2005

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

ISSN: 1525-3066

Center for Policy Research
Working Paper No. 76

SIMULATION-BASED TWO-STEP ESTIMATION WITH ENDOGENOUS REGRESSORS

Kamhon Kan and Chihwa Kao*

Center for Policy Research
Maxwell School of Citizenship and Public Affairs
Syracuse University
426 Eggers Hall
Syracuse, New York 13244-1020
(315) 443-3114 | Fax (315) 443-1081
e-mail: ctrpol@syr.edu

December 2005

$5.00

CENTER FOR POLICY RESEARCH – Fall 2005

Timothy Smeeding, Director
Professor of Economics & Public Administration

Associate Directors

Margaret Austin
Associate Director,
Budget and Administration

Douglas Holtz-Eakin
Professor of Economics
Associate Director, Center for Policy Research

Douglas Wolf
Professor of Public Administration
Associate Director, Aging Studies Program

John Yinger
Professor of Economics and Public Administration
Associate Director, Metropolitan Studies Program

SENIOR RESEARCH ASSOCIATES

Badi Baltagi .... Economics
Pablo Beramendi .... Political Science
Dan Black .... Economics
Lloyd Blanchard .... Public Administration
William Duncombe .... Public Administration
Gary Engelhardt .... Economics
Deborah Freund .... Public Administration
Madonna Harrington Meyer .... Sociology
Christine Himes .... Sociology
William C. Horrace .... Economics
Bernard Jump .... Public Administration
Duke Kao .... Economics
Eric Kingson .... Social Work
Thomas Kniesner .... Economics
Tom Krebs .... Economics
Jeff Kubik .... Economics
Andrew London .... Sociology
Len Lopoo .... Public Administration
Jerry Miner .... Economics
John Moran .... Economics
Jan Ondrich .... Economics
John Palmer .... Public Administration
Lori Ploutz-Snyder .... Health and Physical Education
Grant Reeher .... Political Science
Stuart Rosenthal .... Economics
Ross Rubenstein .... Public Administration
Margaret Usdansky .... Sociology
Michael Wasylenko .... Economics
Janet Wilmoth .... Sociology

GRADUATE ASSOCIATES

Javier Baez .... Economics
Sonali Ballal .... Public Administration
Dana Balter .... Public Administration
Jason Bladen .... Economics
Jesse Bricker .... Economics
Maria Brown .... Social Science
Yong Chen .... Economics
Ginnie Cronin .... Social Science
Ana Dammert .... Economics
Mike Eriksen .... Economics
Katie Fitzpatrick .... Economics
Alexandre Genest .... Public Administration
Julie Anna Golebiewski .... Economics
Nadia Greenhalgh-Stanley .... Economics
Yue Hu .... Economics
Becky Lafrancois .... Economics
Joseph Marchand .... Economics
Larry Miller .... Public Administration
Wendy Parker .... Sociology
Emily Pas .... Economics
Shawn Rohlin .... Economics
Cynthia Searcy .... Public Administration
Talisha Searcy .... Public Administration
Jeff Thompson .... Economics
Wen Wang .... Economics
Yoshikun Yamamoto .... Public Administration

STAFF

Kelly Bogart .... Administrative Secretary
Martha Bonney .... Publications/Events Coordinator
Karen Cimilluca .... Librarian/Office Coordinator
Kim Desmond .... Administrative Secretary
Kati Foley .... Administrative Assistant, LIS
Kitty Nasto .... Administrative Secretary
Candi Patterson .... Computer Consultant
Mary Santy .... Administrative Secretary

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
cdkao@maxwell.syr.edu

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

Abstract

This paper considers models with latent/discrete endogenous regressors and presents a simulation-based two-step (STS) estimator. The endogeneity is corrected by adopting a simulation-based control function approach. The first step consists of simulating of the residuals of the reduced-form equation for endogenous regressors. The second step is a regression model (linear, latent or discrete) with the simulated residual as an additional regressor. In this paper we develop the asymptotic theory for the STS estimator and its rate of convergence.

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 first step of our STS estimator consists of the construction of simulation-based residuals from reduced form equation for the latent/discrete endogenous regressors. The second step is a regression model with generated regressors and the simulated residual as an additional regressor. The proposed STS estimator requires no choices of kernels or bandwidths as in nonparametric/semiparametric estimators, e.g., Das (2005), Christofides et al. (2002), Darolles et al. (2003), Blundell and Powell (2004a, 2004b), Yildiz (2004), Vytlacil and Yildiz (2004). It resembles the control function estimators that are popular in the literature, e.g., Smith and Blundell (1986), Vella (1993, 1998), Rivers and Vuong (1988), Blundell and Smith (1994), Newey et al. (1999), Li and Wooldridge (2002), Blundell and Powell (2004a, 2004b), and Lee (2004), Ma and Koenker (2004).

Regression models with latent/discrete regressors have been studied extensively in the literature, e.g., Amemiya (1978), Heckman (1978), Nelson and Olson (1978), Hsiao (1983), Maddala (1983), Newey (1985), Hsiao and Mountain (1985), Amemiya (1985, Chapter 10), Terza (1987, 1998), Kao and Wu (1990), Vella (1993, 1998), Lee (1994), M-J. Lee (1995), Vella and Verbeek (1999), Angrist (2001), Li and Wooldridge (2002), Wooldridge (2002, 15.7.3), Lewbel (2004), Vytlacil and Yildiz (2004), Yildiz (2004), and Das (2005), to mention only a few. For empirical applications with latent/discrete regressors, see Willis and Rosen (1979), Lee (1978), Evans and Schwab (1995), Evans et al. (1999), Goldman et al. (2001), and many others. This paper contributes to the literature by studying the asymptotic theory of the STS estimator with endogenous latent/discrete regressors. We derive the rate of convergence and the limiting distribution of the STS estimator.

This paper also builds on a growing literature on simulation-based methods, e.g., McFadden (1989), Pakes and Pollard (1989), Duffie and Singleton (1993), Lee (1995, 1997, 1999a, 1999b), Breslaw and McIntosh (1998), Carrasco and Florens (2002), Zhang and Lee (2004).

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:

$$y_{1i} = x_i'\beta_o + \alpha_o \mathbf{y}_{2i}^* + \varepsilon_{1i} \tag{1}$$

$$\mathbf{y}_{2i}^* = z_i^{'}\delta_o + \boldsymbol{\varepsilon}_{2i} \tag{2}$$

with

$$\left[\begin{array}{c} \varepsilon_{1i} \\ \boldsymbol{\varepsilon}_{2i} \end{array}\right] \overset{iid}{\sim} \left(0, \left[\begin{array}{cc} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{array}\right]\right) \tag{3}$$

$i = 1, ..., n$, 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, $\mathbf{y}_{2i}^*$ is an endogenous latent regressor, β_o and δ_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 $\mathbf{y}_{2i}^*$, we observe

$$y_{2i} = \tau\left(\mathbf{y}_{2i}^*\right)$$

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

$$E\left[\boldsymbol{\varepsilon}_{2i} | z_i\right] = 0$$

and the conditional mean restriction

$$E\left[\varepsilon_{1i} | \boldsymbol{\varepsilon}_{2i}, y_{2i}, \mathbf{x}_i\right] = E\left[\varepsilon_{1i} | \boldsymbol{\varepsilon}_{2i}, y_{2i}\right] \tag{4}$$

are satisfied. We further assume

$$E\left[\varepsilon_{1i} | \boldsymbol{\varepsilon}_{2i}\right] = \rho_o \boldsymbol{\varepsilon}_{2i} \tag{5}$$

such that

$$\rho_o = \frac{E\left(\varepsilon_{1i}\boldsymbol{\varepsilon}_{2i}\right)}{E\left(\varepsilon_{1i}^2\right) E\left(\boldsymbol{\varepsilon}_{2i}^2\right)}.$$

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

$$E\left[y_{1i} | y_{2i}\right] = x_i'\beta_o + \alpha_o E\left[\mathbf{y}_{2i}^* | y_{2i}\right] + E\left[\varepsilon_{1i} | y_{2i}\right] \tag{6}$$

and

$$E\left[\mathbf{y}_{2i}^* | y_{2i}\right] = z_i^{'}\delta_o + E\left[\boldsymbol{\varepsilon}_{2i} | y_{2i}\right]. \tag{7}$$

By the law of iterated expectation we get

$$E\left[\varepsilon_{1i} | y_{2i}\right] = E\left[E\left[\varepsilon_{1i} | \boldsymbol{\varepsilon}_{2i}\right] | y_{2i}\right] = \rho_o E\left[\boldsymbol{\varepsilon}_{2i} | y_{2i}\right]. \tag{8}$$

Denote

$$y_{2i}^* = E\left[\mathbf{y}_{2i}^* | y_{2i}\right]$$

and

$$\varepsilon_{2i} = E\left[\boldsymbol{\varepsilon}_{2i}|y_{2i}\right].$$

Plugging equation (8) into equation (6) gives

$$E\left[y_{1i}|y_{2i}\right] = x_i'\beta_o + \alpha_o y_{2i}^* + \rho_o \varepsilon_{2i}$$

or

$$\begin{aligned} y_{1i} &= x_i'\beta_o + \alpha_o y_{2i}^* + \rho_o \varepsilon_{2i} + \left[y_{1i} - E\left[y_{1i}|y_{2i}\right]\right] \\ &= w_i^{'}\theta_o + u_i \end{aligned} \tag{9}$$

where $u_i = \left[y_{1i} - E\left[y_{1i}|y_{2i}\right]\right]$, $w_i = \left(x_i, y_{2i}^*, \varepsilon_{2i}\right)^{'}$ and $\theta_o = \left(\beta_o^{'}, \alpha_o, \rho_o\right)^{'}$. In (9) we use $\varepsilon_{2i} = E\left[\boldsymbol{\varepsilon}_{2i}|y_{2i}\right]$ 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 $\varepsilon_{2i} = E\left[\boldsymbol{\varepsilon}_{2i}|y_{2i}\right]$ as a regressor corrects the inconsistency of the second stage regression. Clearly y_{2i}^* and ε_{2i} in (9) are not observable. The idea of this paper is to substitute simulated moment estimates for $y_{2i}^* = E\left[\mathbf{y}_{2i}^*|y_{2i}\right]$ and $\varepsilon_{2i} = E\left[\boldsymbol{\varepsilon}_{2i}|y_{2i}\right]$ and derive an estimator (e.g., a least squares estimator) for β_o, α_o, and ρ_o. Let $\widetilde{y}_{2i}^*$ and $\widetilde{\varepsilon}_{2i}$ be the simulated moment estimates of $y_{2i}^* = E\left[\mathbf{y}_{2i}^*|y_{2i}\right]$ and $\varepsilon_{2i} = E\left[\boldsymbol{\varepsilon}_{2i}|y_{2i}\right]$, e.g., $\widetilde{y}_{2i}^*$ and $\widetilde{\varepsilon}_{2i}$ 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,

$$\varepsilon_{2i} = E\left[\boldsymbol{\varepsilon}_{2i}|y_{2i}\right]$$

by its simulated moment estimate[2]. For the sample observation, z_i, a simulated moment (simulator) for $\varepsilon_{2i} = E\left[\boldsymbol{\varepsilon}_{2i}|y_{2i}\right]$ is

$$\frac{1}{R}\sum_{j=1}^{R}\overline{\varepsilon}_{2i}^{j} \tag{10}$$

where

$$\overline{\varepsilon}_{2i}^{j} = h\left(\xi_i^j, z_i, \delta_o\right),$$

$\xi_i^1, ..., \xi_i^R$ are R random draws for a random variable ξ and

$$E\left(h\left(\xi_i^j, z_i, \delta_o\right)|\mathbf{x}\right) = E\left[\boldsymbol{\varepsilon}_2|y_2\right].$$

[1] Suppose $y_{2i} = \mathbf{y}_{2i}^*$. We compute $\widehat{\varepsilon}_{2i} = \mathbf{y}_{2i}^* - \widehat{z}_i^{'}\delta_o$, the residual from the first stage regression. Now consider including $\widehat{\varepsilon}_{2i}$ 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 $\varepsilon_{2i} = E\left[\boldsymbol{\varepsilon}_{2i}|y_{2i}\right]$ may not be easily calculated.

