Supplement to "Efficient two-step generalized empirical likelihood estimation and tests with martingale differences"

Fei Jin[1] and Lung-fei Lee[*2]

[1]School of Economics, Fudan University, and Shanghai Institute of International Finance and Economics, Shanghai 200433, China

[2]Department of Economics, The Ohio State University, Columbus, OH 43210, USA

May 5, 2020

Contents

[*]Corresponding author. Tel.: +1 614 292 5508; fax: +1 614 292 3906. E-mail addresses: jin.fei@live.com (Fei Jin), lee.1777@osu.edu (Lung-fei Lee).

D TGEL estimators with a consistent estimator of the transformation matrix

In the presence of an initial consistent estimator $\check{\alpha}$ of nuisance parameters, we may use the moment $C_n(\check{\alpha},\tilde{\beta})g_{ni}(\check{\alpha},\beta)$ instead of $C_n(\check{\alpha},\beta)g_{ni}(\check{\alpha},\beta)$ for a TGEL estimation; in the presence of $\dot{\alpha}_n(\beta)$, we may use $C_n(\dot{\alpha}_n(\check{\beta}),\check{\beta})g_{ni}(\dot{\alpha}_n(\beta),\beta)$ instead of $C_n(\dot{\alpha}_n(\beta),\beta)g_{ni}(\dot{\alpha}_n(\beta),\beta)$, where $\tilde{\beta}$ and $\check{\beta}$ are consistent estimators of β_0. Denote $\tilde{C}_n = C_n(\check{\alpha},\tilde{\beta})$ and $\check{C}_n = C_n(\dot{\alpha}_n(\check{\beta}),\check{\beta})$ for simplicity. The $\tilde{C}_n$ and $\check{C}_n$ are consistent estimators of $\lim_{n\to\infty}\bar{C}_n$, so unknown β only appears in $g_{ni}(\check{\alpha},\beta)$ or $g_{ni}(\dot{\alpha}_n(\beta),\beta)$. Using $\tilde{C}_n$ or $\check{C}_n$ involves one more estimation step for $\tilde{\beta}$ or $\check{\beta}$, but the additional estimation does not seem to cause much computational burden compared with the double optimization of GEL due to its saddle-point characterization.

For later analysis, we use the following $\tilde{\beta}$ and $\breve{\beta}$ for the analysis on higher order biases:

$$\tilde{\beta} = \arg\min_{\beta\in\mathcal{B}} g'_{nb}(\check{\alpha},\beta)\hat{W}_n^{-1} g_{nb}(\check{\alpha},\beta), \tag{D.1}$$

and

$$\breve{\beta} = \arg\min_{\beta\in\mathcal{B}} g'_{nb}(\dot{\alpha}_n(\beta),\beta)\hat{W}_n^{-1} g_{nb}(\dot{\alpha}_n(\beta),\beta), \tag{D.2}$$

where $\hat{W}_n$ is an $m_b \times m_b$ positive definite matrix. We maintain the following assumption to investigate the asymptotic properties of $\tilde{\beta}$ and $\breve{\beta}$.

Assumption D.1. *(i) β_0 is the unique solution to* $\lim_{n\to\infty} \mathbf{E}[g_{nb}(\alpha_0,\beta)] = 0$ *and* $\lim_{n\to\infty} \mathbf{E}[g_{nb}(\alpha_n(\beta),\beta)] = 0$; *(ii) there exists a nonstochastic $\bar{W}_n$ such that $\hat{W}_n = \bar{W}_n + O_p(n^{-1/2})$ and $\lim_{n\to\infty} \bar{W}_n$ is positive definite; (iii) for the case with $\check{\alpha}$,* $\mathrm{rank}(\lim_{n\to\infty} \bar{G}_{nb\beta}) = k_\beta$*, where $\bar{G}_{nb\beta} = \mathbf{E}(\frac{\partial g_{nb}(\gamma_0)}{\partial \beta'})$; for the case with $\dot{\alpha}_n(\beta)$,* $\mathrm{rank}(\lim_{n\to\infty}(\bar{G}_{nb\beta} + \bar{G}_{nb\alpha}\alpha_{n\beta})) = k_\beta$*, where $\alpha_{n\beta} = \frac{\partial \alpha_n(\beta_0)}{\partial \beta'}$.*

With $\tilde{C}_n g_{ni}(\check{\alpha},\beta)$ and $\breve{C}_n g_{ni}(\dot{\alpha}_n(\beta),\beta)$, we have the following TGEL estimators

$$\hat{\beta}_{\text{TGEL3}} = \arg\min_{\beta\in\mathcal{B}} \max_{\mu\in\Lambda_{nd}(\check{\alpha},\beta)} \sum_{i=1}^{n} \rho\Big(\mu'\tilde{C}_n g_{ni}(\check{\alpha},\beta)\Big), \tag{D.3}$$

and

$$\hat{\beta}_{\text{TGEL4}} = \arg\min_{\beta\in\mathcal{B}} \sup_{\mu\in\Lambda_{nd}(\dot{\alpha}_n(\beta),\beta)} \sum_{i=1}^{n} \rho\Big(\mu'\breve{C}_n g_{ni}(\dot{\alpha}_n(\beta),\beta)\Big). \tag{D.4}$$

The corresponding estimates of the auxiliary parameter μ are denoted by, respectively, $\hat{\mu}_{\text{TGEL3}}$ and $\hat{\mu}_{\text{TGEL4}}$. The following assumption contains the identification conditions for $\hat{\beta}_{\text{TGEL3}}$ and $\hat{\beta}_{\text{TGEL4}}$.

Assumption D.2. *β_0 is the unique solution to* $\lim_{n\to\infty} \bar{C}_n \mathbf{E}[g_n(\alpha_0,\beta)] = 0$ *and* $\lim_{n\to\infty} \bar{C}_n \mathbf{E}[g_n(\alpha_n(\beta),\beta)] = 0$.

The TGEL estimators $\hat{\beta}_{\text{TGEL3}}$ and $\hat{\beta}_{\text{TGEL4}}$ have the same first order asymptotic properties as those of $\hat{\beta}_{\text{TGEL}}$ and $\hat{\beta}_{\text{TGEL2}}$ in the main text, which are summarized in the following theorem.[1]

Theorem D.1. *Suppose that Assumptions 1–2, D.1(i)–(ii) and D.2 are satisfied.*

(i) $\sqrt{n}(\hat{\beta}_{\text{TGEL3}}-\beta_0) \xrightarrow{d} N\Big(0, \lim_{n\to\infty}(\bar{D}'_{n\beta}\bar{\Omega}_{nd}^{-1}\bar{D}_{n\beta})^{-1}\Big)$ *and* $\sqrt{n}(\hat{\beta}_{\text{TGEL4}}-\beta_0) \xrightarrow{d} N\Big(0, \lim_{n\to\infty}(\bar{D}'_{n\beta}\bar{\Omega}_{nd}^{-1}\bar{D}_{n\beta})^{-1}\Big)$.

[1] We omit the proof of this theorem since it is similar to that of Theorem 1.

(ii) $2\Big[\sum_{i=1}^n \rho\Big(\hat{\mu}'_{\text{TGEL}_3}\tilde{C}_n g_{ni}(\check{\alpha},\hat{\beta}_{\text{TGEL}_3})\Big) - n\rho(0)\Big] \xrightarrow{d} \chi^2(m_b - k_\beta)$ *and* $2[\sum_{i=1}^n \rho(\hat{\mu}'_{\text{TGEL}_4}\check{C}_n g_{ni}(\dot{\alpha}_n(\hat{\beta}_{\text{TGEL}_4}),\hat{\beta}_{\text{TGEL}_4})) - n\rho(0)] \xrightarrow{d} \chi^2(m_b - k_\beta)$.

(iii) $\hat{\beta}_{\text{TGEL}_3}$ *and* $\hat{\beta}_{\text{TGEL}_4}$ *are generally less efficient relative to* $\hat{\beta}_{\text{GEL}}$*, where* $\hat{\beta}_{\text{GEL}}$ *is the joint GEL estimator of* β*, which is a subvector of* $\hat{\gamma}_{\text{GEL}}$ *in* (2.8). *But if* $m_a = k_\alpha$*, then* $\hat{\beta}_{\text{TGEL}_3}$ *and* $\hat{\beta}_{\text{TGEL}_4}$ *(will be denoted as* $\hat{\beta}_{\text{E-TGEL}_3}$ *and* $\hat{\beta}_{\text{E-TGEL}_4}$*) are asymptotically equivalent to* $\hat{\beta}_{\text{GEL}}$.

(iv) *If* $m_a = k_\alpha$ *and* $\mathbf{E}(\sup_{\alpha\in\mathcal{A},\beta\in\mathcal{N}_\beta}\|\frac{\partial g_n(\gamma)}{\partial\beta'}\|) < \infty$ *for the parameter space* $\mathcal{A}$ *of* α *and a neighborhood* $\mathcal{N}_\beta$ *of* β_0*, then* $\hat{\alpha}_{\text{E-TGEL}_3}$ *and* $\hat{\alpha}_{\text{E-TGEL}_4}$*, where*

$$\hat{\alpha}_{\text{E-TGEL}_3} = \arg\min_{\alpha\in\mathcal{A}} \sup_{\lambda\in\Lambda_{ng}(\alpha,\hat{\beta}_{\text{E-TGEL}_3})} \sum_{i=1}^n \rho\Big(\lambda' g_{ni}(\alpha,\hat{\beta}_{\text{E-TGEL}_3})\Big)$$

and $\hat{\alpha}_{\text{E-TGEL}_4} = \arg\min_{\alpha\in\mathcal{A}}\sup_{\lambda\in\Lambda_{ng}(\alpha,\hat{\beta}_{\text{E-TGEL}_4})}\sum_{i=1}^n \rho(\lambda' g_{ni}(\alpha,\hat{\beta}_{\text{E-TGEL}_4}))$*, are asymptotically equivalent to the joint estimate* $\hat{\alpha}_{\text{GEL}}$ *of* α *from* (2.8).

The proper TGMM estimators to be compared with $\hat{\beta}_{\text{TGEL}_3}$ and $\hat{\beta}_{\text{TGEL}_4}$ are, respectively, $\hat{\beta}_{\text{TGMM}_3} = \arg\min_{\beta\in\mathcal{B}} g_n'(\check{\alpha},\beta)\tilde{C}_n'(\tilde{C}_n\tilde{\Omega}_n\tilde{C}_n')^{-1}\tilde{C}_n g_n(\check{\alpha},\beta)$ and

$$\hat{\beta}_{\text{TGMM}_4} = \arg\min_{\beta\in\mathcal{B}} g_n'(\dot{\alpha}_n(\beta),\beta)\check{C}_n'(\check{C}_n\check{\Omega}_n\check{C}_n')^{-1}\check{C}_n g_n(\dot{\alpha}_n(\beta),\beta),$$

where $\tilde{\Omega}_n = \Omega_n(\check{\alpha},\tilde{\beta})$ and $\check{\Omega}_n = \Omega_n(\dot{\alpha}_n(\check{\beta}),\check{\beta})$.

