INFERENCE ON CONDITIONAL MOMENT RESTRICTION MODELS WITH GENERATED VARIABLES

RYO KIMOTO AND TAISUKE OTSU

Abstract. A seminal work by Domínguez and Lobato (2004) proposed a consistent estimation method for conditional moment restrictions, which does not rely on additional identification assumptions as in the GMM estimator using unconditional moments and is free from any user-chosen number. Their methodology is further extended by Domínguez and Lobato (2015, 2020) for consistent specification testing of conditional moment restrictions, which may involve generated variables. We follow up this literature and derive the asymptotic distribution of Domínguez and Lobato's (2004) estimator that involves generated variables. Our simulation result illustrates that ignoring proxy errors in the generated variables may cause severer distortions for the coverage or size properties of statistical inference on parameters.

1. Introduction

A seminal work by Domínguez and Lobato (2004) (hereafter, DL) proposed a consistent estimation method for conditional moment restrictions, which does not rely on additional identification assumptions as in the GMM estimator based on unconditional moments and is free from any user-chosen number. Their methodology is further extended by Domínguez and Lobato (2015) for consistent specification testing of conditional moment restrictions.

In empirical analysis, it is often the case that econometric models contain latent or theoretical variables, which are unobservable but may be estimated by observable data. Such variables, called generated variables, are often obtained as fitted values or residuals of preliminary regression fitting. Common examples include expected values of prices or sales, total factor productivity, relative quality of firms, among others. A recent paper by Domínguez and Lobato (2020) studied specification testing for conditional moment restrictions, where conditioning variables are generated ones. The results in Domínguez and Lobato (2020) are important in at least two senses: (a) their test can be considered as a generalization of Ramsey's (1969) RESET for possibly nonlinear conditional moment restrictions, and (b) the asymptotic distributions of their specification test statistics same as the case where there is no estimation error in generated variables. The latter is theoretically a very interesting finding since estimation errors for generated variables typically change asymptotic distributions of statistics (Pagan, 1984).

We follow up the analysis in Domínguez and Lobato (2020) by investigating asymptotic properties of the DL estimator when conditional moment restrictions involve generated variables. After accepting the null of specification testing by Domínguez and Lobato (2020), researchers typically proceed to conduct statistical inference on the parameters in the conditional moment restrictions. Then it is critical to characterize the asymptotic distribution of the point estimator by DL. Given the finding (b) in Domínguez and Lobato (2020), it is also of interest to see whether the estimation errors of the generated variables will change the asymptotic distribution of the DL estimator.

In this paper, we derive the asymptotic distributions of the DL estimator of conditional moment restrictions for two cases: (i) only conditioning variables are generated, and (ii) variables in both the conditioning set and moment function are generated. We find that in Case (i), the asymptotic distribution of the DL estimator is not affected by estimation errors for generated variables. On the other hand, in Case (ii), estimation errors for generated variables change the limiting distribution of the DL estimator. In particular, additional components due to such errors emerge in the asymptotic variance formula. We propose a consistent estimator for those additional components and asymptotically valid inference method for parameters in conditional moment restrictions. Our simulation result illustrates that ignoring proxy errors in the generated variables may cause severer distortions for the coverage or size properties of statistical inference on parameters.

2. Main result

Our notation closely follows that of DL. Consider the conditional moment restriction

$$E[h(W_t, \theta_0)|X_t] = 0 \quad \text{a.s.}, \tag{1}$$

for a unique $\theta_0 \in \Theta \subset \mathbb{R}^{d_\theta}$, where $h : \mathbb{R}^{d_W} \times \Theta \to \mathbb{R}$ is a known function up to θ_0 and $X_t \in \mathbb{R}^{d_X}$ is a vector of conditioning variables. Let $I\{A\}$ be the indicator function for an event A. From Billingsley (1995, Theorem 16.10 iii), the restriction (1) holds true if and only if

$$H(\theta_0, x) = E[h(W_t, \theta_0) I\{X_t \le x\}] \quad \text{for almost every } x \in \mathbb{R}^{d_X},$$

where $I\{X_t \le x\} = \prod_{d=1}^{d_X} I\{X_t^{(d)} \le x^{(d)}\}$ for the d-th elements $X_t^{(d)}$ and $x^{(d)}$ of X_t and x, respectively. Thus, the true parameter value θ_0 can be alternatively written as

$$\theta_0 = \arg\min_{\theta \in \Theta} \int H(\theta, x)^2 dP_X(x),$$

where P_X is the probability measure of X_t. By taking its sample analog, DL proposed the following estimator

$$\hat{\theta}_{DL} = \arg\min_{\theta \in \Theta} \frac{1}{n^3} \sum_{\ell=1}^{n} \left(\sum_{t=1}^{n} h(W_t, \theta_0) I\{X_t \le X_\ell\} \right)^2. \tag{2}$$