The superscript j on ξ indicates random draws are independent across different sample observations as in Lee (1995, 1999). However, $\overline{\varepsilon}_{2i}^{j} = h\left(\xi_i^j, z_i, \delta_o\right)$ depends on the unknown parameter θ_o. We then replace $\overline{\varepsilon}_{2i}^{j} = h\left(\xi_i^j, z_i, \delta_o\right)$ by $\varepsilon_{2i}^{j} = h\left(\xi_i^j, z_i, \widehat{\delta}\right)$ so that the simulated moments used in this paper is defined as

$$\widetilde{\varepsilon}_{2i} = \frac{1}{R}\sum_{j=1}^{R} \varepsilon_{2i}^{j} \tag{11}$$

where $\widehat{\delta}$ is a consistent estimator of δ_o. We define

$$\widetilde{y}_{2i}^{*} = z_i^{'}\widehat{\delta} + \widetilde{\varepsilon}_{2i} \tag{12}$$

where $\widehat{\delta}$ 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, $\widetilde{\varepsilon}_{2i} = \frac{1}{R}\sum_{j=1}^{R} \varepsilon_{2i}^{j}$ will be a consistent estimator of $\varepsilon_{2i} = E\left[\boldsymbol{\varepsilon}_{2i} | y_{2i}\right]$. Then we replace y_{2i}^{*} and ε_{2i} by $\widetilde{y}_{2i}^{*}$ and $\widetilde{\varepsilon}_{2i}$ to get

$$\begin{aligned}
y_{1i} &= x_i'\beta_o + \alpha_o y_{2i}^{*} + \rho_o \varepsilon_{2i} + u_i \\
&= x_i'\beta_o + \alpha_o \widetilde{y}_{2i}^{*} + \rho_o \widetilde{\varepsilon}_{2i} + \alpha_o\left(y_{2i}^{*} - \widetilde{y}_{2i}^{*}\right) + \rho_o\left(\varepsilon_{2i} - \widetilde{\varepsilon}_{2i}\right) + u_i \\
&= x_i'\beta_o + \alpha_o \widetilde{y}_{2i}^{*} + \rho_o \widetilde{\varepsilon}_{2i} + \alpha_o z_i^{'}\left(\widehat{\delta} - \delta_o\right) + \alpha_o\left(\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\right) + \rho_o\left(\varepsilon_{2i} - \widetilde{\varepsilon}_{2i}\right) + u_i \\
&= x_i'\beta_o + \alpha_o \widetilde{y}_{2i}^{*} + \rho_o \widetilde{\varepsilon}_{2i} + \alpha_o z_i^{'}\left(\widehat{\delta} - \delta_o\right) + \left(\alpha_o - \rho_o\right)\left(\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\right) + u_i \\
&= x_i'\beta_o + \alpha_o \widetilde{y}_{2i}^{*} + \rho_o \widetilde{\varepsilon}_{2i} + \mu_i + u_i \\
&= x_i'\beta_o + \alpha_o \widetilde{y}_{2i}^{*} + \rho_o \widetilde{\varepsilon}_{2i} + v_i \\
&= \widehat{w}_i^{'}\theta_o + v_i
\end{aligned}$$

where

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

and

$$v_i = \mu_i + u_i.$$

Thus we estimate

$$\begin{aligned}
y_{1i} &= x_i'\beta_o + \alpha_o \widetilde{y}_{2i}^{*} + \rho_o \widetilde{\varepsilon}_{2i} + v_i \\
&= \widehat{w}_i^{'}\theta_o + v_i
\end{aligned} \tag{13}$$

by least squares, for example, where $\widehat{w}_i = \left(x_i, \widetilde{y}_{2i}^{*}, \widetilde{\varepsilon}_{2i}\right)^{'}$ 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, $\widetilde{y}_{2i}^{*}$ and $\widetilde{\varepsilon}_{2i}$. 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 $\mathbf{x}_i = (x_i, z_i)$, $i = 1, ..., n$ are i.i.d. with a compact support $\mathbf{X}$.
2. The parameter space Θ is a compact convex subset of a $k+p$ dimensional Euclidean space and the true parameter vector θ_o is in the interior of Θ.
3. The function $\varepsilon_{2i} = E[\boldsymbol{\varepsilon}_{2i}|y_{2i}]$ is continuous on $\Theta \times \mathbf{X}$.
4. The function $\varepsilon_{2i} = E[\boldsymbol{\varepsilon}_{2i}|y_{2i}]$ is differentiable in θ up to the second order. Those derivatives are continuous on $\Theta \times \mathbf{X}$.

Assumption 2:

1. The random draws of ξ are from a common distribution and are independent of $\mathbf{x}_i = (x_i, z_i)$ and θ.
2. The function $h(\xi, \mathbf{x}, \theta)$ is a continuous unbiased estimator of $\varepsilon_{2i} = E[\boldsymbol{\varepsilon}_{2i}|y_{2i}]$ conditional on $\mathbf{x}_i$.
3. $h(\xi, z, \delta)$ is twice differentiable in θ. Those derivatives are continuous on $\Theta \times \mathbf{X}$.
4. The absolute values of $h(\xi, z, \delta)$ and its first and second derivatives with respect to θ are dominated by square integrable function of ξ uniformly in $\mathbf{x}$ and θ.
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 Θ.

Assumption 3: The number of random draws R for each individual i goes to infinity as n goes to infinity.
Assumption 4: $\mathbf{x}_i = (x_i, z_i)$ is independent of $(\varepsilon_{1i}, \boldsymbol{\varepsilon}_{2i})$
Assumption 5: $Ew_i^{'}w_i < \infty$, $Ew_iw_i^{'} = Q > 0$, $Ew_iw_i^{'}u_i^2 = \Omega$, $Ew_iz_i^{'} = Q_1$.
Assumption 6: $\sqrt{n}\left(\widehat{\delta} - \delta_o\right) \xrightarrow{d} N(0, V_1)$.
Remarks:

1. Assumptions 1-3 are similar the ones in Lee (1995).
2. Assumption 4 is a critical independence condition commonly assumed in the literature, e.g., Yildiz (2004), Vytlacil and Yildiz (2004).
3. In Assumption 6, δ_o, in fact, represents all the parameters in the first stage regression, for example, δ_o may include variance or threshold parameters for limited dependent variable models. $\widehat{\delta}$ could be MLE, GMM or any $\sqrt{n}$ consistent semi-parametric estimator (e.g., Powell, 1984, 1986).

Let $C_{nR} = \min\left\{\sqrt{n}, \sqrt{R}\right\}$. The following lemma describes the asymptotic properties of the simulated residual, $\widetilde{\varepsilon}_{2i}$.

Lemma 1 *Under Assumptions 1-6,*

(a)

$$\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\right)^2 = O_p\left(\frac{1}{C_{nR}^2}\right),$$

(b)

$$\frac{1}{n}\sum_{i=1}^{n} x_i\left(\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\right) = O_p\left(\frac{1}{C_{nR}}\right),$$

(c)

$$\frac{1}{n}\sum_{i=1}^{n} y_{2i}^*\left(\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\right) = O_p\left(\frac{1}{C_{nR}}\right),$$

(d)

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

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

Lemma 2 *Under Assumptions 1-6,*

(a)

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

(b)

$$\frac{1}{n}\sum_{i=1}^{n} x_i\left(\widetilde{y}_{2i}^* - y_i^*\right) = O_p\left(\frac{1}{C_{nR}}\right),$$

(c)

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

(d)

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

(e)

$$\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{y}_{2i}^* - y_{2i}^*\right)\left(\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\right) = O_p\left(\frac{1}{C_{nR}^2}\right).$$

Remarks:

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

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)^2 \\
= \; & O_p\left(\frac{1}{R}\right)+O_p\left(\frac{1}{n}\right) \\
= \; & O_p\left(\frac{1}{C_{nR}^2}\right)
\end{aligned}$$

since $\widetilde{\varepsilon}_{2i}$ does depend on $\left(\widehat{\delta}-\delta_o\right)$. 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, $\widetilde{y}_{2i}^{*}$, and true latent variable, y_{2i}^{*}, and vanishes as $(n,R)\to\infty$. The rate of convergence is determined by C_{nR}^2. Of course, Lemma 1(a) has a similar interpretation.

Our STS estimator can be written as

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

Hence it follows that

$$\begin{aligned}
\left(\widehat{\theta}-\theta_o\right) &= \left(\sum_{i=1}^{n}\widehat{w}_i\widehat{w}_i^{'}\right)^{-1}\sum_{i=1}^{n}w_i\left\{u_i+\alpha_o z_i^{'}\left(\widehat{\delta}-\delta_o\right)+\left(\alpha_o-\rho_o\right)\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)\right\} \\
&= \left(\sum_{i=1}^{n}\widehat{w}_i\widehat{w}_i^{'}\right)^{-1}\sum_{i=1}^{n}\widehat{w}_i u_i \\
&\quad +\left(\sum_{i=1}^{n}\widehat{w}_i\widehat{w}_i^{'}\right)^{-1}\sum_{i=1}^{n}\widehat{w}_i\alpha z_i^{'}\left(\widehat{\delta}-\delta_o\right) \\
&\quad +\left(\sum_{i=1}^{n}\widehat{w}_i\widehat{w}_i^{'}\right)^{-1}\sum_{i=1}^{n}\widehat{w}_i\left(\alpha_o-\rho_o\right)\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right).
\end{aligned}$$

Lemma 3 *Under Assumptions 1-6, we have*

(a)

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{w}_i\widehat{w}_i^{'}=\frac{1}{n}\sum_{i=1}^{n}ww_i^{'}+O_p\left(\frac{1}{C_{nR}}\right),$$

(b)

$$\frac{1}{n}\sum_{i=1}^{n}\|\widehat{w}_i - w_i\|^2 = O_p\left(\frac{1}{C_{nR}^2}\right)$$

(c)