Let the leading order terms of $\sqrt{n}[\dot{\alpha}_n(\check{\beta}) - \alpha_0]$, $\sqrt{n}[C_n(\check{\alpha},\tilde{\beta}) - \bar{C}_n]$ and $\sqrt{n}[C_n(\dot{\alpha}_n(\check{\beta}),\check{\beta}) - \bar{C}_n]$ be, respectively, $\psi_{n\check{\alpha}}$, $\psi_{n\tilde{C}}$ and $\psi_{n\check{C}}$. Then, $\psi_{n\check{\alpha}} = \sqrt{n}(\dot{\alpha}_n(\beta_0) - \alpha_0) + \alpha_{n\beta}'\psi_{n\check{\beta}}$, $\psi_{n\tilde{C}} = \psi_{nC} + \sum_{j=1}^{k_\alpha}\bar{C}_n^{(j)}\psi_{n\check{\alpha}j} + \sum_{j=1}^{k_\beta}\bar{C}_n^{(k_\alpha+j)}\psi_{n\tilde{\beta}j}$, and $\psi_{n\check{C}} = \psi_{nC} + \sum_{j=1}^{k_\alpha}\bar{C}_n^{(j)}\psi_{n\check{\alpha}j} + \sum_{j=1}^{k_\beta}\bar{C}_n^{(k_\alpha+j)}\psi_{n\check{\beta}j}$. The higher order biases of $\hat{\beta}_{\text{TGEL}_3}$ and $\hat{\beta}_{\text{TGEL}_4}$ are investigated in the following theorem.

Theorem D.2. *Suppose that Assumptions 1–3, D.1 and D.2 are satisfied.*

(i) *The bias of* $\hat{\beta}_{\text{TGEL}_3}$ *is*

$$\text{Bias}(\hat{\beta}_{\text{TGEL}_3}) = [B_{nd}^I + (B_{nd}^\Omega + \frac{\rho_3(0)}{2}\tilde{B}_{nd}^\Omega) + (B_{nd}^G - \tilde{B}_{nd}^G)] + [B_{nd}^{C-g} + B_{nd}^{C-\Omega} + B_{nd}^{C-G} + B_{nd}^{\check{\alpha}}],$$

where B_{nd}^I, B_{nd}^Ω, $\tilde{B}_{nd}^\Omega$, B_{nd}^G, $\tilde{B}_{nd}^G$ *and* $B_{nd}^{\check{\alpha}}$ *are the same as corresponding ones for* $\hat{\beta}_{\text{TGEL}}$ *in Theorem 2,* $B_{nd}^{C-g} = -\frac{1}{n}\bar{H}_{nd}\,\mathbf{E}(\sqrt{n}\psi_{n\tilde{C}}g_n + \psi_{n\tilde{C}}\bar{G}_{n\alpha}\psi_{n\check{\alpha}} + \psi_{n\tilde{C}}\bar{G}_{n\beta}\psi_{n\beta})$,

$$B_{nd}^{C-\Omega} = -\frac{1}{n}\bar{H}_{nd}\,\mathbf{E}[(\psi_{n\tilde{C}}\bar{\Omega}_n\bar{C}_n' + \bar{C}_n\bar{\Omega}_n\psi_{n\tilde{C}}')\psi_{n\mu}],$$

and $B_{nd}^{C-G} = \frac{1}{n}\mathbf{E}(\bar{\Sigma}_{nd}\bar{G}_{n\beta}'\psi_{n\tilde{C}}'\psi_{n\mu})$.

(ii) *The bias of* $\hat{\beta}_{\text{TGMM3}}$ *is*

$$\text{Bias}(\hat{\beta}_{\text{TGMM3}}) = [B^I_{nd} + B^{\Omega}_{nd} + B^G_{nd} + B^W_{nd}] + [B^{C-g}_{nd} + B^{C-\Omega}_{nd} + B^{C-G}_{nd} + B^{\check{\alpha}}_{nd}] + B^{W-\check{\alpha}}_{nd},$$

where $B^W_{nd} = -\frac{1}{n}\bar{H}_{nd}\bar{C}_n \sum_{j=1}^{k_\beta}\sum_{i=1}^n \mathbf{E}(g_{ni}g_{ni}^{(k_\alpha+j)'} + g_{ni}^{(k_\alpha+j)}g'_{ni})\bar{C}'_n\bar{P}_{nd}\bar{C}_n\mathbf{E}(g_n g'_{nb})\bar{H}'_{nW\beta}e_{k_\beta j}$, $B^{W-\check{\alpha}}_{nd} = -\frac{1}{n^2}\bar{H}_{nd}\bar{C}_n\sum_{j=1}^{k_\beta}\sum_{i=1}^n \mathbf{E}(g_{ni}g_{ni}^{(k_\alpha+j)'} + g_{ni}^{(k_\alpha+j)}g'_{ni})\bar{C}'_n\bar{P}_{nd}\bar{C}_n\mathbf{E}(\sqrt{n}g_n\psi'_{n\check{\alpha}})\bar{G}'_{nb\alpha}\bar{H}'_{nW\beta}e_{k_\beta j}$, *and other terms are the same as corresponding ones in* (i).

(iii) *The bias of* $\hat{\beta}_{\text{TGEL4}}$ *is*

$$\text{Bias}(\hat{\beta}_{\text{TGEL4}}) = \widetilde{\text{Bias}}(\hat{\beta}_{\text{TGEL3}}) + [(B^{\dot{\alpha}-G_\alpha}_{nd} - \tilde{B}^{\dot{\alpha}-G_\alpha}_{nd}) + B^{\dot{\alpha}-G_\alpha-C}_{nd}],$$

where $B^{\dot{\alpha}-G_\alpha}_{nd} = -\bar{\Sigma}_{nd}\mathbf{E}(\alpha'_{n\beta}G'_{n\alpha}\bar{C}'_n\bar{P}_{nd}\bar{C}_n g_n)$, $\tilde{B}^{\dot{\alpha}-G_\alpha}_{nd} = -\frac{1}{n^2}\bar{\Sigma}_{nd}\sum_{i=1}^n\mathbf{E}(\alpha'_{n\beta}G'_{ni\alpha}\bar{C}'_n\bar{P}_{nd}\bar{C}_n g_{ni})$,

$$B^{\dot{\alpha}-G_\alpha-C}_{nd} = \frac{1}{n}\bar{\Sigma}_{nd}\alpha'_{n\beta}\mathbf{E}\Big[\Big(\sum_{j=1}^{k_\alpha}\bar{G}^{(j)'}_{n\alpha}\bar{C}'_n\psi_{n\dot{\alpha}j} + \bar{G}'_{n\alpha}\psi'_{n\check{C}}\Big)\psi_{n\mu}\Big],$$

and $\widetilde{\text{Bias}}(\hat{\beta}_{\text{TGEL3}})$ *has the same form as that of* $\text{Bias}(\hat{\beta}_{\text{TGEL3}})$ *in* (i) *except that* $\psi_{n\tilde{C}}$ *and* $\psi_{n\check{\alpha}}$ *in* $\text{Bias}(\hat{\beta}_{\text{TGEL3}})$ *are replaced by, respectively,* $\psi_{n\check{C}}$ *and* $\psi_{n\dot{\alpha}}$.

(iv) *The bias of* $\hat{\beta}_{\text{TGMM4}}$ *is*

$$\text{Bias}(\hat{\beta}_{\text{TGMM4}}) = \widetilde{\text{Bias}}(\hat{\beta}_{\text{TGMM3}}) + [B^{\dot{\alpha}-G_\alpha}_{nd} + B^{\dot{\alpha}-G_\alpha-C}_{nd}] + B^{W-\dot{\alpha}}_{nd},$$

where $B^{\dot{\alpha}-G_\alpha}_{nd}$ *and* $B^{\dot{\alpha}-G_\alpha-C}_{nd}$ *are in* (iii),

$$B^{W-\dot{\alpha}}_{nd} = -\frac{1}{n^2}\bar{H}_{nd}\bar{C}_n\sum_{j=1}^{k_\alpha}\sum_{i=1}^n\mathbf{E}(g_{ni}g_{ni}^{(j)'} + g_{ni}^{(j)}g'_{ni})\alpha_{n\beta}\mathbf{E}(\psi_{n\breve{\beta}j}\bar{C}'_n\psi_{n\mu})$$
$$-\frac{1}{n^2}\bar{H}_{nd}\bar{C}_n\sum_{j=1}^{k_\beta}\sum_{i=1}^n\mathbf{E}(g_{ni}g_{ni}^{(k_\alpha+j)'} + g_{ni}^{(k_\alpha+j)}g'_{ni})\bar{C}'_n\mathbf{E}(\psi_{n\mu}\psi_{n\breve{\beta}j}) - B^W_{nd} - B^{W\alpha}_{nd},$$

and $\widetilde{\text{Bias}}(\hat{\beta}_{\text{TGMM3}})$ *has the same form as that of* $\text{Bias}(\hat{\beta}_{\text{TGMM3}})$ *in* (ii) *except that* $\psi_{n\tilde{\beta}}$ *and* $\psi_{n\check{\alpha}}$ *in* $\text{Bias}(\hat{\beta}_{\text{TGMM3}})$ *are replaced by, respectively,* $\psi_{n\breve{\beta}}$ *and* $\psi_{n\dot{\alpha}}$.

E Higher order biases of TGMM estimators with unknown parameters in the transformation matrix

We shall compare the higher order bias of $\hat{\beta}_{\text{TGEL}}$ with that of $\hat{\beta}_{\text{TGMM}} = \arg\min_{\beta\in\mathcal{B}} d_n'(\check{\alpha},\beta)\Omega_{nd}^{-1}(\check{\alpha},\tilde{\beta})d_n(\check{\alpha},\beta)$, and also compare the higher order bias of $\hat{\beta}_{\text{TGEL2}}$ with that of

$$\hat{\beta}_{\text{TGMM2}} = \arg\min_{\beta\in\mathcal{B}} d_n'(\dot{\alpha}_n(\beta),\beta)\Omega_{nd}^{-1}(\dot{\alpha}_n(\breve{\beta}),\breve{\beta})d_n(\dot{\alpha}_n(\beta),\beta),$$

where $d_n(\gamma) = C_n(\gamma)g_n(\gamma)$, $\Omega_{nd}(\gamma) = C_n(\gamma)\Omega_n(\gamma)C_n'(\gamma)$, and $\tilde{\beta}$ and $\breve{\beta}$ are initial consistent estimators of β. Although $\tilde{\beta}$ and $\breve{\beta}$ can be any consistent estimators, we shall use $\tilde{\beta}$ and $\breve{\beta}$ defined in, respectively, (D.1) and (D.2).