DL studied asymptotic properties of this estimator. Notably this estimator is free from any user-specified constant and circumvents a potential inconsistency problem of the GMM estimator based on unconditional moment restrictions implied from (1).

This paper is concerned with the situation, where X_t is unobservable and specified as $X_t = X(Z_t, \beta_0)$ for observables $Z_t \in \mathbb{R}^{d_Z}$ and unknown parameters $\beta_0 \in B \subset \mathbb{R}^{d_\beta}$. We also assume that some estimator $\hat{\beta}$ for β_0 is available to the researcher so that the generated variable $\hat{X}_t = X(Z_t, \hat{\beta})$ can be used as a proxy of X_t. In particular, we focus on the following cases:

- Case (i): W_t is observable (i.e., W_t and X_t have no overlap),
- Case (ii): W_t contains X_t (say, $W_t = (Y_t', X_t')'$).

For each case, we study asymptotic properties of the DL estimator, where X_t is replaced with the generated variable $\hat{X}_t$. Indeed a recent paper by Domínguez and Lobato (2020) studied specification testing of the model in (1) with generated variables. In contrast, this paper focuses on inference for the parameter θ_0.

2.1. **Case (i): W_t is observable.** We first consider the case where W_t and X_t have no overlap and W_t is observable. In this case, the DL estimator is defined as

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \frac{1}{n^3} \sum_{\ell=1}^{n} \left(\sum_{t=1}^{n} h(W_t, \theta) I\{\hat{X}_t \leq \hat{X}_\ell\} \right)^2, \tag{3}$$

i.e., only the variables in the indicator are generated variables. Let $\dot{h}(W_t, \theta) = \frac{\partial}{\partial \theta} h(W_t, \theta)$. To study asymptotic properties of $\hat{\theta}$, we impose the following assumptions.

Assumption.

- **(i):** *$\{W_t, Z_t\}$ is ergodic and strictly stationary.*
- **(ii):** *$E[h(W_t, \theta)|X_t] = 0$ a.s. if and only if $\theta = \theta_0$. $h(w, \cdot)$ is continuous in Θ for a.e. $w \in \mathbb{R}^{d_W}$. Θ is compact and $\theta_0 \in int(\Theta)$. $h(w, \cdot)$ is once continuously differentiable in a neighborhood of θ_0 and satisfies $E[\sup_{\theta \in \mathcal{N}_\theta} |\dot{h}(W_t, \theta)|] < \infty$ for a neighborhood $\mathcal{N}_\theta$ of θ_0. $E[\sup_{\theta \in \Theta} |h(W_t, \theta)|] < \infty$ and $E[h(W_t, \theta_0)^4 ||X_t||^{1+\delta}] < \infty$. $h(W_t, \theta_0)$ is a martingale difference sequence with respect to $\{(W_s, Z_s) : s \leq t\}$.*
- **(iii):** *The density of $X_t = X(Z_t, \beta_0)$ given the past is bounded and continuous. $X(Z_t, \beta)$ is differentiable with respect to β in a neighborhood $\mathcal{N}_\beta$ of β_0, and $\max_{1 \leq t \leq n} \sup_{\beta \in \mathcal{N}_\beta} |\partial X(Z_t, \beta)/\partial \beta'| = o_p(n^{1/2})$. The estimator $\hat{\beta}$ admits the asymptotic linear form $\sqrt{n}(\hat{\beta} - \beta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi_t(\beta_0) + o_p(1)$ for some influence function $\psi_t(\cdot)$ satisfying $E[\psi_t(\beta_0)] = 0$ and $E[||\psi_t(\beta_0)||^2] < \infty$.*

Assumptions (i) and (ii) are same as the ones in DL and used to derive the asymptotic properties of the infeasible estimator $\hat{\theta}_{DL}$ in (2). Assumption (iii) lists additional requirements on the generated variables, which are analogous to the ones in Domínguez and Lobato (2020). The form of the influence function $\psi_t(\cdot)$ changes with the estimator $\hat{\beta}$, but this assumption is typically satisfied for various $\sqrt{n}$-consistent estimators.