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{w}_i u_i &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i u_i + O_p\left(\frac{\sqrt{n}}{C_{nR}}\right)\\ &= O_p(1) + O_p\left(\frac{\sqrt{n}}{C_{nR}}\right),\end{aligned}$$

(d)

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{w}_i \alpha_o z_i^{'}\left(\widehat{\delta}-\delta_o\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i \alpha_o z_i^{'}\left(\widehat{\delta}-\delta_o\right) + O_p\left(\frac{1}{C_{nR}}\right)\\ &= O_p(1) + O_p\left(\frac{1}{C_{nR}}\right),\end{aligned}$$

(e)

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{w}_i\left(\alpha_o-\rho_o\right)\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i\left(\alpha_o-\rho_o\right)\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right) + O_p\left(\frac{\sqrt{n}}{C_{nR}^2}\right)\\ &= O_p\left(\frac{\sqrt{n}}{C_{nR}}\right) + O_p\left(\frac{\sqrt{n}}{C_{nR}^2}\right).\end{aligned}$$

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

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

where

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

$$L_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i \alpha_o z_i^{'}\left(\widehat{\delta}-\delta_o\right),$$

and

$$Q_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i\left(\alpha_o-\rho_o\right)\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right).$$

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

Remarks:

1. The term L_n involves the estimation error, $\left(\widehat{\delta}-\delta_o\right)$, from the first stage regression. The term Q_n involves the errors of the simulated moment, $\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}$. 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 δ_o is known before we simulate the simulated moment. Let's explain this point in details. Suppose that the simulated moment $\widetilde{\varepsilon}_{2i}$ does not depend on $\widehat{\delta}$. Q_n is $O_p\left(\frac{\sqrt{n}}{\sqrt{R}}\right)$ as shown in (30). It implies that $Q_n=O_p\left(\frac{\sqrt{n}}{\sqrt{R}}\right)=o_p\left(1\right)$ as $\frac{n}{R}\to 0$ if $\widetilde{\varepsilon}_{2i}$ does not depend on $\widehat{\delta}$. Therefore the limiting distribution of $\sqrt{n}\left(\widehat{\theta}-\theta_o\right)$ is determined by $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}w_iu_i$ and $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}w_i\alpha_o z_i^{'}\left(\widehat{\delta}-\delta_o\right)$. Hence it follows that

$$\sqrt{n}\left(\widehat{\theta}-\theta_o\right)=Q^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}w_iu_i+\frac{1}{n}\sum_{i=1}^{n}w_i\alpha_o z_i^{'}\sqrt{n}\left(\widehat{\delta}-\delta_o\right)\right]+o_p(1)$$

   Also

$$\left[\begin{array}{c}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}w_iu_i\\ \sqrt{n}\left(\widehat{\delta}-\delta_o\right)\end{array}\right]\xrightarrow{d}N\left(0,\left(\begin{array}{cc}\Omega & Q_2\\ Q_2^{'} & V_1\end{array}\right)\right)$$

   by a central limit theorem and Assumption 5 where

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

   Hence,

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

   where

$$\begin{aligned}V_2&=Q^{-1}\left[\Omega+\alpha_o^2Q_1V_1Q_1^{'}+Q_1V_1Q_2^{'}+Q_2V_1Q_1^{'}\right]Q^{-1}\\&=Q^{-1}\Sigma Q^{-1}\end{aligned}$$

   and

$$\Sigma=\left[\Omega+\alpha_o^2Q_1V_1Q_1^{'}+Q_1V_1Q_2^{'}+Q_2V_1Q_1^{'}\right].$$

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

$$\lim\frac{n}{R}=\infty.$$

It follows that

$$\begin{aligned}
\sqrt{R}\left(\widehat{\theta}-\theta_o\right) &= Q^{-1}\left[O\left(\sqrt{R}\right)\left\{\begin{array}{c} O_p\left(\frac{1}{C_{nR}}\right)+O_p\left(\frac{1}{\sqrt{n}}\right) \\ +O_p\left(\frac{1}{C_{nR}}\right)\end{array}\right\}\right]+o_p\left(1\right) \\
&= Q^{-1}\left[O_p\left(\frac{\sqrt{R}}{C_{nR}}\right)+O_p\left(\frac{\sqrt{R}}{\sqrt{n}}\right)+O_p\left(\frac{\sqrt{R}}{C_{nR}}\right)\right]+o_p\left(1\right) \\
&= Q^{-1}\left[O_p\left(1\right)+O_p\left(\frac{\sqrt{R}}{\sqrt{n}}\right)+O_p\left(1\right)\right]+o_p\left(1\right) \\
&= Q^{-1}\left[O_p\left(1\right)+O_p\left(\frac{\sqrt{R}}{\sqrt{n}}\right)+O_p\left(1\right)\right]+o_p\left(1\right) \\
&= O_p\left(1\right)
\end{aligned}$$

since

$$\lim \frac{R}{n} \rightarrow 0.$$

4. When R increases slower than n, the limiting distribution of

$$\begin{aligned}
\sqrt{n}\left(\widehat{\theta}-\theta_o\right) &= \sqrt{\frac{n}{R}}\sqrt{R}\left(\widehat{\theta}-\theta_o\right) \\
&= \sqrt{\frac{n}{R}}O_p\left(1\right)
\end{aligned}$$

diverges. Only when R increases faster than n is the limiting distribution of $\sqrt{n}\left(\widehat{\theta}-\theta_o\right)$ properly behaved.

5. The *iid* assumption for $(\varepsilon_{1i}, \boldsymbol{\varepsilon}_{2i})$ 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

$$\Omega = Ew_iw_i^{'}u_i^2 = \sigma_u^2 Q.$$

6. Note y_{2i}^* is endogenous in (1) if and only if $E\left(\varepsilon_{1i}\boldsymbol{\varepsilon}_{2i}\right) \neq 0$. We could use the results of Theorem 1 to test

$$H_0 : E\left(\varepsilon_{1i}\boldsymbol{\varepsilon}_{2i}\right) = 0.$$

Therefore testing $H_0 : E\left(\varepsilon_{1i}\boldsymbol{\varepsilon}_{2i}\right) = 0$ 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), 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

$$y_{1i}^* = x_i'\beta_o + \alpha_o \mathbf{y}_{2i}^* + \varepsilon_{1i} \tag{14}$$

$$\mathbf{y}_{2i}^* = z_i^{'}\delta_o + \boldsymbol{\varepsilon}_{2i} \tag{15}$$

where y_{1i}^* and $\mathbf{y}_{2i}^*$ are both latent variables with

$$\begin{bmatrix} \varepsilon_{1i} \\ \boldsymbol{\varepsilon}_{2i} \end{bmatrix} \overset{iid}{\sim} \left(0, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \right). \tag{16}$$

Rather than observing y_{1i}^* and $\mathbf{y}_{2i}^*$, we observe

$$y_{1i} = \tau_1 \left(y_{1i}^* \right)$$

and

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

respectively. The setup includes several different limited dependent variable models. For example, $\tau_1 \left(y_{1i}^* \right)$ could be $\max(0, y_{1i}^*)$ and $\tau_2 \left(\mathbf{y}_{2i}^* \right)$ could be $1 \left(\mathbf{y}_{2i}^* > 0 \right)$, 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

$$E \left[\varepsilon_{1i} | \boldsymbol{\varepsilon}_{2i} \right] = \rho_o \boldsymbol{\varepsilon}_{2i} \tag{17}$$

as in (5) we get

$$E \left[\varepsilon_{1i} | y_{2i} \right] = \rho_o E \left[\boldsymbol{\varepsilon}_{2i} | y_{2i} \right] \tag{18}$$

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

$$E \left[y_{1i}^* | y_{2i} \right] = x_i'\beta_o + \alpha_o E \left[\mathbf{y}_{2i}^* | y_{2i} \right] + E \left[\varepsilon_{1i} | y_{2i} \right] \tag{19}$$

Plugging equation (18) into equation (19) gives

$$E \left[y_{1i}^* | y_{2i} \right] = x_i'\beta_o + \alpha_o E \left[\mathbf{y}_{2i}^* | y_{2i} \right] + \rho_o E \left[\boldsymbol{\varepsilon}_{2i} | y_{2i} \right]$$

or

$$\begin{aligned} y_{1i}^* &= x_i'\beta_o + \alpha_o E \left[\mathbf{y}_{2i}^* | y_{2i} \right] + \rho_o E \left[\boldsymbol{\varepsilon}_{2i} | y_{2i} \right] + u_i \\ &= x_i'\beta_o + \alpha_o y_{2i}^* + \rho_o \varepsilon_{2i} + u_i \\ &= w_i^{'}\theta_o + u_i \end{aligned} \tag{20}$$

where $y_{2i}^{*}=E\left[\mathbf{y}_{2i}^{*}|y_{2i}\right]$, $\varepsilon_{2i}=E\left[\boldsymbol{\varepsilon}_{2i}|y_{2i}\right]$, and $u_i=y_{1i}^{*}-E\left[y_{1i}^{*}|y_{2i}\right]$ which has mean zero and variance $\sigma_u^2<\infty$. Note in (20) it does require to assume the joint distribution of $(\varepsilon_{1i},\varepsilon_{2i})$. If we knew $y_{2i}^{*}=\left[\mathbf{y}_{2i}^{*}|y_{2i}\right]$ and $\varepsilon_{2i}=E\left[\boldsymbol{\varepsilon}_{2i}|y_{2i}\right]$ we could estimate β_o, α_o, and ρ_o by a MLE if we knew the distribution of u_i or a GMM. With y_{2i}^{*} and ε_{2i} unobservable, we can use $\widetilde{y}_{2i}^{*}$ and $\widetilde{\varepsilon}_{2i}$ to approximate y_{2i}^{*} and ε_{2i} where $\widetilde{y}_{2i}^{*}$ and $\widetilde{\varepsilon}_{2i}$ are given in (10) and (12). Thus, we have

$$y_{1i}^{*}=\widehat{w}_i^{'}\theta_o+v_i$$

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

$$\mu_i=\alpha_o z_i^{'}\left(\widehat{\delta}-\delta_o\right)+(\alpha_o-\rho_o)\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right).$$

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

$$\frac{1}{n}\sum_{i=1}^{n}g\left(\widehat{w}_i,\theta\right)=0 \tag{21}$$

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

$$\begin{aligned}\sqrt{n}\left(\widehat{\theta}-\theta_o\right) &= -\left[\frac{1}{n}\sum_{i=1}^{n}\nabla_\theta g\left(\widehat{w}_i,\overline{\theta}\right)\right]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g\left(\widehat{w}_i,\theta_o\right)\\ &= -\left[\frac{1}{n}\sum_{i=1}^{n}\nabla_\theta g\left(\widehat{w}_i,\overline{\theta}\right)\right]^{-1}\\ &\quad\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g\left(w_i,\theta_o\right)+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nabla_w g\left(\overline{w}_i,\theta_o\right)\left(\widehat{w}_i-w_i\right)\right\}\end{aligned} \tag{22}$$

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

Assumption 7:

(1) $\frac{1}{n}\sum_{i=1}^{n}\nabla_w g\left(w_i,\theta_o\right)\nabla_w g\left(w_i,\theta_o\right)^{'}\xrightarrow{p}G_w$,

(2) $\frac{1}{n}\sum_{i=1}^{n}\nabla_\theta g\left(w_i,\theta_o\right)\xrightarrow{p}\mathbf{Q}$,

(3) $\frac{1}{n}\sum_{i=1}^{n}\nabla_{\theta w} g\left(w_i,\theta_o\right)\xrightarrow{p}G_{\theta w}$,

(4) $\frac{1}{n}\sum_{i=1}^{n}\nabla_{y_2} g\left(w_i,\theta_o\right)z_i^{'}\xrightarrow{p}\mathbf{Q}_1$,

[3] The generalized residual approach of Vella (1993, 1998) also has this difficulty of computing MLE.