Denote $\bar{H}_{nW\beta} = (\bar{G}_{nb\beta}'\bar{W}_n^{-1}\bar{G}_{nb\beta})^{-1}\bar{G}_{nb\beta}'\bar{W}_n^{-1}$. In addition, let the leading order terms of $\sqrt{n}(\tilde{\beta}-\beta_0)$ and $\sqrt{n}(\breve{\beta}-\beta_0)$ be, respectively, $\psi_{n\tilde{\beta}}$ and $\psi_{n\breve{\beta}}$. Then, $\psi_{n\tilde{\beta}} = -\bar{H}_{nW\beta}(\sqrt{n}g_{nb}+\bar{G}_{nb\alpha}\psi_{n\check{\alpha}})$ and $\psi_{n\breve{\beta}} = -[(\bar{G}_{nb\alpha}\alpha_{n\beta}+\bar{G}_{nb\beta})'\bar{W}_n^{-1}(\bar{G}_{nb\alpha}\alpha_{n\beta}+\bar{G}_{nb\beta})]^{-1}(\bar{G}_{nb\alpha}\alpha_{n\beta}+\bar{G}_{nb\beta})'\bar{W}_n^{-1}[\sqrt{n}g_{nb}+\bar{G}_{nb\alpha}\sqrt{n}(\dot{\alpha}_n(\beta_0)-\alpha_0)]$.

Theorem E.1. *Suppose that Assumptions 1–3 and D.1 are satisfied.*

(i) The bias of $\hat{\beta}_{\text{TGMM}}$ is

$$\text{Bias}(\hat{\beta}_{\text{TGMM}}) = [B_{nd}^{I}+B_{nd}^{\Omega}+B_{nd}^{G}+B_{nd}^{W}]+[B_{nd}^{C-\beta}+B_{nd}^{C-g}+B_{nd}^{C-\Omega}+B_{nd}^{C-G}+B_{nd}^{\check{\alpha}}]+B_{nd}^{C-\beta-g}+B_{nd}^{C-\beta-W}+B_{nd}^{W-\check{\alpha}},$$

where $B_{nd}^{W} = -\frac{1}{n}\bar{H}_{nd}\bar{C}_n\sum_{j=1}^{k_\beta}\sum_{i=1}^{n}\mathbf{E}(g_{ni}g_{ni}^{(k_\alpha+j)'}+g_{ni}^{(k_\alpha+j)}g_{ni}')\bar{C}_n'\bar{P}_{nd}\bar{C}_n\mathbf{E}(g_ng_{nb}')\bar{H}_{nW\beta}'e_{k_\beta j}$, $B_{nd}^{C-\beta-g} = -\frac{1}{n}\bar{\Sigma}_{nd}[\text{tr}(\bar{C}_n^{(k_\alpha+1)}\bar{\Omega}_n\bar{C}_n'\bar{P}_{nd}),\ldots,\text{tr}(\bar{C}_n^{(k_\alpha+k_\beta)}\bar{\Omega}_n\bar{C}_n'\bar{P}_{nd})]'$,

$$B_{nd}^{C-\beta-W} = -\frac{1}{n}\bar{H}_{nd}\sum_{j=1}^{k_\beta}(\bar{C}_n^{(k_\alpha+j)}\bar{\Omega}_n\bar{C}_n'+\bar{C}_n\bar{\Omega}_n\bar{C}_n^{(k_\alpha+j)'})\mathbf{E}(\psi_{n\mu}\psi_{n\tilde{\beta}j}),$$

$B_{nd}^{W-\check{\alpha}} = -\frac{1}{n^2}\bar{H}_{nd}\bar{C}_n\sum_{j=1}^{k_\beta}\sum_{i=1}^{n}\mathbf{E}(g_{ni}g_{ni}^{(k_\alpha+j)'}+g_{ni}^{(k_\alpha+j)}g_{ni}')\bar{C}_n'\bar{P}_{nd}\bar{C}_n\mathbf{E}(\sqrt{n}g_n\psi_{n\check{\alpha}}')\bar{G}_{nb\alpha}'\bar{H}_{nW\beta}'e_{k_\beta j}$, *and other terms are the same as corresponding ones for $\hat{\beta}_{\text{TGEL}}$ in Theorem 2.*

(ii) The bias of $\hat{\beta}_{\text{TGMM2}}$ is

$$\text{Bias}(\hat{\beta}_{\text{TGMM2}}) = \widetilde{\text{Bias}}(\hat{\beta}_{\text{TGMM}}) + [B_{nd}^{C-\dot{\alpha}}+B_{nd}^{\dot{\alpha}-G_\alpha}+B_{nd}^{\dot{\alpha}-G_\alpha-C}]+[B_{nd}^{C-\dot{\alpha}-g}+B_{nd}^{W-\dot{\alpha}}+B_{nd}^{C-\Omega-\dot{\alpha}}],$$

where $B_{nd}^{C-\dot{\alpha}}$, $B_{nd}^{\dot{\alpha}-G_\alpha}$ *and* $B_{nd}^{\dot{\alpha}-G_\alpha-C}$ *are the same as corresponding ones for* $\hat{\beta}_{\text{TGEL2}}$ *in Theorem 2,* $B_{nd}^{C-\dot{\alpha}-g} = -\frac{1}{n}\bar{\Sigma}_{nd}\Big[\text{tr}[(\sum_{k=1}^{k_\alpha}\bar{C}_n^{(k)}\alpha_{nk}^{(1)})'\bar{P}_{nd}\bar{C}_n\bar{\Omega}_n],\ldots,\text{tr}[(\sum_{k=1}^{k_\alpha}\bar{C}_n^{(k)}\alpha_{nk}^{(k_\beta)})'\bar{P}_{nd}\bar{C}_n\bar{\Omega}_n]\Big]'$,

$$B_{nd}^{W-\dot{\alpha}} = -\frac{1}{n^2}\bar{H}_{nd}\bar{C}_n\sum_{j=1}^{k_\alpha}\sum_{i=1}^{n}\mathbf{E}(g_{ni}g_{ni}^{(j)'} + g_{ni}^{(j)}g_{ni}')\alpha_{n\beta}\,\mathbf{E}(\psi_{n\breve{\beta}j}\bar{C}_n'\psi_{n\mu})$$
$$-\frac{1}{n^2}\bar{H}_{nd}\bar{C}_n\sum_{j=1}^{k_\beta}\sum_{i=1}^{n}\mathbf{E}(g_{ni}g_{ni}^{(k_\alpha+j)'} + g_{ni}^{(k_\alpha+j)}g_{ni}')\bar{C}_n'\,\mathbf{E}(\psi_{n\mu}\psi_{n\breve{\beta}j}) - B_{nd}^{W} - B_{nd}^{W-\check{\alpha}},$$

$B_{nd}^{C-\Omega-\dot{\alpha}} = -\frac{1}{n}\bar{H}_{nd}\sum_{j=1}^{k_\alpha}(\bar{C}_n^{(j)}\bar{\Omega}_n\bar{C}_n' + \bar{C}_n\bar{\Omega}_n\bar{C}_n^{(j)'})\mathbf{E}[\psi_{n\mu}(\psi_{n\check{\alpha}j} - \psi_{n\dot{\alpha}j})]$, *and* $\widetilde{\text{Bias}}(\hat{\beta}_{\text{TGMM}})$ *has the same form as that of* $\text{Bias}(\hat{\beta}_{\text{TGMM}})$ *in (i) except that* $\psi_{n\check{\alpha}}$ *in* $\text{Bias}(\hat{\beta}_{\text{TGMM}})$ *is replaced by* $\psi_{n\dot{\alpha}}$*.*

Compare to the TGEL estimator $\hat{\beta}_{\text{TGEL}}$, $\hat{\beta}_{\text{TGMM}}$ has the extra bias term B_{nd}^{W} from the choice of the initial GMM estimator of β_0. In addition, $\hat{\beta}_{\text{TGMM}}$ has extra bias terms $B_{nd}^{C-\beta-g}$, $B_{nd}^{C-\beta-W}$ and $B_{nd}^{W-\check{\alpha}}$, which are due to, respectively, the correlation of g_n with $\bar{C}_n^{(k_\alpha+j)}g_n$, the two-step nature of GMM in the presence of unknown β in $C_n(\alpha,\beta)$, and the possible use of $\check{\alpha}$ in deriving the initial GMM estimate $\tilde{\beta}$. The $\hat{\beta}_{\text{TGEL}}$ automatically eliminates the bias term $B_{nd}^{C-\beta-g}$ and it does not have the bias term $B_{nd}^{W-\check{\alpha}}$ since it estimates β in one step.

The bias terms $B_{nd}^{C-\dot{\alpha}-g}$, $B_{nd}^{W-\dot{\alpha}}$ and $B_{nd}^{C-\Omega-\dot{\alpha}}$ for $\hat{\beta}_{\text{TGMM2}}$ are due to the derivative of $\dot{\alpha}(\beta)$. The $B_{nd}^{C-\dot{\alpha}-g}$ is related to the correlation of $\bar{C}_n^{(j)}\alpha_{nj}^{(k)}g_n$ with g_n, $B_{nd}^{W-\dot{\alpha}}$ is the difference of $\text{Bias}(\hat{\beta}_{\text{TGMM2}})$ and $\text{Bias}(\hat{\beta}_{\text{TGMM}})$ due to the two-step nature of GMM, and $B_{nd}^{C-\Omega-\dot{\alpha}}$ is the difference of $\text{Bias}(\hat{\beta}_{\text{TGMM2}})$ and $\text{Bias}(\hat{\beta}_{\text{TGMM}})$ due to the estimation of $\bar{C}_n$ in $\bar{C}_n\bar{\Omega}_n\bar{C}_n'$.

F Efficient estimation of the transformation matrix

When we construct the TGEL estimators, we assume the existence of the estimating function $C_n(\gamma)$ for the transformation matrix, but have not discussed explicitly how expectations in $\bar{C}_n$ are estimated. It is possible to have a relatively efficient estimate $C_n(\gamma)$ and we discuss it in this section.