Under these assumptions, the asymptotic distribution of $\hat{\theta}$ is obtained as follows.

Theorem 1. *Suppose W_t is observable. Under Assumptions (i)-(iii), it holds*

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Omega),$$

where

$$\Omega = \left(\int \dot{H} \dot{H}' dP_X \right)^{-1} \left[\int \int \dot{H}(x_1) \dot{H}(x_2)' \Gamma(x_1, x_2) dP_X(x_1) dP_X(x_2) \right] \left(\int \dot{H} \dot{H}' dP_X \right)^{-1},$$

$\dot{H}(x) = E[\dot{h}(W_t, \theta_0) I\{X_t \leq x\}]$, and $\Gamma(x_1, x_2) = E[h(W_t, \theta_0)^2 I\{X_t \leq x_1 \wedge x_2\}]$.

It is interesting to note that the asymptotic distribution of $\hat{\theta}$ is same as that of the infeasible DL estimator $\hat{\theta}_{DL}$. In other words, the proxy error $\hat{X}_t - X_t$ is asymptotically negligible in this case. By taking sample counterparts, the asymptotic variance Ω can be consistently estimated by

$$\hat{\Omega} = \frac{1}{n^2}\sum_{l=1}^{n}\sum_{l_1=1}^{n}\widehat{\dot{H}}(\hat{X}_l)\left[\frac{1}{n}\sum_{t=1}^{n}h(W_t,\theta_0)^2 I\{X_t \le \hat{X}_l \wedge \hat{X}_{l_1}\}\right]\widehat{\dot{H}}(\hat{X}_{l_1})',$$

where $\widehat{\dot{H}}(a) = \frac{1}{n}\sum_{t=1}^{n}\dot{h}(W_t,\hat{\theta})I\{\hat{X}_t \le a\}$.

2.2. **Case (ii): W_t contains X_t.** We next consider the case where $W_t = (Y_t', X_t')'$ contains unobservable X_t. In this case, the DL estimator is defined as

$$\tilde{\theta} = \arg\min_{\theta\in\Theta}\frac{1}{n^3}\sum_{\ell=1}^{n}\left(\sum_{t=1}^{n}h(Y_t,\hat{X}_t,\theta)I\{\hat{X}_t \le \hat{X}_\ell\}\right)^2, \tag{4}$$

i.e., the generated variables $\hat{X}_t$ appear not only in the indicator function $I\{\cdot\}$ but also in the moment function $h(\cdot)$. The asymptotic distribution of $\tilde{\theta}$ is obtained as follows.

Theorem 2. *Suppose $W_t = (Y_t', X_t')'$. Under Assumptions (i)-(iii), it holds*

$$\sqrt{n}(\tilde{\theta} - \theta_0) \xrightarrow{d} N(0, A'VA),$$

where $\dot{H}_\beta(x) = E[\{\partial h(Y_t, X(Z_t,\beta_0),\theta_0)/\partial\beta\}I\{X(Z_t,\beta_0) \le x\}]$,

$$A' = \left[\left(\int \dot{H}\dot{H}'dP_X\right)^{-1} : \left(\int \dot{H}\dot{H}'dP_X\right)^{-1}\int \dot{H}\dot{H}_\beta' dP_X\right],$$

$$V = \begin{bmatrix} \int\int \dot{H}(x_1)\dot{H}(x_2)'\Gamma(x_1,x_2)dP_X(x_1)dP_X(x_2) & \int \dot{H}\dot{H}_\beta' dP_X \\ \int \dot{H}_\beta \dot{H}' dP_X & E[\psi_t(\beta_0)\psi_t(\beta_0)'] \end{bmatrix}.$$

It should be noted that in this case the influence of the proxy error $\hat{X}_t - X_t$ changes the asymptotic distribution of the estimator. Letting $A' = [A_1' : A_2']$ and $V = \begin{bmatrix} V_1 & C \\ C' & V_2 \end{bmatrix}$ based on the above expressions, the asymptotic variance of $\tilde{\theta}$ can be written as

$$A'VA = A_1'V_1A_1 + A_2'C'A_1 + A_1'CA_2 + A_2'V_2A_2.$$

Intuitively the first term $A_1'V_1A_1$ is equivalent to the asymptotic variance Ω in Theorem 1 for $\hat{\theta}$ or the infeasible estimator $\hat{\theta}_{DL}$, and the other three terms are additional ones due to the proxy error $\hat{X}_t - X_t$.