(5) $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g\left(w_i,\theta_o\right) \xrightarrow{p} N\left(0,\mathbf{\Omega}\right)$, where

$$\mathbf{\Omega}=E\left[g\left(w_i,\theta_o\right)g\left(w_i,\theta_o\right)^{'}\right],$$

Theorem 2: Under Assumptions 1-7, and $\frac{n}{R}\to 0$ as $(n,R)\to\infty$, we have:

$$\sqrt{n}\left(\widehat{\theta}-\theta_o\right)=Q^{-1}\left\{\mathbf{S}_n+\mathbf{L}_n+\mathbf{Q}_{1n}+\mathbf{Q}_{2n}\right\}$$

where

$$\mathbf{S}_n=\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g\left(w_i,\theta_o\right),$$

$$\mathbf{L}_n=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nabla_{y_2} g\left(w_i,\theta_o\right) z_i^{'}\left(\widehat{\delta}-\delta\right),$$

$$\mathbf{Q}_{1n}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nabla_{y_2} g\left(w_i,\theta_o\right)\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right),$$

and

$$\mathbf{Q}_{2n}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nabla_{\varepsilon_2} g\left(w_i,\theta_o\right)\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right).$$

Furthermore, $\mathbf{S}_n=O_p\left(1\right)$, $\mathbf{L}_n=O_p\left(1\right)$, $\mathbf{Q}_{1n}=O_p\left(\frac{\sqrt{n}}{C_{nR}}\right)=O_p\left(\frac{\sqrt{n}}{\sqrt{n}}\right)=O_p\left(1\right)$, and $\mathbf{Q}_{2n}=O_p\left(\frac{\sqrt{n}}{C_{nR}}\right)=O_p\left(\frac{\sqrt{n}}{\sqrt{n}}\right)=O_p\left(1\right)$ since $R>n$.

Remarks:

1. Suppose that the simulated moment does not depend on the first stage estimation. It follows that

$$\begin{aligned}\sqrt{n}\left(\widehat{\theta}-\theta_o\right) &= -\mathbf{Q}^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g\left(w_i,\theta_o\right)+\frac{1}{n}\sum_{i=1}^{n}\nabla_{y_2} g\left(w_i,\theta_o\right) z_i^{'}\sqrt{n}\left(\widehat{\delta}-\delta_o\right)\right\}+o_p\left(1\right)\\ &\xrightarrow{d} N\left(0,\mathbf{V}_2\right)\end{aligned}$$

since

$$\left[\begin{array}{c}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g\left(w_i,\theta_o\right)\\ \sqrt{n}\left(\widehat{\delta}-\delta_o\right)\end{array}\right]\xrightarrow{d} N\left(0,\left(\begin{array}{cc}\mathbf{\Omega} & \mathbf{Q}_2\\ \mathbf{Q}_2^{'} & V_1\end{array}\right)\right)$$

where

$$Cov\left[\sqrt{n}\left(\widehat{\delta}-\delta_o\right),\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g\left(w_i,\theta_o\right)\right]=\mathbf{Q}_2,$$

$$\begin{aligned}\mathbf{V}_2 &= \mathbf{Q}^{-1}\left[\Omega+\mathbf{Q}_1V_1\mathbf{Q}_1^{'}+\mathbf{Q}_1V_1\mathbf{Q}_2^{'}+\mathbf{Q}_2V_1\mathbf{Q}_1^{'}\right]\mathbf{Q}^{-1},\\ &= \mathbf{Q}^{-1}\mathbf{\Sigma}\mathbf{Q}^{-1}\end{aligned}$$

and

$$\mathbf{\Sigma}=\Omega+\mathbf{Q}_1V_1\mathbf{Q}_1^{'}+\mathbf{Q}_1V_1\mathbf{Q}_2^{'}+\mathbf{Q}_2V_1\mathbf{Q}_1^{'}.$$

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 $\mathbf{y}_{2i}$ be the dummy endogenous regressor. We consider

$$y_{1i}^{*} = x_i'\beta_o + \alpha_o \mathbf{y}_{2i} + \varepsilon_{1i} \tag{23}$$

$$\mathbf{y}_{2i}^{*} = z_i^{'}\delta_o + \boldsymbol{\varepsilon}_{2i} \tag{24}$$

where

$$y_{1i} = \tau_1\left(y_{1i}^{*}\right)$$

and

$$y_{2i} = 1 \text{ if } \mathbf{y}_{2i}^{*} > 0;\ y_{2i} = 1 \text{ otherwise.}$$

Again we assume

$$E\left[\varepsilon_{1i}|\boldsymbol{\varepsilon}_{2i}\right] = \rho_o \boldsymbol{\varepsilon}_{2i}$$

and

$$E\left[\varepsilon_{1i}|y_{2i}\right] = \rho_o E\left[\boldsymbol{\varepsilon}_{2i}|y_{2i}\right]$$

such that

$$\rho_o = \frac{E\left(\varepsilon_{1i}\boldsymbol{\varepsilon}_{2i}\right)}{E\left(\varepsilon_{1i}^2\right)E\left(\boldsymbol{\varepsilon}_{2i}^2\right)}.$$

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

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

and

$$E\left[\mathbf{y}_{2i}^{*}|y_{2i}\right] = z_i^{'}\delta_o + E\left[\boldsymbol{\varepsilon}_{2i}|y_{2i}\right]. \tag{26}$$

Then rewrite (25) as

$$\begin{aligned} y_{1i}^{*} &= x_i'\beta_o + \alpha_o y_{2i} + \rho_o E\left[\boldsymbol{\varepsilon}_{2i}|y_{2i}\right] + u_i \\ &= x_i'\beta_o + \alpha_o y_{2i} + \rho_o \varepsilon_{2i} + u_i \\ &= w_i^{'}\theta_o + u_i \end{aligned} \tag{27}$$

where $\varepsilon_{2i} = E[\boldsymbol{\varepsilon}_{2i}|y_{2i}]$, $u_i = y_{1i}^* - E[y_{1i}^*|y_{2i}]$, $w_i = \left(x_i', y_{2i}, \varepsilon_{2i}\right)'$ 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 $\varepsilon_{2i} = E[\boldsymbol{\varepsilon}_{2i}|y_{2i}]$ can not be observed, we can estimate $\varepsilon_{2i} = E[\boldsymbol{\varepsilon}_{2i}|y_{2i}]$ by $\widetilde{\varepsilon}_{2i}$ as in (10). Thus

$$\begin{aligned} y_{1i}^* &= x_i'\beta_o + \alpha_o y_{2i} + \rho_o \widetilde{\varepsilon}_{2i} + \rho_o(\varepsilon_{2i} - \widetilde{\varepsilon}_{2i}) + u_i \\ &= \widehat{w}_i'\theta_o + \nu_i \end{aligned} \tag{28}$$

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

$$\frac{1}{n}\sum_{i=1}^{n} g(\widehat{w}_i, \theta) = 0 \tag{29}$$

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

$$\begin{aligned} \sqrt{n}\left(\widehat{\theta} - \theta_o\right) &= -\left[\frac{1}{n}\sum_{i=1}^{n} \nabla_\theta g\left(\widehat{w}_i, \overline{\theta}\right)\right]^{-1} \\ &\quad \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(w_i, \theta_o) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \nabla_w g(\overline{w}_i, \theta_o)(\widehat{w}_i - w_i)\right\}. \end{aligned}$$

Theorem 3: Under Assumptions 1-7 and $\frac{n}{R} \to 0$ as $(n,\ R) \to \infty$, we have:

$$\sqrt{n}\left(\widehat{\theta} - \theta_o\right) = -\mathbf{Q}^{-1}[\mathbf{S}_n + \mathbf{Q}_{2n}] + o_p(1)$$

where $\mathbf{S}_n$ and $\mathbf{Q}_{2n}$ 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., $\sqrt{n}\left(\widehat{\theta} - \theta_o\right)$ diverges, unless the number of simulated random variables, R, goes to infinity with a rate faster than the sample size, n, i.e., $\frac{n}{R} \to 0$ as $(n, R) \to \infty$. 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 $\widetilde{\varepsilon}_{2i} = \frac{1}{R}\sum_{j=1}^{R}\varepsilon_{2i}^{j}$. Let

$$\overline{\varepsilon}_{2i}^{j} = \frac{1}{R}\sum_{j=1}^{R}\varepsilon_{2i}^{j}\left(\delta_o\right)$$

and

$$\begin{aligned} q_i &= \left(\overline{\varepsilon}_{2i}^{j} - \varepsilon_{2i}\right)^2 \\ &= \left(\frac{1}{R}\sum_{j=1}^{R}\left(\varepsilon_{2i}^{j}\left(\delta_o\right) - \varepsilon_{2i}\right)\right)^2. \end{aligned}$$

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

$$\begin{aligned} E\left(q_i\right)^2 &= E\left(\frac{1}{R}\sum_{j=1}^{R}\left(\varepsilon_{2i}^{j}\left(\delta_o\right) - \varepsilon_{2i}\right)\right)^4 \\ &\leq \frac{c}{R^4}R^2\sum_{j=1}^{R}E\left(\varepsilon_{2}^{j}\left(\delta_o\right) - \varepsilon_{2}\right)^4 \\ &= \frac{c}{R^2}\sum_{j=1}^{R}E\left(\varepsilon_{2}^{j}\left(\delta_o\right) - \varepsilon_{2}\right)^4 \\ &= \frac{c}{R}E\left(\varepsilon_{2}^{j}\left(\delta_o\right) - \varepsilon_{2}\right)^4 \end{aligned}$$

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

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

since

$$E\left(\varepsilon_{2}^{j}\left(\delta_o\right) - \varepsilon_{2}\right)^4 = O\left(1\right)$$

from Assumption 2. Because ϵ is arbitrary,

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

It follows that

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{R}\sum_{j=1}^{R}\left(\varepsilon_{2i}^{j}(\delta_o)-\varepsilon_{2i}\right)\right)^2 \\
= & \frac{1}{n}\sum_{i=1}^{n}\left(\overline{\varepsilon}_{2i}-\varepsilon_{2i}\right)^2 \\
= & O_p\left(\frac{1}{R}\right). \qquad (30)
\end{aligned}$$