Define the following implied probabilities from the two-step CU estimator (e.g., Newey and Smith, 2004):

$$\dot{\pi}_{ni}(\gamma) = \frac{1}{n} - \frac{1}{n}[d_{ni}(\gamma) - d_n(\gamma)]'V_n^{-1}(\gamma)d_n(\gamma), \tag{F.1}$$

for $i = 1,\ldots,n$, where $V_n(\gamma) = \frac{1}{n}\sum_{i=1}^n [d_{ni}(\gamma) - d_n(\gamma)][d_{ni}(\gamma) - d_n(\gamma)]'$.[2] We may use $\dot{\pi}_{ni}(\gamma)$'s to efficiently estimate expectations. As an illustration, consider a function $h(z)$ of a data vector z, and the estimate of its expectation $\mathbf{E}[h(z)]$ by

$$\sum_{i=1}^n \dot{\pi}_{ni}(\gamma_0)h(z_{ni}) = \frac{1}{n}\sum_{i=1}^n h(z_{ni}) - \frac{1}{n}\sum_{i=1}^n h(z_{ni})[d_{ni}(\gamma_0) - d_n(\gamma_0)]'V_n^{-1}(\gamma_0)d_n(\gamma_0), \tag{F.2}$$

where z_i's are i.i.d. observations on z. This estimate of $\mathbf{E}[h(z)]$ with implied probabilities is the empirical counterpart of

$$a_n = \frac{1}{n}\sum_{i=1}^n h(z_{ni}) - \mathrm{cov}[h(z), \bar{C}_n g_n(\gamma_0)]\mathrm{var}^{-1}[\bar{C}_n g_n(\gamma_0)]\bar{C}_n g_n(\gamma_0).$$

The a_n is not correlated with $\bar{C}_n g_n(\gamma_0)$ since it is the minimum variance estimator of $\mathbf{E}[h(z)]$ among estimators with the form $\frac{1}{n}\sum_{i=1}^n h(z_{ni}) - t'\bar{C}_n g_n(\gamma_0)$ for some $m_b \times 1$ vector t (Antoine et al., 2007).[3] Thus, we may use the implied probabilities to efficiently estimate $\bar{C}_n$. For example, when $m_a = k_\alpha$, we may let $C_n(\gamma) = [\sum_{i=1}^n \dot{\pi}_{ni}(\gamma)\frac{\partial g_{nbi}(\gamma)}{\partial \alpha'}][\sum_{i=1}^n \dot{\pi}_{ni}(\gamma)\frac{\partial g_{nai}(\gamma)}{\partial \alpha'}]^{-1}$. With such a $C_n(\gamma)$ for $\hat{\beta}_{\text{TGEL}}$ and $\hat{\beta}_{\text{TGEL2}}$, $C_n(\check{\alpha}, \hat{\beta}_{\text{TGEL}})$ is uncorrelated with $\sqrt{n}d_n(\check{\alpha}, \hat{\beta}_{\text{TGEL}})$, and $C_n(\dot{\alpha}_n(\hat{\beta}_{\text{TGEL2}}), \hat{\beta}_{\text{TGEL2}})$ is uncorrelated with $\sqrt{n}d_n(\dot{\alpha}_n(\hat{\beta}_{\text{TGEL2}}), \hat{\beta}_{\text{TGEL2}})$. For $\hat{\beta}_{\text{TGEL3}}$ and $\hat{\beta}_{\text{TGEL4}}$, $C_n(\check{\alpha}, \beta)$ and $C_n(\dot{\alpha}_n(\beta), \beta)$ need to be evaluated at an estimator of β that is asymptotically as efficient as the joint GMM estimator. The implied probabilities in (F.1) can be negative. Antoine et al. (2007) define nonnegative shrunk implied probabilities, which do not lose their asymptotic efficiency nature. The nonnegative shrunk probabilities are $\ddot{\pi}_{ni}(\gamma) = \frac{1}{1+\delta_n(\gamma)}\dot{\pi}_{ni}(\gamma) + \frac{\delta_n(\gamma)}{1+\delta_n(\gamma)}\frac{1}{n}$, where $\delta_n(\gamma) = -n\min\{\min_{1\le i\le n}\dot{\pi}_{ni}(\gamma), 0\}$.

Our Monte Carlo results show that using CU implied probabilities or CU shrunk probabilities for the transformation matrix does not lead to significant improvement in the finite sample performance of TGEL estimators, so we have not considered them in the main text.

[2]The GEL estimator $\hat{\gamma}_{\text{GEL}} = \arg\min_{\gamma\in\Gamma}\sup_{\lambda\in\Lambda_{ng}(\gamma)}\sum_{i=1}^n \rho(\lambda' g_{ni}(\gamma))$ is equivalent to the minimum discrepancy estimator $\arg\min_{\gamma\in\Gamma,\pi_1,\ldots,\pi_n}\sum_{i=1}^n h(\pi_i)$, subject to $\sum_{i=1}^n \pi_i g_{ni}(\gamma) = 0$ and $\sum_{i=1}^n \pi_i = 1$ for some convex function $h(\pi)$ of a scalar π. For EL, $h(\pi) = -\ln(\pi)$; for ET, $h(\pi) = \pi\ln(\pi)$; for CU, $h(\pi) = \pi^2$. The implied probabilities π_i's are $\rho_1(\hat{\lambda}'_{\text{GEL}} g_{ni}(\hat{\gamma}_{\text{GEL}}))/[\sum_{j=1}^n \rho_1(\hat{\lambda}'_{\text{GEL}} g_{nj}(\hat{\gamma}_{\text{GEL}}))]$. Thus we can show the CU implied probabilities have closed forms.

[3]It is not so if $g_{ni}(\gamma)$'s are not i.i.d.

G More tests

G.1 Tests in the TGMM framework

For comparison purposes, we consider several tests in the TGMM framework. The TGMM distance difference test statistic is

$$\mathcal{R}_{\text{TGMM}} = n[d_n'(\check{\alpha},\hat{\beta}_{\text{RTGMM}})\Omega_{nd}^{-1}(\check{\alpha},\hat{\beta}_{\text{RTGMM}})d_n(\check{\alpha},\hat{\beta}_{\text{RTGMM}}) - d_n'(\check{\alpha},\hat{\beta}_{\text{TGMM}})\Omega_{nd}^{-1}(\check{\alpha},\hat{\beta}_{\text{TGMM}})d_n(\check{\alpha},\hat{\beta}_{\text{TGMM}})], \tag{G.1}$$

where $\hat{\beta}_{\text{RTGMM}}$ is the restricted GMM estimator that solves $\min_{\beta\in\mathcal{B}} d_n'(\check{\alpha},\beta)\Omega_{nd}^{-1}(\check{\alpha},\tilde{\beta})d_n(\check{\alpha},\beta)$ subject to $r(\beta)=0$. The Wald test statistic $\mathcal{W}_{\text{TGMM}}$ with the TGMM estimate $\hat{\beta}_{\text{TGMM}}$ has the same form as that of $\mathcal{W}_{\text{TGEL}}$ in the main text, with $\hat{\beta}_{\text{TGEL}}$ in $\mathcal{W}_{\text{TGEL}}$ replaced by $\hat{\beta}_{\text{TGMM}}$.

The gradient test statistic in the TGMM framework has the form

$$\begin{aligned}\mathcal{G}_{\text{TGMM}} = {} & nd_n'(\check{\alpha},\hat{\beta}_{\text{RTGMM}})\Omega_{nd}^{-1}(\check{\alpha},\hat{\beta}_{\text{RTGMM}})D_{n\beta}(\check{\alpha},\hat{\beta}_{\text{RTGMM}})\Sigma_{nd}(\check{\alpha},\hat{\beta}_{\text{RTGMM}}) \\ & \cdot D_{n\beta}'(\check{\alpha},\hat{\beta}_{\text{RTGMM}})\Omega_{nd}^{-1}(\check{\alpha},\hat{\beta}_{\text{RTGMM}})d_n(\check{\alpha},\hat{\beta}_{\text{RTGMM}}),\end{aligned} \tag{G.2}$$

which requires the use of $\hat{\beta}_{\text{RTGMM}}$.

Let $\Psi_n(\gamma)$ and $\Psi_{ni}(\gamma)$ be as given in Section 3 of the main text. Since $\frac{1}{n}\sum_{i=1}^n \Psi_{ni}(\check{\gamma}_r)\Psi_{ni}'(\check{\gamma}_r)$ is a consistent estimator of the limiting variance of $\sqrt{n}\Psi_n(\gamma_0)$, in the form of an outer-product-of-gradients (OPG) test, as in Davidson and MacKinnon (2004), the $C(\alpha)$-type gradient test statistic is

$$\mathcal{OPG}_{\text{T}} = \Big(\sum_{i=1}^n \Psi_{ni}(\check{\alpha},\check{\beta}_r)\Big)'\Big(\sum_{i=1}^n \Psi_{ni}(\check{\alpha},\check{\beta}_r)\Psi_{ni}'(\check{\alpha},\check{\beta}_r)\Big)^{-1}\Big(\sum_{i=1}^n \Psi_{ni}(\check{\alpha},\check{\beta}_r)\Big). \tag{G.3}$$

Another advantage of the $C(\alpha)$-type gradient test is its robustness to unknown heteroskedasticity, because $\Omega_n(\gamma_0) = \frac{1}{n}\sum_{i=1}^n g_{ni}(\gamma_0)g_{ni}'(\gamma_0)$ may capture unknown heteroskedasticity in $g_{ni}(\gamma_0)$ (Lee and Yu, 2012).