By taking the sample counterparts, the asymptotic variance of $\tilde{\theta}$ can be consistently estimated by $\hat{V}_\theta = (\hat{A}_1' : \hat{A}_2') \begin{bmatrix} \hat{V}_1 & \hat{C} \\ \hat{C}' & \hat{V}_2 \end{bmatrix} \begin{pmatrix} \hat{A}_1 \\ \hat{A}_2 \end{pmatrix}$, where $\widehat{\dot{H}}(a) = \frac{1}{n}\sum_{t=1}^n \dot{h}(Y_t, X(Z_t, \hat{\beta}), \hat{\theta}) I\{\hat{X}_t \leq a\}$,

$$\begin{aligned}
\hat{A}_1' &= \left(\frac{1}{n} \sum_{l=1}^n \widehat{\dot{H}}(\hat{X}_l) \widehat{\dot{H}}(\hat{X}_l)' \right)^{-1}, \\
\hat{A}_2' &= \left(\frac{1}{n} \sum_{l=1}^n \widehat{\dot{H}}(\hat{X}_l) \widehat{\dot{H}}(\hat{X}_l)' \right)^{-1} \frac{1}{n} \sum_{l=1}^n \widehat{\dot{H}}(\hat{X}_l) \left(\frac{1}{n} \sum_{t=1}^n \frac{\partial h(Y_t, X(Z_t, \hat{\beta}), \hat{\theta})}{\partial \beta} I\{\hat{X}_t \leq \hat{X}_l\} \right)', \\
\hat{V}_1 &= \frac{1}{n^2} \sum_{l=1}^n \sum_{l_1=1}^n \widehat{\dot{H}}(\hat{X}_l) \left[\frac{1}{n} \sum_{t=1}^n h(Y_t, X(Z_t, \hat{\beta}), \theta_0)^2 I\{X_t \leq \hat{X}_l \wedge \hat{X}_{l_1}\} \right] \widehat{\dot{H}}(\hat{X}_{l_1})', \\
\hat{V}_2 &= \frac{1}{n} \sum_{t=1}^n \psi_t(\hat{\beta}) \psi_t(\hat{\beta})', \qquad \hat{C} = \frac{1}{n} \sum_{t=1}^n \widehat{\dot{H}}(X_t) h(Y_t, X(Z_t, \hat{\beta}), \hat{\theta}) \psi_t(\hat{\beta})'.
\end{aligned}$$

3. Simulation

This section illustrates the theoretical results obtained in the last section by a simulation study. Based on DL, we consider the following data generating process:

$$Y_t = \theta_0^2 X_t + \theta_0 X_t^2 + U_t, \qquad X_t = \beta_0^{(1)} + \beta_0^{(2)} Z_t,$$

where $(Z_t, U_t) \sim N(0, I_2)$ for $t = 1, \ldots, 100$. We set $\theta_0 = 5/4$ and $(\beta_0^{(1)}, \beta_0^{(2)}) = (0, 1)$. The generated variable $\hat{X}_t = \hat{\beta}_0^{(1)} + \hat{\beta}_0^{(2)} Z_t$ is given by the OLS fitted values for the regression from $\tilde{X}_t = \beta_0^{(1)} + \beta_0^{(2)} Z_t + \xi(Z_t) V_t$ on Z_t with $V_t \sim N(0, \sigma^2)$. To assess the effect of the noise in the generated variables, we consider $\xi(Z_t) = 1$ (homoskedastic case) and $\xi(Z_t) = \sqrt{0.1 + 0.2 Z_t + 0.3 Z_t^2}$ (heteroskedastic case). For σ^2, we consider $\sigma^2 \in \{0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00\}$.