Now use C^r inequality to get

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)^2 \\
= & \frac{1}{n}\sum_{i=1}^{n}\left(\overline{\varepsilon}_{2i}-\varepsilon_{2i}+\widetilde{\varepsilon}_{2i}-\overline{\varepsilon}_{2i}\right)^2 \\
\leq & 2\left[\frac{1}{n}\sum_{i=1}^{n}\left(\overline{\varepsilon}_{2i}-\varepsilon_{2i}\right)^2+\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\varepsilon}_{2i}-\overline{\varepsilon}_{2i}\right)^2\right].
\end{aligned}$$

By a mean value theorem

$$\widetilde{\varepsilon}_{2i}-\overline{\varepsilon}_{2i} = \nabla\widetilde{\varepsilon}_{2i}\left(\overline{\delta}\right)\left(\widehat{\delta}-\delta_o\right)$$

with ∇ being the gradient operator and $\overline{\delta}$ lies between $\widehat{\delta}$ and δ_o. Then

$$\begin{aligned}
\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\varepsilon}_{2i}-\overline{\varepsilon}_{2i}\right)^2\right\| &= \left\|\frac{1}{n}\sum_{i=1}^{n}\left[\nabla\widetilde{\varepsilon}_{2i}\left(\overline{\delta}\right)\left(\widehat{\delta}-\delta_o\right)\right]^2\right\| \\
&\leq \left\|\widehat{\delta}-\delta_o\right\|^2\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\nabla\widetilde{\varepsilon}_{2i}\left(\overline{\delta}\right)\right\|^4\right)^{1/2} \\
&= O_p\left(\frac{1}{n}\right)O_p(1) \\
&= O_p\left(\frac{1}{n}\right).
\end{aligned}$$

Here we use the results that

$$\widehat{\theta}-\theta_o = O_p\left(\frac{1}{\sqrt{n}}\right)$$

the consistency $\overline{\delta}$ and

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\nabla\widetilde{\varepsilon}_{2i}\left(\overline{\delta}\right)\right\|^{4}=O_{p}\left(1\right).$$

It follows that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)^{2}\\
=\;& O_{p}\left(\frac{1}{R}\right)+O_{p}\left(\frac{1}{n}\right)\\
=\;& O_{p}\left(\frac{1}{C_{nR}^{2}}\right)
\end{aligned}$$

This proves part (a). Consider (b). By the Cauchy-Schwarz inequality

$$\begin{aligned}
\left\|\frac{1}{n}\sum_{i=1}^{n}x_{i}\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)\right\| &\leq \left(\frac{1}{n}\sum_{i=1}^{n}\left\|x_{i}\right\|^{2}\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)^{2}\right)^{1/2}\\
&= O_{p}\left(1\right)O_{p}\left(\frac{1}{C_{nR}}\right)\\
&= O_{p}\left(\frac{1}{C_{nR}}\right)
\end{aligned}$$

since

$$\frac{1}{n}\sum_{i=1}^{n}\left\|x_{i}\right\|^{2}=O_{p}\left(1\right).$$

Consider (c).

$$\begin{aligned}
\left\|\frac{1}{n}\sum_{i=1}^{n}y_{2i}^{*}\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)\right\| &\leq \left(\frac{1}{n}\sum_{i=1}^{n}\left\|y_{2i}^{*}\right\|^{2}\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)^{2}\right)^{1/2}\\
&= O_{p}\left(1\right)O_{p}\left(\frac{1}{C_{nR}}\right)\\
&= O_{p}\left(\frac{1}{C_{nR}}\right)
\end{aligned}$$

since

$$\frac{1}{n}\sum_{i=1}^{n}\left\|y_{2i}^{*}\right\|^{2}=O_{p}\left(1\right).$$

Consider (d).

$$\begin{aligned}
\left\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{2i}\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)\right\| &\leq \left(\frac{1}{n}\sum_{i=1}^{n}\|\varepsilon_{2i}\|^{2}\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)^{2}\right)^{1/2} \\
&= O_p\left(1\right)O_p\left(\frac{1}{C_{nR}}\right) \\
&= O_p\left(\frac{1}{C_{nR}}\right)
\end{aligned}$$

since

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

■

B Proof of Lemma 2

Proof. Note

$$\widetilde{y}_{2i}^{*}-y_{2i}^{*}=z_i^{'}\left(\widehat{\delta}-\delta_o\right)+\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right).$$

By the C^r inequality

$$\begin{aligned}
\|\widetilde{y}_{2i}^{*}-y_{2i}^{*}\|^{2} &= \left\|z_i^{'}\left(\widehat{\delta}-\delta_o\right)+\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)\right\|^{2} \\
&\leq 2\left(\left\|z_i^{'}\left(\widehat{\delta}-\delta_o\right)\right\|^{2}+\|\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\|^{2}\right) \\
&= 2\left(a_i+b_i\right)
\end{aligned}$$

where

$$a_i=\left\|z_i^{'}\left(\widehat{\delta}-\delta_o\right)\right\|^{2}$$

and

$$b_i=\|\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\|^{2}.$$

It follows that

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

Now

$$\left\|z_i^{'}\left(\widehat{\delta}-\delta_o\right)\right\|^{2}\leq\|z_i\|^{2}\left\|\left(\widehat{\delta}-\delta_o\right)\right\|^{2}.$$

Thus

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left\|z_i'\left(\widehat{\delta}-\delta_o\right)\right\|^2 &\leq \left\|\left(\widehat{\delta}-\delta_o\right)\right\|^2\frac{1}{n}\sum_{i=1}^{n}\|z_i\|^2 \\
&= \left[O_p\left(\frac{1}{\sqrt{n}}\right)\right]^2 O_p(1) \\
&= O_p\left(\frac{1}{n}\right)
\end{aligned}$$

because

$$\left(\widehat{\delta}-\delta_o\right)=O_p\left(\frac{1}{\sqrt{n}}\right)$$

and

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

by Assumptions 4 and 5. For (b), we have that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}b_i &= \frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)^2 \\
&= O_p\left(\frac{1}{C_{nR}^2}\right)
\end{aligned}$$

from Lemma 1. Combining these results, we have

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}(a_i+b_i) &= O_p\left(\frac{1}{n}\right)+O_p\left(\frac{1}{C_{nR}^2}\right) \\
&= O_p\left(\frac{1}{C_{nR}^2}\right).
\end{aligned}$$

This proves part (a). By the Cauchy-Schwarz inequality:

$$\left\|\frac{1}{n}\sum_{i=1}^{n}x_i\left(\widetilde{y}_{2i}^*-y_{2i}^*\right)\right\|\leq\left(\frac{1}{n}\sum_{i=1}^{n}\|x_i\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{y}_{2i}^*-y_{2i}^*\right)^2\right)^{1/2}$$

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

$$\begin{aligned}
\left\|\frac{1}{n}\sum_{i=1}^{n}y_{2i}^*\left(\widetilde{y}_{2i}^*-y_{2i}^*\right)\right\| &\leq \left(\frac{1}{n}\sum_{i=1}^{n}\|y_{2i}^*\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{y}_{2i}^*-y_{2i}^*\right)^2\right)^{1/2} \\
&= O_p(1)\,O_p\left(\frac{1}{C_{nR}}\right)=O_p\left(\frac{1}{C_{nR}}\right).
\end{aligned}$$

Consider (d).

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{y}_{2i}^{*}-y_{2i}^{*}\right)\varepsilon_{2i}\right\| &\leq \left(\frac{1}{n}\sum_{i=1}^{n}\|(\widetilde{y}_{2i}^{*}-y_{2i}^{*})\|^{2}\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{2i}^{2}\right)^{1/2} \\ &= O_p\left(\frac{1}{C_{nR}}\right)O_p\left(1\right)=O_p\left(\frac{1}{C_{nR}}\right).\end{aligned}$$

Consider (e).

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{y}_{2i}^{*}-y_{2i}^{*}\right)\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)\right\| &\leq \left(\frac{1}{n}\sum_{i=1}^{n}\|(\widetilde{y}_{2i}^{*}-y_{2i}^{*})\|^{2}\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)^{2}\right)^{1/2} \\ &= O_p\left(\frac{1}{C_{nR}}\right)O_p\left(\frac{1}{C_{nR}}\right) \\ &= O_p\left(\frac{1}{C_{nR}^{2}}\right).\end{aligned}$$

■

C Proof of Lemma 3

Proof. Note

$$\sum_{i=1}^{n}\widehat{w}_i\widehat{w}_i^{'}=\sum_{i=1}^{n}\begin{bmatrix} x_ix_i^{'} & x_i\widetilde{y}_{2i}^{*} & x_i\widetilde{\varepsilon}_{2i} \\ \widetilde{y}_{2i}^{*}x_i^{'} & \widetilde{y}_{2i}^{*2} & \widetilde{y}_{2i}^{*}\widetilde{\varepsilon}_{2i} \\ \widetilde{\varepsilon}_{2i}x_i^{'} & \widetilde{y}_{2i}^{*}\widetilde{\varepsilon}_{2i} & \widetilde{\varepsilon}_{2i}^{2}\end{bmatrix}.$$

Hence using the results of Lemmas 1 and 2 we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}x_i\widetilde{y}_{2i}^{*} &= \frac{1}{n}\sum_{i=1}^{n}\left[x_iy_{2i}^{*}+x_i\left(\widetilde{y}_{2i}^{*}-y_{2i}^{*}\right)\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}x_iy_{2i}^{*}+\frac{1}{n}\sum_{i=1}^{n}x_i\left(\widetilde{y}_{2i}^{*}-y_{2i}^{*}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}x_iy_{2i}^{*}+O_p\left(\frac{1}{C_{nR}}\right),\end{aligned}$$

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}x_i\widetilde{\varepsilon}_{2i} &= \frac{1}{n}\sum_{i=1}^{n}x_i\varepsilon_{2i}+\frac{1}{n}\sum_{i=1}^{n}x_i\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}x_i\varepsilon_{2i}+O_p\left(\frac{1}{C_{nR}}\right),\end{aligned}$$

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\widetilde{y}_{2i}^{*2} &= \frac{1}{n}\sum_{i=1}^{n}\left(y_{2i}^{*}+\widetilde{y}_{2i}^{*}-y_{2i}^{*}\right)^{2} \\
&= \frac{1}{n}\sum_{i=1}^{n}y_{2i}^{*2}-2\frac{1}{n}\sum_{i=1}^{n}y_{2i}^{*}\left(\widetilde{y}_{2i}^{*}-y_{2i}^{*}\right)+\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{y}_{2i}^{*}-y_{2i}^{*}\right)^{2} \\
&= \frac{1}{n}\sum_{i=1}^{n}y_{2i}^{*2}+O_p\left(\frac{1}{C_{nR}}\right)+O_p\left(\frac{1}{C_{nR}^{2}}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}y_{2i}^{*2}+O_p\left(\frac{1}{C_{nR}}\right),
\end{aligned}$$

