat{\theta} - \theta_0) = -Q^{-1} \left\{ O_p(1) + O_p \left( \frac{\sqrt{n}}{\sqrt{n}} \right) \right\} + o_p(1)
\]

\[
= -Q^{-1} \{ O_p(1) + O_p(1) \} + o_p(1).
\]

Hence the limiting distribution of \( \sqrt{n}(\hat{\theta} - \theta_0) \) is determined by \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(w_i, \theta_0) \) and \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{w Y_i} (\overline{w}_i, \theta_0) (\overline{w}_i - w_i) \).

Note

\[
\nabla_{w Y_i} (\overline{w}_i, \theta_0) (\overline{w}_i - w_i) = \nabla_{x_i} (\overline{w}_i, \theta_0) (x_i - x_i) + \nabla_{y_2} (\overline{w}_i, \theta_0) (y_{2i} - y_{2i})
\]

\[
+ \nabla_{\varepsilon_2 Y_i} (\overline{w}_i, \theta_0) (\varepsilon_{2i} - \varepsilon_{2i})
\]

\[
= \nabla_{y_2} (\overline{w}_i, \theta_0) (y_{2i} - y_{2i}) + \nabla_{\varepsilon_2 Y_i} (\overline{w}_i, \theta_0) (\varepsilon_{2i} - \varepsilon_{2i}).
\]

By the triangle inequality we have

\[
\left\| \frac{1}{n} \sum_{i=1}^{n} \nabla_{w Y_i} (\overline{w}_i, \theta_0) (\overline{w}_i - w_i) \right\| \leq \left\| \frac{1}{n} \sum_{i=1}^{n} \nabla_{y_2} (\overline{w}_i, \theta_0) (y_{2i} - y_{2i}) \right\|
\]

\[
+ \left\| \frac{1}{n} \sum_{i=1}^{n} \nabla_{\varepsilon_2 Y_i} (\overline{w}_i, \theta_0) (\varepsilon_{2i} - \varepsilon_{2i}) \right\| = I + II.
\]

Consider I.

\[
\left\| \frac{1}{n} \sum_{i=1}^{n} \nabla_{y_2} (\overline{w}_i, \theta_0) (y_{2i} - y_{2i}) \right\| \leq \left( \frac{1}{n} \sum_{i=1}^{n} \| \nabla_{y_2} (\overline{w}_i, \theta_0) \|^2 \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} \| \overline{y}_{2i} - y_{2i} \|^2 \right)^{1/2}
\]

29
which is bounded by $O_p(1)O_p\left(\frac{1}{\sqrt{nR}}\right) = O_p\left(\frac{1}{\sqrt{nR}}\right)$ using Assumption 6 and Lemma 2. Next we consider II.

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \nabla_{\varepsilon^2} \left( w_i, \theta_o \right) \left( \bar{\varepsilon}_{2i} - \varepsilon_{2i} \right) \right\| \leq \left( \frac{1}{n} \sum_{i=1}^{n} \| \nabla_{y^2} \left( w_i, \theta_o \right) \|^2 \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} \| \bar{\varepsilon}_{2i} - \varepsilon_{2i} \|^2 \right)^{1/2}$$

$$= O_p(1)O_p\left(\frac{1}{C_nR}\right)$$

$$= O_p\left(\frac{1}{C_nR}\right)$$

since

$$\frac{1}{n} \sum_{i=1}^{n} \| \nabla_{y^2} \left( w_i, \theta_o \right) \|^2 = O_p(1)$$

and

$$\frac{1}{n} \sum_{i=1}^{n} \| \bar{\varepsilon}_{2i} - \varepsilon_{2i} \|^2 = O_p\left(\frac{1}{C_nR}\right).$$