Under this setup, we compute the point estimator $\tilde{\theta}$ in (4) and associated confidence intervals based on the variance estimator $\hat{V}_\theta$ obtained in the last section and the variance estimator $\hat{A}_1' \hat{V}_1 \hat{A}_1$, which does not take into account for the generated variables (called "Adjusted" and "Unadjusted" in Table 1, respectively). Table 1 reports the biases and standard deviations of $\tilde{\theta}$, and coverage frequencies of the adjusted and unadjusted confidence intervals based on 20,000 Monte Carlo replications.

From Table 1, we can clearly see that as the noise level σ^2 for the generated variable increases, the unadjusted standard errors tend to be too small and exhibit severe under-coverages. On the other hand, the adjusted standard error based on the asymptotic distribution in Theorem 2 works well in terms of coverages across all cases.

Table 1. Simulation results

σ^2	$\tilde{\theta}$ Bias	SD	Homoskedastic Unadjusted	Adjusted	Heteroskedastic Unadjusted	Adjusted
0.25	0.0022	0.0706	86.605%	95.625%	87.1%	95.43%
0.50	0.0066	0.1110	68.8%	95.15%	70.195%	94.845%
0.75	0.0144	0.1613	55.485%	94.76%	56.8%	94.59%
1.00	0.0261	0.2221	46.785%	94.395%	47.815%	94.265%
1.25	0.0431	0.3043	42.195%	93.955%	42.295%	93.78%
1.50	0.0658	0.4060	39.395%	93.45%	39.135%	93.25%
1.75	0.0899	0.4931	38.04%	92.89%	37.085%	92.8%
2.00	0.1211	0.6444	37.93%	92.325%	36.54%	92.215%

Appendix A. Mathematical appendix

A.1. **Proof of Theorem 1.** The consistency $\hat{\theta} \xrightarrow{p} \theta_0$ can be shown by adapting the argument in the proof of Theorem 1 of DL.

Similar to DL, an expansion of the first order condition of $\hat{\theta}$ yields

$$\sqrt{n}(\hat{\theta}-\theta_0) = G_n^{-1}\frac{1}{n}\sum_{l=1}^n \left(\frac{1}{n}\sum_{t=1}^n \dot{h}(W_t,\hat{\theta})I\{\hat{X}_t \le \hat{X}_l\}\right)\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n h(W_t,\theta_0)I\{\hat{X}_t \le \hat{X}_l\}\right)$$

$$\begin{aligned}
&= G_n^{-1}\frac{1}{n}\sum_{l=1}^n \left(\frac{1}{n}\sum_{t=1}^n \dot{h}(W_t,\hat{\theta})I\{\hat{X}_t \le \hat{X}_l\}\right)\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n h(W_t,\theta_0)[I\{\hat{X}_t \le \hat{X}_l\} - I\{X_t \le X_l\}]\right)\\
&\quad + G_n^{-1}\frac{1}{n}\sum_{l=1}^n \left(\frac{1}{n}\sum_{t=1}^n \dot{h}(W_t,\hat{\theta})[I\{\hat{X}_t \le \hat{X}_l\} - I\{X_t \le X_l\}]\right)\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n h(W_t,\theta_0)I\{X_t \le X_l\}\right)\\
&\quad + G_n^{-1}\frac{1}{n}\sum_{l=1}^n \left(\frac{1}{n}\sum_{t=1}^n \dot{h}(W_t,\hat{\theta})I\{X_t \le X_l\}\right)\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n h(W_t,\theta_0)I\{X_t \le X_l\}\right)\\
&=: T_1 + T_2 + T_3,
\end{aligned}$$

where $G_n = \frac{1}{n}\sum_{l=1}^n \left(\frac{1}{n}\sum_{t=1}^n \dot{h}(W_t,\hat{\theta})I\{\hat{X}_t \le \hat{X}_l\}\right)\left(\frac{1}{n}\sum_{t=1}^n \dot{h}(W_t,\tilde{\theta})' I\{\hat{X}_t \le \hat{X}_l\}\right)$, and $\tilde{\theta}$ is on the line joining $\hat{\theta}$ and θ_0. Note that Assumption (iii) guarantees $\max_{1\le l \le n}|\hat{X}_l - X_l| = o_p(1)$. For T_1 and T_2, observe that for any arbitrarily small $\bar{\delta} > 0$,