Theorem G.1. *Suppose that Assumptions 1–2 and 4 hold. $\mathcal{R}_{\text{TGMM}}$, $\mathcal{W}_{\text{TGMM}}$, $\mathcal{G}_{\text{TGMM}}$ and $\mathcal{OPG}_{\text{T}}$ are all asymptotically equivalent with the asymptotic distribution $\chi^2(k_r, \lim_{n\to\infty} c'R'(R\bar{\Sigma}_{nd}R')^{-1}Rc)$.*

G.2 Tests in the ordinary GEL and GMM frameworks

Let $\boldsymbol{\rho}_n(\gamma,\lambda)=\sum_{i=1}^n \rho(\lambda' g_{ni}(\gamma))$, $\hat{\gamma}_{\text{rGEL}}$ be the restricted GEL estimator that solves

$$\min_{\gamma\in\Gamma}\sup_{\lambda\in\Lambda_{ng}(\gamma)}\boldsymbol{\rho}_n(\gamma,\lambda)\quad \text{s.t.}\quad r(\beta)=0,$$

and $\hat{\lambda}_{\text{rGEL}}=\arg\max_{\lambda\in\Lambda_{ng}(\hat{\gamma}_{\text{rGEL}})}\boldsymbol{\rho}_n(\hat{\gamma}_{\text{rGEL}},\lambda)$. Then the GEL ratio test is

$$\mathcal{R}_{\text{GEL}}=2[\boldsymbol{\rho}_n(\hat{\gamma}_{\text{rGEL}},\hat{\lambda}_{\text{rGEL}})-\boldsymbol{\rho}_n(\hat{\gamma}_{\text{GEL}},\hat{\lambda}_{\text{GEL}})],\tag{G.4}$$

and the corresponding GMM distance difference test statistic is

$$\mathcal{R}_{\text{GMM}}=n[g_n'(\hat{\gamma}_{\text{rGMM}})\Omega_n^{-1}(\hat{\gamma}_{\text{rGMM}})g_n(\hat{\gamma}_{\text{rGMM}})-g_n'(\hat{\gamma}_{\text{GMM}})\Omega_n^{-1}(\hat{\gamma}_{\text{GMM}})g_n(\hat{\gamma}_{\text{GMM}})].\tag{G.5}$$

The Wald tests with the GEL and GMM estimates are, respectively,

$$\mathcal{W}_{\text{GEL}}=n\cdot r'(\hat{\beta}_{\text{GEL}})[R_\gamma(\hat{\beta}_{\text{GEL}})\Sigma_n(\check{\alpha},\hat{\beta}_{\text{GEL}})R_\gamma'(\hat{\beta}_{\text{GEL}})]^{-1}r(\hat{\beta}_{\text{GEL}}),\tag{G.6}$$

and

$$\mathcal{W}_{\text{GMM}}=n\cdot r'(\hat{\beta}_{\text{GMM}})[R_\gamma(\hat{\beta}_{\text{GMM}})\Sigma_n(\check{\alpha},\hat{\beta}_{\text{GMM}})R_\gamma'(\hat{\beta}_{\text{GMM}})]^{-1}r(\hat{\beta}_{\text{GMM}}),\tag{G.7}$$

where $R_\gamma=\frac{\partial r(\beta)}{\partial\gamma'}$ and $\Sigma_n(\gamma)=[G_n'(\gamma)\Omega_n^{-1}(\gamma)G_n(\gamma)]^{-1}$. The score-type test with the restricted GEL estimator has the test statistic

$$\mathcal{S}_{\text{GEL}}=\frac{1}{n}\frac{\partial\boldsymbol{\rho}_n(\hat{\gamma}_{\text{rGEL}},\hat{\lambda}_{\text{rGEL}})}{\partial\gamma'}\Sigma_n(\hat{\gamma}_{\text{rGEL}})\frac{\partial\boldsymbol{\rho}_n(\hat{\gamma}_{\text{rGEL}},\hat{\lambda}_{\text{rGEL}})}{\partial\gamma}.\tag{G.8}$$

Let $\hat{\gamma}_{\text{rGMM}}$ be the restricted OGMM estimator of γ that solves $\min_{\gamma\in\Gamma}g_n'(\gamma)\Omega_n^{-1}(\tilde{\gamma}_{\text{GMM}})g_n(\gamma)$ subject to $r(\beta)=0$, where $\tilde{\gamma}_{\text{GMM}}$ is an initial consistent estimator of γ_0. Then the gradient test statistic in the GMM framework is

$$\mathcal{G}_{\text{GMM}}=ng_n'(\hat{\gamma}_{\text{rGMM}})\Omega_n^{-1}(\hat{\gamma}_{\text{rGMM}})G_n(\hat{\gamma}_{\text{rGMM}})\Sigma_n(\hat{\gamma}_{\text{rGMM}})G_n'(\hat{\gamma}_{\text{rGMM}})\Omega_n^{-1}(\hat{\gamma}_{\text{rGMM}})g_n(\hat{\gamma}_{\text{rGMM}}),\tag{G.9}$$

where $G_n(\gamma)=\frac{\partial g_n(\gamma)}{\partial\gamma'}$. For any $\sqrt{n}$-consistent restricted estimate $\check{\gamma}_r$ of γ such that $r(\beta)=0$,

$$\hat{\Upsilon}_n(\gamma)=R_\gamma(\check{\gamma}_r)\Sigma_n(\check{\gamma}_r)G_n'(\check{\gamma}_r)\Omega_n^{-1}(\check{\gamma}_r)g_n(\gamma)$$

is a $C(\alpha)$-type statistic, which satisfies $\sqrt{n}\hat{\Upsilon}_n(\check{\gamma}_r)=\sqrt{n}\hat{\Upsilon}_n(\gamma_0)+o_p(1)$. Let $\Upsilon_{ni}(\gamma)$ be the vector obtained by replacing $g_n(\gamma)$ in the above $\hat{\Upsilon}_n(\gamma)$ by $g_{ni}(\gamma)$. Then the OPG test is

$$\mathcal{OPG}=\Big(\sum_{i=1}^n\Upsilon_{ni}(\check{\gamma}_r)\Big)'\Big(\sum_{i=1}^n\Upsilon_{ni}(\check{\gamma}_r)\Upsilon_{ni}'(\check{\gamma}_r)\Big)^{-1}\Big(\sum_{i=1}^n\Upsilon_{ni}(\check{\gamma}_r)\Big).\tag{G.10}$$

In the GEL framework, we may also implement the gradient test. Let

$$\check{\lambda}_r = \arg \max_{\lambda \in \Lambda_{n\Upsilon}(\check{\gamma}_r)} \sum_{i=1}^{n} \rho\Big(\lambda' \Upsilon_{ni}(\check{\gamma}_r)\Big),$$

where $\Lambda_{n\Upsilon}(\check{\gamma}_r) = \{\lambda : \lambda'\Upsilon_{ni}(\check{\gamma}_r) \in \mathcal{V}, i = 1, \ldots, n\}$. Then we have the following GEL gradient test statistic

$$\mathcal{G}_{\text{GEL}} = 2\Big[\sum_{i=1}^{n} \rho\Big(\check{\lambda}_r' \Upsilon_{ni}(\check{\gamma}_r)\Big) - n\rho(0)\Big]. \tag{G.11}$$

Theorem G.2. *Suppose that* $\sqrt{n}(\check{\gamma}_r - \gamma_0) = O_p(1)$ *and Assumptions 1, 4 and D.1 hold. Then* $\mathcal{R}_{\text{GEL}}$, $\mathcal{R}_{\text{GMM}}$, $\mathcal{W}_{\text{GEL}}$, $\mathcal{W}_{\text{GMM}}$, $\mathcal{S}_{\text{GEL}}$, $\mathcal{G}_{\text{GMM}}$, $\mathcal{OPG}$, $\mathcal{G}_{\text{GEL}}$ *are all asymptotically equivalent and they have the asymptotic distribution* $\chi^2(k_r, \lim_{n\to\infty} c'R'(R_\gamma \bar{\Sigma}_n R_\gamma')^{-1} Rc)$*, where* $\bar{\Sigma}_n = (\bar{G}_n' \bar{\Omega}_n^{-1} \bar{G}_n)^{-1}$ *and* $R_\gamma = \frac{\partial r(\beta_0)}{\partial \gamma'} = [0, R]$.

From the above theorem, under the null (as $c = 0$), all the test statistics considered follow the asymptotic central chi-squared distribution with k_r degrees of freedom. If $c \neq 0$, then the above tests have the same local power asymptotically. We omit the proof of Theorem G.2, since it is similar to that of Theorem 3 in the main text. The only difference is on the use of different moment functions. The set of moment functions in Theorem G.2 consists of the original ones.

H Lemmas

Lemma H.1. *(i) Under Assumptions 1(ii)–(v), 2(iii)–(iv) and D.1(i)–(ii),* $\tilde{\beta} = \beta_0 + o_p(1)$ *and* $\check{\beta} = \beta_0 + o_p(1)$.

(ii) Under Assumptions 1(ii)–(v), (viii)–(ix), (xi), 2(iii)–(iv), 3(i)–(ii), and D.1, $\sqrt{n}(\tilde{\beta} - \beta_0) = \psi_{n\tilde{\beta}} + O_p(n^{-1/2})$*, where* $\psi_{n\tilde{\beta}} = -\bar{H}_{nW\beta}(\sqrt{n} g_{nb} + \bar{G}_{nb\alpha} \psi_{n\check{\alpha}})$ *with* $\bar{H}_{nW\beta} = (\bar{G}_{nb\beta}' \bar{W}_n^{-1} \bar{G}_{nb\beta})^{-1} \bar{G}_{nb\beta}' \bar{W}_n^{-1}$*, and* $\sqrt{n}(\check{\beta} - \beta_0) = \psi_{n\check{\beta}} + O_p(n^{-1/2})$*, where* $\psi_{n\check{\beta}} = -[(\bar{G}_{nb\alpha} \frac{\partial \alpha_n(\beta_0)}{\partial \beta'} + \bar{G}_{nb\beta})' \bar{W}_n^{-1} (\bar{G}_{nb\alpha} \frac{\partial \alpha_n(\beta_0)}{\partial \beta'} + \bar{G}_{nb\beta})]^{-1} (\bar{G}_{nb\alpha} \frac{\partial \alpha_n(\beta_0)}{\partial \beta'} + \bar{G}_{nb\beta})' \bar{W}_n^{-1} [\sqrt{n} g_{nb} + \bar{G}_{nb\alpha} \sqrt{n}(\dot{\alpha}_n(\beta_0) - \alpha_0)]$.

(iii) Let $\hat{\beta} = \beta_0 + n^{-1/2} \psi_{n\beta} + O_p(n^{-1})$*, where* $\psi_{n\beta} = O_p(1)$*. Under Assumptions 2(iv) and 3(i),* $\sqrt{n}(\dot{\alpha}_n(\hat{\beta}) - \alpha_0) = \sqrt{n}(\dot{\alpha}_n(\beta_0) - \alpha_0) + \frac{\partial \alpha_n(\beta_0)}{\partial \beta'} \psi_{n\beta} + O_p(n^{-1/2})$.