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\widetilde{y}_{2i}^{*}\widetilde{\varepsilon}_{2i} &= \frac{1}{n}\sum_{i=1}^{n}\left(y_{2i}^{*}+\widetilde{y}_{2i}^{*}-y_{2i}^{*}\right)\left(\varepsilon_{2i}+\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}y_{2i}^{*}\varepsilon_{2i}+\frac{1}{n}\sum_{i=1}^{n}y_{2i}^{*}\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right) \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{y}_{2i}^{*}-y_{2i}^{*}\right)\varepsilon_{2i}+\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{y}_{2i}^{*}-y_{2i}^{*}\right)\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}y_{2i}^{*}\varepsilon_{2i}+O_p\left(\frac{1}{C_{nR}}\right)+O_p\left(\frac{1}{C_{nR}}\right)+O_p\left(\frac{1}{C_{nR}^{2}}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}y_{2i}^{*}\varepsilon_{2i}+O_p\left(\frac{1}{C_{nR}}\right),
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\widetilde{\varepsilon}_{2i}^{2} &= \frac{1}{n}\sum_{i=1}^{n}\left(\varepsilon_{2i}+\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)^{2} \\
&= \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{2i}^{2}+2\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{2i}\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)+\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)^{2} \\
&= \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{2i}^{2}+O_p\left(\frac{1}{C_{nR}}\right)+O_p\left(\frac{1}{C_{nR}^{2}}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{2i}^{2}+O_p\left(\frac{1}{C_{nR}}\right).
\end{aligned}$$

Thus

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{w}_i\widehat{w}_i^{'}=\frac{1}{n}\sum_{i=1}^{n}ww_i^{'}+O_p\left(\frac{1}{C_{nR}}\right).$$

This proves (a). Consider (b). It is easy to see that

$$\widehat{w}_i-w_i=\begin{pmatrix} x_i-x_i \\ \widetilde{y}_{2i}^{*}-y_{2i}^{*} \\ \widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\end{pmatrix}$$

and

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left\|\widehat{w}_i-w_i\right\|^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{y}_{2i}^{*}-y_{2i}^{*}\right)^2+\frac{1}{n}\sum_{i=1}^{n}\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)^2\\ &= O_p\left(\frac{1}{C_{nR}^2}\right)+O_p\left(\frac{1}{C_{nR}^2}\right)\\ &= O_p\left(\frac{1}{C_{nR}^2}\right).\end{aligned}$$

Consider (c). Note

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{w}_i u_i=\frac{1}{n}\sum_{i=1}^{n}w_i u_i+\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{w}_i-w_i\right)u_i.$$

For the second term, given that the two variables $\widehat{w}_i - w_i$ 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).[4] Therefore

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{w}_i-w_i\right)u_i\right\| &\leq \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\widehat{w}_i-w_i\right\|^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}u_i^2}\,O_p\left(\frac{1}{\sqrt{n}}\right)\\ &= O_p\left(\frac{1}{C_{nR}}\right)O_p\left(1\right)O_p\left(\frac{1}{\sqrt{n}}\right)\\ &= O_p\left(\frac{1}{C_{nR}\sqrt{n}}\right)\end{aligned}$$

by part (b) and

$$\frac{1}{n}\sum_{i=1}^{n}u_i^2=O_p\left(1\right).$$

[4]Suppose we have two $I(0)$ processes x_i and y_i and we would like to examine the convergence speed of

$$\frac{1}{n}\sum_{i=1}^{n}x_i y_i.$$

Assume the law of large number holds here so that

$$\frac{1}{n}\sum_{i=1}^{n}x_i y_i\xrightarrow{p}E\left[xy\right]$$

which is $o_p\left(1\right)$ if x_i and y_i are uncorrelated and $O_p\left(1\right)$ if x_i and y_i are correlated. Hence

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^{n}x_i y_i\right\| &\leq \left(\frac{1}{n}\sum_{i=1}^{n}\left\|x_i\right\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|y_i\right\|^2\right)^{1/2}\\ &= O_p\left(1\right)O_p\left(1\right)\end{aligned}$$

if we assume $\frac{1}{n}\sum_{i=1}^{n}\left\|x_i\right\|^2<\infty$ and $\frac{1}{n}\sum_{i=1}^{n}\left\|y_i\right\|^2<\infty$. However we know $\frac{1}{n}\sum_{i=1}^{n}x_i y_i = O_p\left(1\right)$ only when $E\left[x_i y_i\right]\neq 0$. This means a correction term should be added in order to get a sharper bound when $E\left[x_i y_i\right]=0$ and the correction term is $O_p\left(\frac{1}{\sqrt{n}}\right)$ as shown in Trapani (2004).

Hence,

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{w}_i u_i &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i u_i + O_p\left(\frac{1}{C_{nR}}\right) \\
&= O_p(1) + O_p\left(\frac{1}{C_{nR}}\right) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i u_i + O_p\left(\frac{1}{C_{nR}}\right).
\end{aligned}$$

Consider (d).

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{w}_i \alpha_o z_i^{'}\left(\widehat{\delta}-\delta_o\right) &= \sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\left(w_i+\widehat{w}_i-w_i\right)\alpha_o z_i^{'}\left(\widehat{\delta}-\delta_o\right) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} w_i\alpha_o z_i^{'}\left(\widehat{\delta}-\delta_o\right)+\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{w}_i-w_i\right)\alpha_o z_i^{'}\left(\widehat{\delta}-\delta_o\right)\right) \\
&= \sqrt{n}\left(I+II\right).
\end{aligned}$$

We consider each term in turn.

$$\begin{aligned}
\|I\| &\leq \|\alpha_o\|\left\|\widehat{\delta}-\delta_o\right\|\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|w_i z_i^{'}\right\|^2} \\
&= O_p\left(\frac{1}{\sqrt{n}}\right)O_p(1) \\
&= O_p\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

since

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

$$\begin{aligned}
\|II\| &\leq \|\alpha_o\|\left\|\widehat{\delta}-\delta_o\right\|\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|\widehat{w}_i-w_i\right\|^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left\|z_i\right\|^2} \\
&= O_p\left(\frac{1}{\sqrt{n}}\right)O_p\left(\frac{1}{C_{nR}}\right)O_p(1) \\
&= O_p\left(\frac{1}{C_{nR}\sqrt{n}}\right).
\end{aligned}$$

since

$$\widehat{\delta}-\delta_o = O_p\left(\frac{1}{\sqrt{n}}\right)$$

and

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

by Assumptions 4 and 5. Hence

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{w}_i\alpha_o z_i^{'}\left(\widehat{\delta}-\delta_o\right) &= \sqrt{n}\left(O_p\left(\frac{1}{\sqrt{n}}\right)+O_p\left(\frac{1}{C_{nR}\sqrt{n}}\right)\right). \\ &= O_p(1)+O_p\left(\frac{1}{C_{nR}}\right).\end{aligned}$$

Consider (e). Note

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\widehat{w}_i(\alpha_o-\rho_o)(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}) \\ =\ &\frac{1}{n}\sum_{i=1}^{n}(w_i+\widehat{w}_i-w_i)(\alpha_o-\rho_o)(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}) \\ =\ &\frac{1}{n}\sum_{i=1}^{n}w_i(\alpha_o-\rho_o)(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}) \\ &+\frac{1}{n}\sum_{i=1}^{n}(\widehat{w}_i-w_i)(\alpha_o-\rho_o)(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i})\end{aligned}$$

Now

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^{n}w_i(\alpha_o-\rho_o)(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i})\right\| &\leq \|\alpha_o-\rho_o\|\left(\frac{1}{n}\sum_{i=1}^{n}\|w_i\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i})^2\right)^{1/2} \\ &= \|\alpha_o-\rho_o\|\times I\times II.\end{aligned}$$

We begin with I. Note

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

and

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

Next

$$\begin{aligned}&\left\|\frac{1}{n}\sum_{i=1}^{n}(\widehat{w}_i-w_i)(\alpha_o-\rho_o)(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i})\right\| \\ \leq\ &\|\alpha_o-\rho_o\|\left(\frac{1}{n}\sum_{i=1}^{n}\|\widehat{w}_i-w_i\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\|\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\|^2\right)^{1/2} \\ =\ &O_p\left(\frac{1}{C_{nR}^2}\right)\end{aligned}$$

since

$$\frac{1}{n}\sum_{i=1}^{n}\|\widehat{w}_i - w_i\|^2 = O_p\left(\frac{1}{C_{nR}^2}\right)$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\|\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\|^2 = O_p\left(\frac{1}{C_{nR}^2}\right)$$

Hence

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{w}_i\left(\alpha_o - \rho_o\right)\left(\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\right) &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} w_i\left(\alpha_o - \rho_o\right)\left(\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\right) + O_p\left(\frac{1}{C_{nR}^2}\right)\right) \\
&= O_p\left(\frac{\sqrt{n}}{C_{nR}}\right) + O_p\left(\frac{\sqrt{n}}{C_{nR}^2}\right).
\end{aligned}$$

■

D Proof of Theorem 1

Proof. Lemma 4 implies that

$$\begin{aligned}
\sqrt{n}\left(\widehat{\theta} - \theta_o\right) &= \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{w}_i\widehat{w}_i^{'}\right)^{-1}\left[\begin{array}{c} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{w}_i u_i + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{w}_i\alpha_o z_i^{'}\left(\widehat{\delta} - \delta_o\right) \\ +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{w}_i\left(\alpha_o - \rho_o\right)\left(\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\right) \end{array}\right] \\
&= \left[\frac{1}{n}\sum_{i=1}^{n} ww_i^{'} + O_p\left(\frac{1}{C_{nR}}\right)\right]^{-1} \\
&\quad \left[\begin{array}{c} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i u_i + O_p\left(\frac{1}{C_{nR}}\right) \\ +\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i\alpha_o z_i^{'}\left(\widehat{\delta} - \delta_o\right) + O_p\left(\frac{1}{C_{nR}}\right) \\ +\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i\left(\alpha_o - \rho_o\right)\left(\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\right) + O_p\left(\frac{\sqrt{n}}{C_{nR}^2}\right) \end{array}\right] \\
&= \left[\frac{1}{n}\sum_{i=1}^{n} ww_i^{'} + O_p\left(\frac{1}{C_{nR}}\right)\right]^{-1}\left[\left\{\begin{array}{c} O_p\left(1\right) + O_p\left(\frac{1}{C_{nR}}\right) \\ +O_p\left(1\right) + O_p\left(\frac{1}{C_{nR}}\right) \\ +O_p\left(\frac{\sqrt{n}}{C_{nR}}\right) + O_p\left(\frac{\sqrt{n}}{C_{nR}^2}\right) \end{array}\right\}\right].
\end{aligned}$$

■

E Proof of Theorem 2

Proof. Note by a Taylor expansion

$$\begin{aligned}
& \sqrt{n}\left(\widehat{\theta}-\theta_o\right) \\
= & -\left[\frac{1}{n}\sum_{i=1}^{n}\nabla_{\theta}g\left(\widehat{w}_i,\overline{\theta}\right)\right]^{-1} \\
& \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g\left(w_i,\theta_o\right)+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nabla_{w}g\left(\overline{w}_i,\theta_o\right)\left(\widehat{w}_i-w_i\right)\right\}.
\end{aligned}$$