$$\begin{aligned}
|T_1| &\le \frac{1}{n}\sum_{l=1}^n \left|\frac{1}{n}\sum_{t=1}^n \dot{h}(W_t,\hat{\theta})I\{\hat{X}_t \le \hat{X}_l\}\right| \sup_{x,\delta\in[-\bar{\delta},\bar{\delta}]}\left|\frac{1}{\sqrt{n}}\sum_{t=1}^n h(W_t,\theta_0)[I\{X_t \le x+\delta\} - I\{X_t \le x\}]\right|,\\
|T_2| &\le \frac{1}{n}\sum_{l=1}^n \sup_{x,\delta\in[-\bar{\delta},\bar{\delta}]}\left|\frac{1}{n}\sum_{t=1}^n \dot{h}(W_t,\hat{\theta})[I\{X_t \le x+\delta\} - I\{X_t \le x\}]\right|\left|\frac{1}{\sqrt{n}}\sum_{t=1}^n h(W_t,\theta_0)I\{X_t \le X_l\}\right|.
\end{aligned}$$

Thus, by applying the arguments in the proof of Dominguez and Lobato (2020, Proposition 1), we obtain $T_1 \xrightarrow{p} 0$ and $T_2 \xrightarrow{p} 0$. Similarly we can show that

$$\begin{aligned}
T_3 &= \left[\frac{1}{n}\sum_{l=1}^n \left(\frac{1}{n}\sum_{t=1}^n \dot{h}(W_t,\hat{\theta})I\{X_t \le X_l\}\right)\left(\frac{1}{n}\sum_{t=1}^n \dot{h}(W_t,\tilde{\theta})' I\{X_t \le X_l\}\right)\right]^{-1}\\
&\quad \times \frac{1}{n}\sum_{l=1}^n \left(\frac{1}{n}\sum_{t=1}^n \dot{h}(W_t,\hat{\theta})I\{X_t \le X_l\}\right)\left(\frac{1}{\sqrt{n}}\sum_{t=1}^n h(W_t,\theta_0)I\{X_t \le X_l\}\right) + o_p(1).
\end{aligned}$$

Therefore, the conclusion follows by Lemmas 1 and 2 of DL.

A.2. **Proof of Theorem 2.** The consistency results $\hat{\theta} \xrightarrow{p} \theta_0$ and $\hat{\beta} \xrightarrow{p} \beta_0$ can be shown by adapting the argument in the proof of Theorem 1 of DL and Assumption (iii), respectively.

By expanding the first order condition of $\hat{\theta}$ around $(\hat{\theta}, \hat{\beta}) = (\theta_0, \beta_0)$, we obtain

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{l=1}^{n}\left(\frac{1}{n}\sum_{t=1}^{n}\dot{h}(Y_t, X(Z_t, \hat{\beta}), \hat{\theta})I\{\hat{X}_t \leq \hat{X}_l\}\right)\left(\frac{1}{n}\sum_{t=1}^{n}h(Y_t, X(Z_t, \hat{\beta}), \hat{\theta})I\{\hat{X}_t \leq \hat{X}_l\}\right) \\ &= \hat{g}_n + \hat{G}_{\theta,n}(\hat{\theta} - \theta_0) + \hat{G}_{\beta,n}(\hat{\beta} - \beta_0), \end{aligned}$$

where

$$\begin{aligned} \hat{g}_n &= \frac{1}{n}\sum_{l=1}^{n}\dot{H}_{1,n}(\hat{X}_l)\left(\frac{1}{n}\sum_{t=1}^{n}h(Y_t, X(Z_t, \beta_0), \theta_0)I\{\hat{X}_t \leq \hat{X}_l\}\right), \\ \hat{G}_{\theta,n} &= \frac{1}{n}\sum_{l=1}^{n}\dot{H}_{1,n}(\hat{X}_l)\left(\frac{1}{n}\sum_{t=1}^{n}\dot{h}(Y_t, X(Z_t, \tilde{\beta}), \tilde{\theta})I\{\hat{X}_t \leq \hat{X}_l\}\right)', \\ \hat{G}_{\beta,n} &= \frac{1}{n}\sum_{l=1}^{n}\dot{H}_{1,n}(\hat{X}_l)\left(\frac{1}{n}\sum_{t=1}^{n}\dot{h}_\beta(Y_t, X(Z_t, \tilde{\beta}), \tilde{\theta})I\{\hat{X}_t \leq \hat{X}_l\}\right)', \\ \dot{H}_{1,n}(x) &= \frac{1}{n}\sum_{t=1}^{n}\dot{h}(Y_t, X(Z_t, \hat{\beta}), \hat{\theta})I\{\hat{X}_t \leq x\}, \qquad \dot{h}_\beta(Y_t, X(Z_t, \beta), \theta) = \frac{\partial h(Y_t, X(Z_t, \beta), \theta)}{\partial \beta}. \end{aligned}$$