Proof. (i) Let $\bar{g}_n(\alpha, \beta) = \mathbf{E}[g_n(\alpha, \beta)]$. By the triangle inequality,

$$\sup_{\beta \in \mathcal{B}} \|g_n(\check{\alpha}, \beta) - \bar{g}_n(\alpha_0, \beta)\| \le \sup_{\beta \in \mathcal{B}} \|g_n(\check{\alpha}, \beta) - \bar{g}_n(\check{\alpha}, \beta)\| + \sup_{\beta \in \mathcal{B}} \|\bar{g}_n(\check{\alpha}, \beta) - \bar{g}_n(\alpha_0, \beta)\|$$

$$\leq \sup_{\gamma\in\Gamma}\|g_n(\gamma)-\bar{g}_n(\gamma)\|+\sup_{\beta\in\mathcal{B}}\|\bar{g}_n(\check{\alpha},\beta)-\bar{g}_n(\alpha_0,\beta)\|.$$

The first term in the last line is $o_p(1)$ under Assumption 1(iv), and the second term is $o_p(1)$ since $\check{\alpha}=\alpha_0+o_p(1)$ and $\bar{g}_n(\gamma)$ is continuous on Γ uniformly in n. Thus, $\sup_{\beta\in\mathcal{B}}\|g_n(\check{\alpha},\beta)-\bar{g}_n(\alpha_0,\beta)\|=o_p(1)$. Under Assumption 1($v$), $\sup_{\beta\in\mathcal{B}}\|\bar{g}_n(\alpha_0,\beta)\|=O_p(1)$. In particular, $\sup_{\beta\in\mathcal{B}}\|g_{nb}(\check{\alpha},\beta)-\bar{g}_{nb}(\alpha_0,\beta)\|=o_p(1)$ and $\sup_{\beta\in\mathcal{B}}\|\bar{g}_{nb}(\alpha_0,\beta)\|=O_p(1)$, where $\bar{g}_{nb}(\gamma)=\mathbf{E}[g_{nb}(\gamma)]$. Therefore, under the identification condition in Assumption D.1, $\tilde{\beta}=\beta_0+o_p(1)$ by Theorem 3.4 in White (1994).

For $\breve{\beta}$, by the triangle inequality,

$$\begin{aligned}
&\sup_{\beta\in\mathcal{B}}\|g_n(\dot{\alpha}_n(\beta),\beta)-\bar{g}_n(\alpha_n(\beta),\beta)\|\\
&\leq \sup_{\beta\in\mathcal{B}}\|g_n(\dot{\alpha}_n(\beta),\beta)-\bar{g}_n(\dot{\alpha}_n(\beta),\beta)\|+\sup_{\beta\in\mathcal{B}}\|\bar{g}_n(\dot{\alpha}_n(\beta),\beta)-\bar{g}_n(\alpha_n(\beta),\beta)\|\\
&\leq \sup_{\gamma\in\Gamma}\|g_n(\gamma)-\bar{g}_n(\gamma)\|+\sup_{\beta\in\mathcal{B}}\|\bar{g}_n(\dot{\alpha}_n(\beta),\beta)-\bar{g}_n(\alpha_n(\beta),\beta)\|\cdot I\Big(\sup_{\beta\in\mathcal{B}}\|\dot{\alpha}_n(\beta)-\alpha_n(\beta)\|\geq c\Big)\\
&\quad+\sup_{\beta\in\mathcal{B}}\|\bar{g}_n(\dot{\alpha}_n(\beta),\beta)-\bar{g}_n(\alpha_n(\beta),\beta)\|\cdot I\Big(\sup_{\beta\in\mathcal{B}}\|\dot{\alpha}_n(\beta)-\alpha_n(\beta)\|< c\Big).
\end{aligned}\tag{H.1}$$

As Γ is compact and $\bar{g}_n(\gamma)$ is equicontinuous, $\bar{g}_n(\gamma)$ is uniformly equicontinuous on Γ. Then for any $c_1>0$, there exists some $c_2>0$, such that $\|\bar{g}_n(\dot{\alpha}_n(\beta),\beta)-\bar{g}_n(\alpha_n(\beta),\beta)\|\leq\frac{c_1}{2}$ for any n and any β satisfying $\|\dot{\alpha}_n(\beta)-\alpha_n(\beta)\|\leq c_2$. Thus,

$$\begin{aligned}
&\mathrm{P}\Big(\sup_{\beta\in\mathcal{B}}\|\bar{g}_n(\dot{\alpha}_n(\beta),\beta)-\bar{g}_n(\alpha_n(\beta),\beta)\|\geq c_1\Big)\\
&=\mathrm{P}\Big(\sup_{\beta\in\mathcal{B}}\Big[\|\bar{g}_n(\dot{\alpha}_n(\beta),\beta)-\bar{g}_n(\alpha_n(\beta),\beta)\|\cdot I(\|\dot{\alpha}_n(\beta)-\alpha_n(\beta)\|>c_2)\\
&\qquad+\|\bar{g}_n(\dot{\alpha}_n(\beta),\beta)-\bar{g}_n(\alpha_n(\beta),\beta)\|\cdot I(\|\dot{\alpha}_n(\beta)-\alpha_n(\beta)\|\leq c_2)\Big]\geq c_1\Big)\\
&\leq \mathrm{P}\Big(\sup_{\beta\in\mathcal{B}}[\|\bar{g}_n(\dot{\alpha}_n(\beta),\beta)-\bar{g}_n(\alpha_n(\beta),\beta)\|\cdot I(\|\dot{\alpha}_n(\beta)-\alpha_n(\beta)\|\geq c_2)]\geq\frac{c_1}{2}\Big)\\
&\leq \mathrm{P}\Big(\sup_{\beta\in\mathcal{B}}\|\dot{\alpha}_n(\beta)-\alpha_n(\beta)\|\geq c_2\Big).
\end{aligned}$$

As $\sup_{\beta\in\mathcal{B}}\|\dot{\alpha}_n(\beta)-\alpha_n(\beta)\|=o_p(1)$, $\sup_{\beta\in\mathcal{B}}\|\bar{g}_n(\dot{\alpha}_n(\beta),\beta)-\bar{g}_n(\alpha_n(\beta),\beta)\|=o_p(1)$. Then, by (H.1), $\sup_{\beta\in\mathcal{B}}\|g_n(\dot{\alpha}_n(\beta),\beta)-\bar{g}_n(\alpha_n(\beta),\beta)\|=o_p(1)$. In particular, $\sup_{\beta\in\mathcal{B}}\|g_{nb}(\dot{\alpha}_n(\beta),\beta)-\bar{g}_{nb}(\alpha_n(\beta),\beta)\|=o_p(1)$. Furthermore, $\sup_{\beta\in\mathcal{B}}\|\bar{g}_n(\alpha_n(\beta),\beta)\|\leq\sup_{\gamma\in\Gamma}\|\bar{g}_n(\gamma)\|=O_p(1)$. Since $\lim_{n\to\infty}\bar{g}_{nb}(\alpha_n(\beta),\beta)$ is uniquely zero at $\beta=\beta_0$ and $\hat{W}_n=\bar{W}_n+o_p(1)$, where $\lim_{n\to\infty}\bar{W}_n$ is nonsingular, the consistency of $\breve{\beta}$ follows.

(ii) Applying the MVT to the first order condition for $\tilde{\beta}$ yields

$$0 = G'_{nb\beta}(\check{\alpha},\tilde{\beta})\hat{W}_n^{-1}g_{nb}(\check{\alpha},\tilde{\beta}) = G'_{nb\beta}(\check{\alpha},\tilde{\beta})\hat{W}_n^{-1}[g_{nb} + G_{nb\beta}(\ddot{\gamma})(\tilde{\beta}-\beta_0) + G_{nb\alpha}(\ddot{\gamma})(\check{\alpha}-\alpha_0)],$$

where $G_{nb\beta}(\gamma) = \frac{\partial g_{nb}(\gamma)}{\partial \beta'}$, $G_{nb\alpha}(\gamma) = \frac{\partial g_{nb}(\gamma)}{\partial \alpha'}$, and $\ddot{\gamma}$ lies between γ_0 and $(\check{\alpha}',\tilde{\beta}')'$. This relation implies that

$$\sqrt{n}(\tilde{\beta}-\beta_0) = -[G'_{nb\beta}(\check{\alpha},\tilde{\beta})\hat{W}_n^{-1}G_{nb\beta}(\ddot{\gamma})]^{-1}G'_{nb\beta}(\check{\alpha},\tilde{\beta})\hat{W}_n^{-1}[\sqrt{n}g_{nb} + G_{nb\alpha}(\ddot{\gamma})\sqrt{n}(\check{\alpha}-\alpha_0)].$$

By Assumption 1(xi), $\sqrt{n}g_{nb} = O_p(1)$. Under Assumption 1(ix), $G_{nb\beta}(\check{\alpha},\tilde{\beta}) = \bar{G}_{nb\beta} + o_p(1) = O_p(1)$ and $G_{nb\alpha}(\ddot{\gamma}) = \bar{G}_{nb\alpha} + o_p(1) = O_p(1)$, where $\bar{G}_{nb\beta}$ has full column rank for large enough n under Assumption D.1(iii). Then, by $\sqrt{n}(\check{\alpha}-\alpha_0) = O_p(1)$ and $\hat{W}_n = \bar{W}_n + O_p(n^{-1/2})$, $\sqrt{n}(\tilde{\beta}-\beta_0) = O_p(1)$. Thus, a first order Taylor expansion of the first order condition for $\tilde{\beta}$ yields

$$0 = G'_{nb\beta}(\check{\alpha},\tilde{\beta})\hat{W}_n^{-1}g_{nb}(\check{\alpha},\tilde{\beta}) = G'_{nb\beta}(\check{\alpha},\tilde{\beta})\hat{W}_n^{-1}[g_{nb} + G_{nb\beta}(\gamma_0)(\tilde{\beta}-\beta_0) + G_{nb\alpha}(\gamma_0)(\check{\alpha}-\alpha_0) + O_p(n^{-1})],$$

where $G_{nb\beta}(\check{\alpha},\tilde{\beta}) = G_{nb\beta}(\gamma_0) + O_p(n^{-1/2}) = \bar{G}_{nb\beta} + O_p(n^{-1/2})$ and $G_{nb\alpha}(\gamma_0) = \bar{G}_{nb\alpha} + O_p(n^{-1/2})$ under Assumption 3(iii). With $\sqrt{n}(\check{\alpha}-\alpha_0) = \psi_{n\check{\alpha}} + O_p(n^{-1/2})$, we have $\sqrt{n}(\tilde{\beta}-\beta_0) = \psi_{n\tilde{\beta}} + O_p(n^{-1/2})$.