By the Cauchy-Schwarz inequality

$$\begin{aligned}
& \left\|\frac{1}{n}\sum_{i=1}^{n}\nabla_{w}g\left(\overline{w}_i,\theta_o\right)\left(\widehat{w}_i-w_i\right)\right\| \\
\leq & \left(\frac{1}{n}\sum_{i=1}^{n}\left\|\nabla_{w}g\left(\overline{w}_i,\theta_o\right)\right\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\widehat{w}_i-w_i\right\|^2\right)^{1/2} \\
= & O_p\left(1\right)O_p\left(\frac{1}{C_{nR}}\right) \\
= & O_p\left(\frac{1}{C_{nR}}\right)
\end{aligned}$$

since

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\nabla_{w}g\left(\overline{w}_i,\theta_o\right)\right\|^2=O_p\left(1\right)$$

by Assumption 6 and

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\widehat{w}_i-w_i\right\|^2=O_p\left(\frac{1}{C_{nR}^2}\right)$$

from Lemma 3 (b). Expanding $\nabla_{\theta}g\left(\widehat{w}_i,\theta_o\right)$ around w_i gives

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\nabla_{\theta}g\left(\widehat{w}_i,\theta_o\right) &= \frac{1}{n}\sum_{i=1}^{n}\nabla_{\theta}g\left(w_i,\theta_o\right)+\frac{1}{n}\sum_{i=1}^{n}\nabla_{\theta w}g\left(\overline{w}_i,\theta_o\right)\left(\widehat{w}_i-w_i\right) \\
&= I+II.
\end{aligned}$$

Consider II.

$$\begin{aligned}
\left\|\frac{1}{n}\sum_{i=1}^{n}\nabla_{\theta w}g\left(\overline{w}_i,\theta_o\right)\left(\widehat{w}_i-w_i\right)\right\| &\leq \left(\frac{1}{n}\sum_{i=1}^{n}\left\|\nabla_{\theta w}g\left(\overline{w}_i,\theta_o\right)\right\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\widehat{w}_i-w_i\right\|^2\right)^{1/2} \\
&= O_p\left(1\right)O_p\left(\frac{1}{C_{nR}}\right) \\
&= O_p\left(\frac{1}{C_{nR}}\right) \\
&= o_p\left(1\right)
\end{aligned}$$

since

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\nabla_{\theta w}g\left(\overline{w}_i,\theta_o\right)\right\|^2=O_p\left(1\right)$$

by Assumption 6 and the consistency of $\overline{w}_i$. Because $\overline{\theta}$ is a consistent estimator of θ_o, it follows that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\nabla_{\theta}g\left(\widehat{w}_i,\overline{\theta}\right) &= \frac{1}{n}\sum_{i=1}^{n}\nabla_{\theta}g\left(w_i,\theta_o\right)+o_p\left(1\right)\\ &= \mathbf{Q}+o_p\left(1\right).\end{aligned}$$

It follows that

$$\begin{aligned}\sqrt{n}\left(\widehat{\theta}-\theta_o\right) &= -\mathbf{Q}^{-1}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}g\left(w_i,\theta_o\right)+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nabla_{w}g\left(\overline{w}_i,\theta_o\right)\left(\widehat{w}_i-w_i\right)\right\}\\ &= -\mathbf{Q}^{-1}\left\{O_p\left(1\right)+\sqrt{n}O_p\left(\frac{1}{C_{nR}}\right)\right\}+o_p\left(1\right).\end{aligned}$$

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

$$\begin{aligned}\sqrt{n}\left(\widehat{\theta}-\theta_o\right) &= -\mathbf{Q}^{-1}\left\{O_p\left(1\right)+O_p\left(\frac{\sqrt{n}}{\sqrt{n}}\right)\right\}+o_p\left(1\right)\\ &= -\mathbf{Q}^{-1}\left\{O_p\left(1\right)+O_p\left(1\right)\right\}+o_p\left(1\right).\end{aligned}$$

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

Note

$$\begin{aligned}\nabla_{w}g\left(\overline{w}_i,\theta_o\right)\left(\widehat{w}_i-w_i\right) &= \nabla_{x}g\left(\overline{w}_i,\theta_o\right)\left(x_i-x_i\right)+\nabla_{y_2}g\left(\overline{w}_i,\theta_o\right)\left(\widetilde{y}_{2i}^*-y_{2i}^*\right)\\ &\quad+\nabla_{\varepsilon_2}g\left(\overline{w}_i,\theta_o\right)\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)\\ &= \nabla_{y_2}g\left(\overline{w}_i,\theta_o\right)\left(\widetilde{y}_{2i}^*-y_{2i}^*\right)+\nabla_{\varepsilon_2}g\left(\overline{w}_i,\theta_o\right)\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right).\end{aligned}$$

By the triangle inequality we have

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=1}^{n}\nabla_{w}g\left(\overline{w}_i,\theta_o\right)\left(\widehat{w}_i-w_i\right)\right\| &\leq \left\|\frac{1}{n}\sum_{i=1}^{n}\nabla_{y_2}g\left(\overline{w}_i,\theta_o\right)\left(\widetilde{y}_{2i}^*-y_{2i}^*\right)\right\|\\ &\quad+\left\|\frac{1}{n}\sum_{i=1}^{n}\nabla_{\varepsilon_2}g\left(\overline{w}_i,\theta_o\right)\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)\right\|\\ &= I+II.\end{aligned}$$

Consider I.

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\nabla_{y_2}g\left(\overline{w}_i,\theta_o\right)\left(\widetilde{y}_{2i}^*-y_{2i}^*\right)\right\|\leq\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\nabla_{y_2}g\left(\overline{w}_i,\theta_o\right)\right\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\widetilde{y}_{2i}^*-y_{2i}^*\right\|^2\right)^{1/2}$$

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

$$\begin{aligned}
\left\|\frac{1}{n}\sum_{i=1}^{n}\nabla_{\varepsilon_2} g\left(\overline{w}_i,\theta_o\right)\left(\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right)\right\| &\leq \left(\frac{1}{n}\sum_{i=1}^{n}\left\|\nabla_{y_2} g\left(\overline{w}_i,\theta_o\right)\right\|^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\widetilde{\varepsilon}_{2i}-\varepsilon_{2i}\right\|^2\right)^{1/2} \\
&= O_p(1)\,O_p\left(\frac{1}{C_{nR}}\right) \\
&= O_p\left(\frac{1}{C_{nR}}\right)
\end{aligned}$$

since

$$\frac{1}{n}\sum_{i=1}^{n}\left\|\nabla_{y_2} g\left(\overline{w}_i,\theta_o\right)\right\|^2 = O_p(1)$$

and

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

Combining I and II, we get

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\nabla_w g\left(\overline{w}_i,\theta_o\right)\left(\widehat{w}_i-w_i\right) \\
=\ & O\left(\sqrt{n}\right)O_p\left(\frac{1}{C_{nR}}\right)+O\left(\sqrt{n}\right)O_p\left(\frac{1}{C_{nR}}\right) \\
=\ & O\left(\sqrt{n}\right)O_p\left(\frac{1}{\sqrt{n}}\right)+O\left(\sqrt{n}\right)O_p\left(\frac{1}{\sqrt{n}}\right) \\
=\ & O_p(1)+O_p(1) \\
=\ & O_p(1)
\end{aligned}$$

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

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{y_2} g\left(w_i, \theta_o\right)\left(\widetilde{y}_{2i}^{*} - y_{2i}^{*}\right) \\
= & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{y_2} g\left(w_i, \theta_o\right)\left(\widetilde{y}_{2i}^{*} - y_{2i}^{*}\right) \\
= & \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{y_2} g\left(w_i, \theta_o\right)\left[z_i^{'}\left(\widehat{\delta} - \delta\right) + \left(\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\right)\right] \\
= & \frac{1}{n} \sum_{i=1}^{n} \nabla_{y_2} g\left(w_i, \theta_o\right) z_i^{'} \sqrt{n}\left(\widehat{\delta} - \delta\right) \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{y_2} g\left(w_i, \theta_o\right)\left(\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\right) \\
= & O_p(1) + O\left(\sqrt{n}\right) O_p\left(\frac{1}{C_{nR}}\right) \\
= & O_p(1) + O_p\left(\frac{\sqrt{n}}{C_{nR}}\right) \\
= & O_p(1) + O_p(1) \\
= & O_p(1)
\end{aligned}$$

if $\frac{n}{R} \to 0$. Hence

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{y_2} g\left(w_i, \theta_o\right)\left(\widetilde{y}_{2i}^{*} - y_{2i}^{*}\right) \\
= & \frac{1}{n} \sum_{i=1}^{n} \nabla_{y_2} g\left(w_i, \theta_o\right) z_i^{'} \sqrt{n}\left(\widehat{\delta} - \delta_o\right) \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{y_2} g\left(w_i, \theta_o\right)\left(\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\right) + o_p(1) \\
= & \mathbf{Q}_1 \sqrt{n}\left(\widehat{\delta} - \delta_o\right) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{y_2} g\left(w_i, \theta_o\right)\left(\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\right) + o_p(1) .
\end{aligned}$$

It follows that

$$\begin{aligned}
& \sqrt{n}\left(\widehat{\theta} - \theta_o\right) \\
= & -\mathbf{Q}^{-1}\left\{\begin{array}{c}
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g\left(w_i, \theta_o\right) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{y_2} g\left(w_i, \theta_o\right) z_i^{'}\left(\widehat{\delta} - \delta\right) \\
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{y_2} g\left(w_i, \theta_o\right)\left(\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\right) \\
+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{\varepsilon_2} g\left(w_i, \theta_o\right)\left(\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\right)
\end{array}\right\} + o_p(1) .
\end{aligned}$$

This proves Theorem 2. ■

F Proof of Theorem 3

Proof.

Note

$$
\begin{aligned}
& \left\| \frac{1}{n} \sum_{i=1}^{n} \nabla_w g\left(\overline{w}_i, \theta_o\right)\left(\widehat{w}_i - w_i\right) \right\| \\
\leq \quad & \left(\frac{1}{n} \sum_{i=1}^{n} \left\| \nabla_w g\left(\overline{w}_i, \theta_o\right) \right\|^2 \right)^{1/2} \left(\frac{1}{n} \sum_{i=1}^{n} \left\| \widehat{w}_i - w_i \right\|^2 \right)^{1/2}
\end{aligned}
$$

and

$$
\begin{aligned}
& \frac{1}{n} \sum_{i=1}^{n} \left\| \widehat{w}_i - w_i \right\|^2 \\
= \quad & \frac{1}{n} \sum_{i=1}^{n} \left(\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\right)^2 \\
= \quad & O_p\left(\frac{1}{C_{nR}^2} \right).
\end{aligned}
$$

Hence

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_w g\left(\overline{w}_i, \theta_o\right)\left(\widehat{w}_i - w_i\right) = O_p\left(\frac{\sqrt{n}}{C_{nR}} \right).
$$

It is easy to see that

$$
\sqrt{n}\left(\widehat{\theta} - \theta_o\right) = -\mathbf{Q}^{-1} \left\{ O_p(1) + O_p\left(\frac{\sqrt{n}}{C_{nR}} \right) \right\} + o_p(1).
$$

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

$$
\begin{aligned}
& \sqrt{n}\left(\widehat{\theta} - \theta_o\right) \\
= \quad & -\mathbf{Q}^{-1} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g\left(w_i, \theta_o\right) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_w g\left(w_i, \theta_o\right)\left(\widehat{w}_i - w_i\right) \right] + o_p(1) \\
= \quad & -\mathbf{Q}^{-1} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} g\left(w_i, \theta_o\right) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla_{\varepsilon_2} g\left(w_i, \theta_o\right)\left(\widetilde{\varepsilon}_{2i} - \varepsilon_{2i}\right) \right] + o_p(1)
\end{aligned}
$$

This proves Theorem 3. ■

References

[1] Ahn, H., and Powell, J. L. (1993), "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58, 3-29.