By applying the same argument in the proof of Theorem 1, we have

$$\sqrt{n}\hat{g}_n = \sqrt{n}g_n + o_p(1), \quad \hat{G}_{\theta,n} = G_{\theta,n} + o_p(1), \quad \hat{G}_{\beta,n} = G_{\beta,n} + o_p(1),$$

where

$$\begin{aligned} g_n &= \frac{1}{n}\sum_{l=1}^{n}\dot{H}_{2,n}(X_l)\left(\frac{1}{n}\sum_{t=1}^{n}h(Y_t, X(Z_t, \beta_0), \theta_0)I\{X_t \leq X_l\}\right), \\ G_{\theta,n} &= \frac{1}{n}\sum_{l=1}^{n}\dot{H}_{2,n}(X_l)\left(\frac{1}{n}\sum_{t=1}^{n}\dot{h}(Y_t, X(Z_t, \tilde{\beta}), \tilde{\theta})I\{X_t \leq X_l\}\right)', \\ G_{\beta,n} &= \frac{1}{n}\sum_{l=1}^{n}\dot{H}_{2,n}(X_l)\left(\frac{1}{n}\sum_{t=1}^{n}\dot{h}_\beta(Y_t, X(Z_t, \tilde{\beta}), \tilde{\theta})I\{X_t \leq X_l\}\right)', \\ \dot{H}_{2,n}(x) &= \frac{1}{n}\sum_{t=1}^{n}\dot{h}(Y_t, X(Z_t, \beta_0), \theta_0)I\{X_t \leq x\}. \end{aligned}$$

Solving for $(\hat{\theta} - \theta_0)$ yields

$$\sqrt{n}(\hat{\theta} - \theta_0) = -[G_{\theta,n}^{-1} : G_{\theta,n}^{-1}G_{\beta,n}]\begin{bmatrix} \sqrt{n}g_n \\ \sqrt{n}(\hat{\beta} - \beta_0) \end{bmatrix} + o_p(1).$$

Now Lemma 1 of DL implies $G_{\theta,n} \xrightarrow{p} \int \dot{H}\dot{H}' dP_X$. Also an adaptation of Lemma 1 of DL implies $G_{\beta,n} \xrightarrow{p} \int \dot{H}\dot{H}_\beta' dP_X$. Finally, letting $\eta_t(\cdot) = E[h(Y_t, X(Z_t, \beta_0), \theta_0), \psi_t(\beta_0)']' I\{X_t \leq \cdot\}]$, an adaptation of Lemma 2 of DL implies that $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\eta_t(\cdot)$ weakly converges to a centered Gaussian process with the covariance kernel $E[\eta_t(x_1)\eta_t(x_2)']$. Therefore, the continuous mapping theorem yields the conclusion.

References

[1] Billingsley, P. (1995) *Probability and Measure*, New York: Wiley.

[2] Domínguez, M. and I. N. Lobato (2004) Consistent estimation of models defined by conditional moment restrictions, *Econometrica*, 72, 1601-1615.

[3] Domínguez, M. and I. N. Lobato (2015) A simple omnibus overidentification specification test for time series econometric models, *Econometric Theory*, 31, 891-910.

[4] Domínguez, M. and I. N. Lobato (2020) Specification testing with estimated variables, *Econometric Reviews*, 39, 476-494.

[5] Pagan, A. (1984) Econometric issues in the analysis of regressions with generated regressors, *International Economic Review*, 25, 221-247.

[6] Ramsey, J. B. (1969) Tests for specification errors in classical linear least squares regression analysis, *Journal of the Royal Statistical Society*, B 31, 350-371.

Department of Economics, The Pennsylvania State University, University Park, PA 16801, USA.

Email address: ryo.kimoto@psu.edu

Department of Economics, London School of Economics, Houghton Street, London, WC2A 2AE, UK.

Email address: t.otsu@lse.ac.uk