For $\breve{\beta}$, applying the MVT to its first order condition

$$[\frac{\partial \dot{\alpha}'_n(\breve{\beta})}{\partial \beta}\frac{\partial g'_{nb}(\dot{\alpha}_n(\breve{\beta}),\breve{\beta})}{\partial \alpha} + \frac{\partial g'_{nb}(\dot{\alpha}_n(\breve{\beta}),\breve{\beta})}{\partial \beta}]\hat{W}_n^{-1}g_{nb}(\dot{\alpha}_n(\breve{\beta}),\breve{\beta}) = 0$$

yields

$$\begin{aligned} 0 = &\Big[\frac{\partial \dot{\alpha}'_n(\breve{\beta})}{\partial \beta}\frac{\partial g'_{nb}(\dot{\alpha}_n(\breve{\beta}),\breve{\beta})}{\partial \alpha} + \frac{\partial g'_{nb}(\dot{\alpha}_n(\breve{\beta}),\breve{\beta})}{\partial \beta}\Big]\hat{W}_n^{-1} \\ &\cdot\Big[g_{nb}(\dot{\alpha}_n(\beta_0),\beta_0) + \frac{\partial g_{nb}(\dot{\alpha}_n(\ddot{\beta}),\ddot{\beta})}{\partial \alpha'}\frac{\partial \dot{\alpha}_n(\ddot{\beta})}{\partial \beta'}(\breve{\beta}-\beta_0) + \frac{\partial g_{nb}(\dot{\alpha}_n(\ddot{\beta}),\ddot{\beta})}{\partial \beta'}(\breve{\beta}-\beta_0)\Big], \end{aligned}$$

where $g_{nb}(\dot{\alpha}_n(\beta_0),\beta_0) = g_{nb}(\alpha_0,\beta_0) + \frac{\partial g_{nb}(\ddot{\alpha},\beta_0)}{\partial \alpha'}(\dot{\alpha}_n(\beta_0)-\alpha_0)$, $\ddot{\beta}$ lies between β_0 and $\breve{\beta}$, and $\ddot{\alpha}$ lies between α_0 and $\dot{\alpha}_n(\beta_0)$. Thus,

$$\begin{aligned} \sqrt{n}(\breve{\beta}-\beta_0) = -&\Big[\Big(\frac{\partial \dot{\alpha}'_n(\breve{\beta})}{\partial \beta}\frac{\partial g'_{nb}(\dot{\alpha}_n(\breve{\beta}),\breve{\beta})}{\partial \alpha} + \frac{\partial g'_{nb}(\dot{\alpha}_n(\breve{\beta}),\breve{\beta})}{\partial \beta}\Big)\hat{W}_n^{-1}\Big(\frac{\partial g_{nb}(\dot{\alpha}_n(\ddot{\beta}),\ddot{\beta})}{\partial \alpha'}\frac{\partial \dot{\alpha}_n(\ddot{\beta})}{\partial \beta'} + \frac{\partial g_{nb}(\dot{\alpha}_n(\ddot{\beta}),\ddot{\beta})}{\partial \beta'}\Big)\Big]^{-1} \\ &\cdot\Big(\frac{\partial \dot{\alpha}'_n(\breve{\beta})}{\partial \beta}\frac{\partial g'_{nb}(\dot{\alpha}_n(\breve{\beta}),\breve{\beta})}{\partial \alpha} + \frac{\partial g'_{nb}(\dot{\alpha}_n(\breve{\beta}),\breve{\beta})}{\partial \beta}\Big)\hat{W}_n^{-1}[\sqrt{n}g_{nb}(\alpha_0,\beta_0) + \frac{\partial g_{nb}(\ddot{\alpha},\beta_0)}{\partial \alpha'}\sqrt{n}(\dot{\alpha}_n(\beta_0)-\alpha_0)]. \end{aligned}$$

Since $\dot{\alpha}_n(\ddot{\beta}) = \alpha_n(\ddot{\beta}) + o_p(1) = \alpha_0 + o_p(1)$, $\frac{\partial g_{nb}(\dot{\alpha}_n(\ddot{\beta}),\ddot{\beta})}{\partial \alpha'} = \bar{G}_{nb\alpha} + o_p(1)$ and $\frac{\partial g_{nb}(\dot{\alpha}_n(\ddot{\beta}),\ddot{\beta})}{\partial \beta'} = \bar{G}_{nb\beta} + o_p(1)$. In addition, $\frac{\partial \dot{\alpha}_n(\ddot{\beta})}{\partial \beta'} = \frac{\partial \alpha_n(\ddot{\beta})}{\partial \beta'} + o_p(1) = \frac{\partial \alpha_n(\beta_0)}{\partial \beta'} + o_p(1)$. Then, $\sqrt{n}(\breve{\beta}-\beta_0) = \psi_{n\breve{\beta}} + o_p(1) = O_p(1)$. With this result, a first order Taylor expansion of the first order condition for $\breve{\beta}$ yields

$$0=\Big[\frac{\partial\dot{\alpha}_n'(\breve{\beta})}{\partial\beta}\frac{\partial g_{nb}'(\dot{\alpha}_n(\breve{\beta}),\breve{\beta})}{\partial\alpha}+\frac{\partial g_{nb}'(\dot{\alpha}_n(\breve{\beta}),\breve{\beta})}{\partial\beta}\Big]\hat{W}_n^{-1}$$
$$\times\Big[g_{nb}(\dot{\alpha}_n(\beta_0),\beta_0)+\frac{\partial g_{nb}(\dot{\alpha}_n(\beta_0),\beta_0)}{\partial\alpha'}\frac{\partial\dot{\alpha}_n(\beta_0)}{\partial\beta'}(\breve{\beta}-\beta_0)+\frac{\partial g_{nb}(\dot{\alpha}_n(\beta_0),\beta_0)}{\partial\beta'}(\breve{\beta}-\beta_0)+O_p(n^{-1})\Big],$$

where $\frac{\partial g_{nb}(\dot{\alpha}_n(\ddot{\beta}),\ddot{\beta})}{\partial\alpha'}=\frac{\partial g_{nb}(\alpha_0,\beta_0)}{\partial\alpha'}+\sum_{j=1}^{k_\beta}[\sum_{k=1}^{k_\alpha}\frac{\partial^2 g_{nb}(\dot{\alpha}_n(\bar{\beta}),\bar{\beta})}{\partial\alpha'\partial\alpha_k}\frac{\partial\check{\alpha}_{nk}(\bar{\beta})}{\partial\beta_j}+\frac{\partial^2 g_{nb}(\dot{\alpha}_n(\bar{\beta}),\bar{\beta})}{\partial\alpha'\partial\beta_j}](\breve{\beta}_j-\beta_{0j})=\bar{G}_{nb\alpha}+O_p(n^{-1/2})$ for some $\bar{\beta}$ between $\breve{\beta}$ and β_0, and $\frac{\partial g_{nb}(\dot{\alpha}_n(\ddot{\beta}),\ddot{\beta})}{\partial\beta'}=\bar{G}_{nb\beta}+O_p(n^{-1/2})$ similarly. Hence, $\sqrt{n}(\breve{\beta}-\beta_0)=\psi_{n\breve{\beta}}+O_p(n^{-1/2})$.

(iii) By a first order Taylor expansion, $\dot{\alpha}_n(\hat{\beta})=\dot{\alpha}_n(\beta_0)+\frac{\partial\dot{\alpha}_n(\beta_0)}{\partial\beta'}(\hat{\beta}-\beta_0)+O_p(n^{-1})=\dot{\alpha}_n(\beta_0)+n^{-1/2}\frac{\partial\alpha_n(\beta_0)}{\partial\beta'}\psi_{n\beta}+O_p(n^{-1})$. Thus, $\sqrt{n}(\dot{\alpha}_n(\hat{\beta})-\alpha_0)=\sqrt{n}(\dot{\alpha}_n(\beta_0)-\alpha_0)+\frac{\partial\alpha_n(\beta_0)}{\partial\beta'}\psi_{n\beta}+O_p(n^{-1/2})$. ■

Lemma H.2. *For $h_{ni}(\beta)=\tilde{C}_n g_{ni}(\check{\alpha},\beta)$, $d_{ni}(\check{\alpha},\beta)$, $\check{C}_n g_{ni}(\dot{\alpha}_n(\beta),\beta)$, $d_{ni}(\dot{\alpha}_n(\beta),\beta)$, $g_{ni}(\check{\alpha},\beta)$ or $g_{ni}(\dot{\alpha}_n(\beta),\beta)$, and any ζ with $1/\eta<\zeta<1/2$, where $\eta>2$ in Assumption 1(v), let μ be a column vector conformable with $h_{ni}(\beta)$, $\Lambda_{nh}=\{\mu:\|\mu\|\le n^{-\zeta}\}$ and $\Lambda_{nh}(\beta)=\{\mu:\mu'h_{ni}(\beta)\in\mathcal{V},i=1,\dots,n\}$. Suppose that Assumptions 1(v) and 2(i)–(iii) hold. In addition, Assumptions 1(ii)–(iv), 2(iv) and D.1(i)–(ii) hold for $h_{ni}(\beta)=\tilde{C}_n g_{ni}(\check{\alpha},\beta)$ or $\check{C}_n g_{ni}(\dot{\alpha}_n(\beta),\beta)$. Then, $\sup_{\beta\in\mathcal{B},\mu\in\Lambda_{nh},1\le i\le n}|\mu'h_{ni}(\beta)|=o_p(1)$ and $\Lambda_{nh}\subset\Lambda_{nh}(\beta)$ for all $\beta\in\mathcal{B}$ with probability approaching one (w.p.a.1).*

Proof. We first prove the result for $h_{ni}(\beta)=\tilde{C}_n g_{ni}(\check{\alpha},\beta)$. By the Cauchy-Schwarz inequality,

$$\sup_{\beta\in\mathcal{B},\mu\in\Lambda_{nh},1\le i\le n}\|\mu'\tilde{C}_n g_{ni}(\check{\alpha},\beta)\|\le\sup_{\mu\in\Lambda_{nh}}\|\mu\|\cdot\|\tilde{C}_n\|\sup_{\gamma\in\Gamma,1\le i\le n}\|g_{ni}(\gamma)\|,\qquad\text{(H.2)}$$

where $\sup_{\gamma\in\Gamma,1\le i\le n}\|g_{ni}(\gamma)\|$ satisfies

$$\mathbf{E}\Big(\sup_{\gamma\in\Gamma,1\le i\le n}\|g_{ni}(\gamma)\|\Big)\le\Big[\mathbf{E}\Big(\sup_{\gamma\in\Gamma,1\le i\le n}\|g_{ni}(\gamma)\|\Big)^\eta\Big]^{1/\eta}\le\Big[\sum_{i=1}^n\mathbf{E}\Big(\sup_{\gamma\in\Gamma}\|g_{ni}(\gamma)\|^\eta\Big)\Big]^{1/\eta}=O(n^{1/\eta})\qquad\text{(H.3)}$$

by Jensen's inequality for a concave function. Thus, $\sup_{\gamma\in\Gamma,1\le i\le n}\|g_{ni}(\gamma)\|=O_p(n^{1/\eta})$. By Lemma H.1$(i)$, $\tilde{\beta}=\beta_0+o_p(1)$. Then under Assumption 2(i)–(ii), $\tilde{C}_n=\bar{C}_n+o_p(1)=O_p(1)$. Hence,

$$\sup_{\beta\in\mathcal{B},\mu\in\Lambda_{nh},1\le i\le n}\|\mu'\tilde{C}_n g_{ni}(\check{\alpha},\beta)\|=O_p(n^{1/\eta-\zeta})=o_p(1).$$

It follows that, w.p.a.1, $\mu'h_{ni}(\beta)\in\mathcal{V}$ for all $\beta\in\mathcal{B}$ and $\|\mu\|\le n^{-\zeta}$.