[2] Amemiya, T., (1978), "The Estimation of a Simultaneous Equation Generalized Probit Models," *Econometrica*, 46, 1193-1205.

[3] Amemiya, T. (1979), "The Estimation of a Simultaneous Tobit Model," *International Economic Review*, 20, 169-181.

[4] Amemiya, T. (1985), *Advanced Econometrics*, Cambridge, Harvard University Press.

[5] Angrist, J. (2001), "Estimation of Limited-Dependent Variable Models with Binary Endogenous Regressors: Simple Strategies for Empirical Practice," *Journal of Business and Economic Statistics*, 19, 2-16.

[6] Bai, J., and Ng, S. (2002), "Determining the Number of Factors in Approximate Factor Models," *Econometrica*, 70, 191-221.

[7] Blundell, R. W., and Smith, R. (1994), "Coherency and Estimation in Simultaneous Models with Censored or Qualitative Dependent Variables," *Journal of Econometrics*, 64, 355-373.

[8] Blundell, R. W., and Powell, J. L. (2004a), "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, 71, 655-679.

[9] Blundell, R. W., and Powell, J. L. (2004b), "Censored Regression Quantiles with Endogenous Regressors," Working paper, University College, London.

[10] Blundell, R. W., and Smith, R. J. (1989), "Estimation in a Class of Simultaneous Equation Limited Dependent Variable Models," *Review of Economic Studies*, 56, 37-58.

[11] Borsch-Supan, A., and Hajivassiliou, V. A. (1993), "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models," *Journal of Econometrics*, 58, 347-368.

[12] Breslaw, J. A., and McIntosh, J. (1998), "Simulated Latent Variable Estimation of Models with Ordered Categorical Data," *Journal of Econometrics*, 25-47.

[13] Carrasco, M., and Florens, J. P. (2002), "Simulation-Based Method of Moments and Efficiency," *Journal of Business & Economic Statistics*, 20, 482-492.

[14] Chen, S., and Khan, S. (2003), "Semiparametric Estimation of Heteroskedastic Sample Selection Models," *Econometric Theory*, 19, 1040-1064.

[15] Christofides, L. N., Li, Q., Liu, Z., and Min, I. (2003), "Recent Two-Stage Sample Selection Procedures with an Application to the Gender Wage Gap," *Journal of Business & Economic Statistics*, 21, 396-405.

[16] Darolles, S., Florens, S., and Renault, E. (2002), "Nonparametric Instrumental Regression," GREMAQ, University of Toulouse.

[17] Das, M. (2005), "Instrumental Variables Estimators of Nonparametric Models with Discrete Endogenous Regressors," *Journal of Econometrics*, 124, 335-361.

[18] Das, M., Newey, W. K., and Vella, F. (2003), "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies*, 70, 33-58.

[19] Dhrymes, P. J. (1970), *Econometrics: Statistical Foundations and Applications*, New York: Springer-Verlag.

[20] Duffie, D., and Singleton, K. L. (1993), "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929-952.

[21] Evans, W. N., and Schwab, R. M. (1995), "Finishing High School and Starting College: Do Catholic Schools make a Difference," *Quarterly Journal of Economics*, 110, 941-974.

[22] Evans, W., Farelly, M. C., and Montgomery, E. (1999), "Do Workplace Smoking Bans Reduce Smoking," *American Economic Review*, 89, 728-747.

[23] Geweke, J. (1991), "Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints," In *Computer Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface*, pp. 571-578, Alexandria, Virginia: American Statistical Association.

[24] Goldman, D., Bhattacharya, J., McCaffrey, D., Duan, N., Leibowitz, A., and Morton, S. (2001), "The Effect of Insurance on Mortality in an HIV+Population in Care," *Journal of the American Statistical Association*, 96, 883-894.

[25] Hajivassiliou, V., McFadden, D., and Ruud, P. (1996), "Simulation of Multivariate Normal Rectangle Probabilities and Their Derivatives," *Journal of Econometrics*, 72, 85-134.

[26] Heckman, J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931-959.

[27] Heckman, J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153-162.

[28] Hsiao, C. (1983), "Regression Analysis with a Categorized Explanatory Variable," in *Studies in Econometrics, Time Series, and Multivariate Statistics*, eds. S. Karlin, T. Amemiya, and L. Goodman, New York, Academic Press, pp. 934-129.

[29] Hsiao, C., and Mountain, D. (1985), "Estimating the Short-Run Income Elasticity of Demand for Electricity by Using Cross-Sectional Categorized Data," *Journal of the American Statistical Association*, 80, 259-265.

[30] Kao, C., and Wu, C. (1990), "Two-Step Estimation of Linear Models with Ordinal Unobserved Variables: The Case of Corporate Bonds," *Journal of Business & Economic Statistics*, 8, 317-325.

[31] Keane, M. P. (1993), "A Computationally Practical Simulation Estimator for Panel Data," *Econometrica*, 62, 95-116.

[32] Lee, L-F. (1978), "Unionism and Wage Rates: A Simultaneous Equation Model with Qualitative and Limited Dependent Variables," *International Economic Review*, 19, 415-433.

[33] Lee, L-F. (1979), "Identification and Estimation in Binary Choice Models with Limited (Censored) Dependent Variables," *Econometrica*, 47, 977-996.

[34] Lee, L-F. (1994), "Semiparametric Instrumental Variable Estimation of Simultaneous Equation Sample Selection Models," *Journal of Econometrics*, 63, 341-388.

[35] Lee, L-F. (1995), "Asymptotic Bias in Simulated Maximum Likelihood Estimation of Discrete Choice Models," *Econometric Theory*, 11, 437-483.

[36] Lee, L-F. (1999), "Simulation Estimation of Dynamic Switching Regression and Dynamic Disequilibrium Models - Some Monte Carlo Results," *Journal of Econometrics*, 78, 179-204.

[37] Lee, L-F. (1999a), "Estimation of Dynamic and ARCH Tobit Models," *Journal of Econometrics*, 92, 335-390.

[38] Lee, L-F. (1999b), "Statistical Inference with Simulated Likelihood Functions," *Econometric Theory*, 15, 337-360.

[39] Lee, M-J. (1995), "Semi-Parametric Estimation of Simultaneous Equations with Limited Dependent Variables: A Case Study of Female Labour Supply," *Journal of Applied Econometrics*, 10, 187-200.

[40] Lee, S. (2004), "Endogeneity in Quantile Regression Models: A Control Function Approach," Cemmap working paper CWP08/04.

[41] Lewbel, A. (2004), "Simple Estimators for Hard Problems: Endogeneity in Discrete Choice Related Models," Working paper, Boston College.

[42] Li, Q., and Wooldridge, J. M. (2002), "Semiparametric Estimation of Partially Linear Models for Dependent Data with Generated Regressors," *Econometric Theory*, 18, 625-645.

[43] Ma, L, and Koenker, R. (2004), "Quantile Regression Methods for Recursive Structural Equation Models," Working paper.

[44] Maddala, G. S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge, MIT Press.

[45] McFadden, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica*, 57, 995-1026.

[46] Nelson, F., and Olson, L. (1978), "Specification and Estimation of a Simultaneous Equation Model with Limited Dependent Variables," *International Economic Review*, 19, 695-710.

[47] Newey, W. K. (1985), "Semiparametric Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables," *Annales de L'INSEE*, 59/60, 219-236.

[48] Newey, W., Powell, J., and Vella, F. (1999), "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 67, 565-604.

[49] Pagan, A. (1984), "Econometric Issues in the Analysis of Regressions with Generated Regressors," *International Economic Review*, 25, 221-247.

[50] Pagan, A. (1986), "Two Stage and Related Estimators and Their Applications," *Review of Economic Studies*, 53, 517-538.

[51] Powell, J. L. (1984), "Least Absolute Deviations Estimation of the Censored Regression Model," *Journal of Econometrics*, 54, 1435-1460.

[52] Powell, J. L. (1986), "Symmetrically Trimmed Least Squares Estimation of Tobit Models," *Econometrica*, 54, 1435-1460.

[53] Pakes, A. D., and Pollard, D. (1989), "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027-1057.

[54] Rivers, D., and Vuang, Q. (1988), "Limited Information Estimators and Exogeneity. Tests for Simultaneous Probit Models," *Journal of Econometrics*, 39, 347-366.

[55] Robinson, P. M. (1988), "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931-954.

[56] Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York, Wiley.

[57] Smith, R. J., and Blundell, R. W. (1986), "An Exogeneity. Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply," *Econometrica*, 54, 679-685.

[58] Stern, S. (1992), "A Method for Smoothing Simulated Moments of Discrete Probabilities in Multivariate Probit Models," *Econometrica*, 60, 943-952.

[59] Terza, J. V. (1987), "Estimating Linear Models with Ordinal Qualitative Regressors," *Journal of Econometrics*, 34, 275-291.

[60] Terza, J. V. (1998), "Estimating Count Data Models with Endogenous Switching: Sampling Selection and Endogenous Treatment Effects," *Journal of Econometrics*, 84, 129-154.

[61] Trapani, L. (2004), "On the Use of the Cauchy-Schwarz Inequality for Asymptotic Theory," Working paper, Cass Business School, London.

[62] Vella, F. (1993), "A Simple Estimator for Simultaneous with Censored Endogenous Regressors," *International Economic Review*, 34, 441-457.

[63] Vella, F. (1998), "Estimating Models with Sample Selection Bias: A Survey," *Journal of Human Resources*, 33, 127-169.

[64] Vella, F., and Verbeek, M. (1999a), "Two-step Estimation of Panel Data Models with Censored Endogenous Variables and Selection Bias," *Journal of Econometrics*, 90, 239-263.

[65] Vella, F., and Verbeek, M. (1999b), "Estimating and Interpreting Models with Endogenous Treatment," *Journal of Business & Economic Statistics*, 17, 473-478..

[66] Vytlacil, E., and Yildiz, N. (2004), "Dummy Endogenous Variables in Weakly Separable Models," Working paper, Department of Economics, Stanford University.

[67] Willis, R. and Rosen, S. (1979), "Education and Self-Selection," *Journal of Political Economy*, 87, S1-S36.

[68] Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MIT Press.

[69] Yildiz, N. (2004), "Estimation of Binary Choice Models with Linear Index and Dummy Endogenous Variables," Working paper, Department of Economics, Stanford University.

[70] Zhang, W., and Lee, L. F. (2004) "Simulation Estimation of Dynamic Discrete Choice Panel Models with Accelerated Importance Samplers," *Econometrics Journals*, 7, 120-142.