Combining I and II, we get

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_\omega \left( w_i, \theta_o \right) \left( \bar{\omega}_i - w_i \right)$$

$$= O(\sqrt{n})O_p\left(\frac{1}{C_nR}\right) + O(\sqrt{n})O_p\left(\frac{1}{C_nR}\right)$$

$$= O(\sqrt{n})O_p\left(\frac{1}{\sqrt{n}}\right) + O(\sqrt{n})O_p\left(\frac{1}{\sqrt{n}}\right)$$

$$= O_p(1) + O_p(1)$$

$$= O_p(1)$$
if \( n \frac{R}{n} \to 0 \) as \((n, R) \to \infty\). Furthermore,

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla g (w_i, \theta_o) (\tilde{y}_{2i}^{*} - y_{2i}^{*}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla g (w_i, \theta_o) (\tilde{y}_{2i}^{*} - y_{2i}^{*}) \\
= \frac{1}{n} \sum_{i=1}^{n} \nabla g (w_i, \theta_o) \left[ z_i' (\tilde{\delta} - \hat{\delta}) + (\tilde{\varepsilon}_{2i} - \varepsilon_{2i}) \right] \\
= \frac{1}{n} \sum_{i=1}^{n} \nabla g (w_i, \theta_o) z_i' \sqrt{n} (\tilde{\delta} - \hat{\delta}) \\
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla g (w_i, \theta_o) (\tilde{\varepsilon}_{2i} - \varepsilon_{2i}) \\
= Op(1) + O \left( \sqrt{n} \right) Op \left( \frac{1}{\sqrt{nR}} \right) \\
= Op(1) + O \left( \sqrt{n} \right) \frac{1}{\sqrt{nR}} \\
= Op(1) + Op(1) \\
= Op(1)
\]

if \( n \frac{R}{n} \to 0 \). Hence

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla g (w_i, \theta_o) (\tilde{y}_{2i}^{*} - y_{2i}^{*}) \\
= \frac{1}{n} \sum_{i=1}^{n} \nabla g (w_i, \theta_o) \left[ z_i' (\tilde{\delta} - \hat{\delta}) + (\tilde{\varepsilon}_{2i} - \varepsilon_{2i}) \right] \\
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla g (w_i, \theta_o) (\tilde{\varepsilon}_{2i} - \varepsilon_{2i}) + op(1) \\
= Q_1 \sqrt{n} (\tilde{\delta} - \hat{\delta}) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla g (w_i, \theta_o) (\tilde{\varepsilon}_{2i} - \varepsilon_{2i}) + op(1).
\]

It follows that

\[
\sqrt{n} (\hat{\theta} - \theta_o) \\
= - Q^{-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g (w_i, \theta_o) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla g (w_i, \theta_o) z_i' (\tilde{\delta} - \hat{\delta}) \right\} + op(1).
\]

This proves Theorem 2.
F Proof of Theorem 3

Proof. Note

\[
\left\| \frac{1}{n} \sum_{i=1}^{n} \nabla_w g (\bar{w}_i, \theta_0) (\hat{w}_i - w_i) \right\| \leq \left( \frac{1}{n} \sum_{i=1}^{n} \| \nabla_w g (\bar{w}_i, \theta_0) \|^2 \right)^{1/2} \left( \frac{1}{n} \sum_{i=1}^{n} \| \hat{w}_i - w_i \|^2 \right)^{1/2}
\]

and

\[
\frac{1}{n} \sum_{i=1}^{n} \| \hat{w}_i - w_i \|^2 = \frac{1}{n} \sum_{i=1}^{n} (\bar{\varepsilon}_{2i} - \varepsilon_{2i})^2 = O_p \left( \frac{1}{C_{nR}} \right).
\]

Hence

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_w g (\bar{w}_i, \theta_0) (\hat{w}_i - w_i) = O_p \left( \frac{\sqrt{n}}{C_{nR}} \right).
\]

It is easy to see that

\[
\sqrt{n} \left( \hat{\theta} - \theta_0 \right) = -Q^{-1} \left\{ O_p (1) + O_p \left( \frac{\sqrt{n}}{C_{nR}} \right) \right\} + o_p (1).
\]

Suppose if \( \frac{\varepsilon}{nR} \to 0 \) as \((n, R) \to \infty\) then we get

\[
\sqrt{n} \left( \hat{\theta} - \theta_0 \right) = -Q^{-1} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g (w_i, \theta_0) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_w g (\bar{w}_i, \theta_0) (\hat{w}_i - w_i) \right] + o_p (1)
\]

\[
= -Q^{-1} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g (w_i, \theta_0) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla \varepsilon \varepsilon g (w_i, \theta_0) (\bar{\varepsilon}_{2i} - \varepsilon_{2i}) \right] + o_p (1)
\]

This proves Theorem 3.

References