For $h_{ni}(\beta)=d_{ni}(\check{\alpha},\beta)$, an inequality similar to (H.2) shows that the result holds if $\sup_{\beta\in\mathcal{B}}\|C_n(\check{\alpha},\beta)\|=O_p(1)$. Since $\sup_{\alpha\in\mathcal{N}_\alpha,\beta\in\mathcal{B}}\|C_n(\alpha,\beta)-\bar{C}_n(\alpha,\beta)\|=o_p(1)$ and $\sup_{\alpha\in\mathcal{N}_\alpha,\beta\in\mathcal{B}}\|\bar{C}_n(\alpha,\beta)\|=O(1)$, we have $\sup_{\beta\in\mathcal{B}}\|C_n(\check{\alpha},\beta)\|=O_p(1)$.

For $h_{ni}(\beta) = \check{C}_n g_{ni}(\dot{\alpha}_n(\beta), \beta)$, note that $\check{C}_n = \bar{C}_n + o_p(1) = O_p(1)$ as $\tilde{C}_n = \bar{C}_n + o_p(1) = O_p(1)$. Then,

$$\sup_{\beta\in\mathcal{B},\mu\in\Lambda_{nh},1\le i\le n} \|\mu'\check{C}_n g_{ni}(\dot{\alpha}_n(\beta),\beta)\| \le \sup_{\mu\in\Lambda_{nh}} \|\mu\|\cdot\|\check{C}_n\|\cdot \sup_{\beta\in\mathcal{B},1\le i\le n} \|g_{ni}(\dot{\alpha}_n(\beta),\beta)\|$$
$$\le \sup_{\mu\in\Lambda_{nh}} \|\mu\|\cdot\|\check{C}_n\|\cdot \sup_{\gamma\in\Gamma,1\le i\le n} \|g_{ni}(\gamma)\| = O_p(n^{1/\eta-\zeta}) = o_p(1).$$

For $h_{ni}(\beta) = C_n(\dot{\alpha}_n(\beta),\beta) g_{ni}(\dot{\alpha}_n(\beta),\beta)$,

$$\sup_{\beta\in\mathcal{B},\mu\in\Lambda_{nh},1\le i\le n} \|\mu' C_n(\dot{\alpha}_n(\beta),\beta) g_{ni}(\dot{\alpha}_n(\beta),\beta)\| \le \sup_{\mu\in\Lambda_{nh}} \|\mu\|\cdot \sup_{\gamma\in\Gamma}\|C_n(\gamma)\| \sup_{\gamma\in\Gamma,1\le i\le n} \|g_{ni}(\gamma)\|.$$

Then the result holds since $\sup_{\gamma\in\Gamma}\|C_n(\gamma)\| = O_p(1)$.

The proofs for the results with $h_{ni}(\beta) = g_{ni}(\check{\alpha},\beta)$ and $h_{ni}(\beta) = g_{ni}(\dot{\alpha}_n(\beta),\beta)$ are similar to those for $h_{ni}(\beta) = \tilde{C}_n g_{ni}(\check{\alpha},\beta)$ and $h_{ni}(\beta) = \check{C}_n g_{ni}(\dot{\alpha}_n(\beta),\beta)$. ■

With Lemma H.2, the following Lemmas H.3–H.4 hold by arguments similar to those for Lemmas A.2–A.3 in Newey and Smith (2004).

Lemma H.3. *Let $h_{ni}(\beta) = \tilde{C}_n g_{ni}(\check{\alpha},\beta)$, $d_{ni}(\check{\alpha},\beta)$, $\check{C}_n g_{ni}(\dot{\alpha}_n(\beta),\beta)$, $d_{ni}(\dot{\alpha}_n(\beta),\beta)$, $g_{ni}(\check{\alpha},\beta)$ or $g_{ni}(\dot{\alpha}_n(\beta),\beta)$, μ be a column vector conformable with $h_{ni}(\beta)$, and $\Lambda_{nh}(\beta) = \{\mu : \mu' h_{ni}(\beta) \in \mathcal{V}, i = 1,\ldots,n\}$. Suppose that $\bar{\beta} \in \mathcal{B}$, $\bar{\beta} = \beta_0 + o_p(1)$, $\frac{1}{n}\sum_{i=1}^n h_{ni}(\bar{\beta}) = O_p(n^{-1/2})$, and Assumptions 1(v)–(vii) and 2(i)–(iv) hold. In addition, Assumptions 1(ii)–(iv) and D.1(i)–(ii) hold for $h_{ni}(\beta) = \tilde{C}_n g_{ni}(\check{\alpha},\beta)$ or $\check{C}_n g_{ni}(\dot{\alpha}_n(\beta),\beta)$. Then, $\bar{\mu} = \arg\max_{\mu\in\Lambda_{nh}(\bar{\beta})} \sum_{i=1}^n \rho(\mu' h_{ni}(\bar{\beta}))$ exists w.p.a.1, $\bar{\mu} = O_p(n^{-1/2})$, and $\sup_{\mu\in\Lambda_{nh}(\bar{\beta})} \sum_{i=1}^n \rho(\mu' h_{ni}(\bar{\beta})) \le \rho(0) + O_p(n^{-1})$.*

Lemma H.4. *If Assumptions 1(v)–(vii) and 2(i)–(iv) hold, then $\|d_n(\check{\alpha},\hat{\beta}_{\text{TGEL}})\| = O_p(n^{-1/2})$ and $\|d_n(\dot{\alpha}_n(\hat{\beta}_{\text{TGEL2}}),\hat{\beta}_{\text{TGEL2}})\| = O_p(n^{-1/2})$. If Assumptions 1(ii)–(iv) and D.1(i)–(ii) also hold, $\|\tilde{C}_n g_n(\check{\alpha},\hat{\beta}_{\text{TGEL3}})\| = O_p(n^{-1/2})$ and $\|\check{C}_n g_n(\dot{\alpha}_n(\hat{\beta}_{\text{TGEL4}}),\hat{\beta}_{\text{TGEL4}})\| = O_p(n^{-1/2})$.*

Lemma H.5. *(i) Under Assumptions 1(ii)–(vi), 2(iii)–(iv) and D.1(i)–(ii), $\tilde{\Omega}_n = \bar{\Omega}_n + o_p(1)$ and $\check{\Omega}_n = \bar{\Omega}_n + o_p(1)$.*

(ii) Under Assumptions 1(ii)–(vi), (viii)–(ix), (xi), 2(iii)–(iv), 3(i), (iii) and D.1, $\sqrt{n}(\tilde{\Omega}_n - \bar{\Omega}_n) = \psi_{n\tilde{\Omega}} + O_p(n^{-1/2})$, where $\psi_{n\tilde{\Omega}} = \sqrt{n}(\frac{1}{n}\sum_{i=1}^n g_{ni}g_{ni}' - \bar{\Omega}_n) + \frac{1}{n}\sum_{j=1}^{k_\alpha}\sum_{i=1}^n \mathbf{E}(g_{ni}g_{ni}^{(j)\prime} + g_{ni}^{(j)}g_{ni}')\psi_{n\check{\alpha}j} + \frac{1}{n}\sum_{j=1}^{k_\beta}\sum_{i=1}^n \mathbf{E}(g_{ni}g_{ni}^{(k_\alpha+j)\prime} + g_{ni}^{(k_\alpha+j)}g_{ni}')\psi_{n\tilde{\beta}j}$, with $\psi_{n\check{\alpha}j}$ and $\psi_{n\tilde{\beta}j}$ denoting, respectively, the jth elements of $\psi_{n\check{\alpha}}$ defined in Assumption 3(i) and $\psi_{n\tilde{\beta}}$ defined in Lemma H.1(ii).

(iii) *Under Assumptions 1(ii)–(v), (viii)–(ix), (xi), 2(iii)–(iv), 3(i)–(ii) and D.1,* $\tilde{C}_n = \bar{C}_n + n^{-1/2}\psi_{n\tilde{C}} + O_p(n^{-1})$*, where* $\psi_{n\tilde{C}} = \psi_{nC} + \sum_{j=1}^{k_\alpha} \bar{C}_n^{(j)}\psi_{n\check{\alpha}j} + \sum_{j=1}^{k_\beta} \bar{C}_n^{(k_\alpha+j)}\psi_{n\tilde{\beta}j}$*.*

(iv) *Under Assumptions 1(ii)–(vi), (viii)–(ix), (xi), 2(iii)–(iv), 3(i)–(iii) and D.1,* $\sqrt{n}(\tilde{C}_n\tilde{\Omega}_n\tilde{C}_n' - \bar{\Omega}_{nd}) = \psi_{n\tilde{\Omega}_d} + O_p(n^{-1/2})$*, where* $\psi_{n\tilde{\Omega}_d} = \dot{\psi}_{n\tilde{\Omega}_d} + \ddot{\psi}_{n\tilde{\Omega}_d}$*,*

$$\dot{\psi}_{n\tilde{\Omega}_d} = \sqrt{n}\bar{C}_n\Big(\frac{1}{n}\sum_{i=1}^n g_{ni}g_{ni}' - \bar{\Omega}_n\Big)\bar{C}_n' - \frac{1}{n}\bar{C}_n\sum_{j=1}^{k_\beta}\sum_{i=1}^n \mathbf{E}(g_{ni}g_{ni}^{(k_\alpha+j)'} + g_{ni}^{(k_\alpha+j)}g_{ni}')\bar{C}_n'e_{k_\beta j}'\bar{H}_{nW\beta}\sqrt{n}g_{nb},$$

and

$$\ddot{\psi}_{n\tilde{\Omega}_d} = \bar{C}_n\bar{\Omega}_n\psi_{n\tilde{C}}' + \psi_{n\tilde{C}}\bar{\Omega}_n\bar{C}_n' + \frac{1}{n}\sum_{j=1}^{k_\alpha}\sum_{i=1}^n \psi_{n\check{\alpha}j}\bar{C}_n\mathbf{E}(g_{ni}g_{ni}^{(j)'} + g_{ni}^{(j)}g_{ni}')\bar{C}_n'$$
$$- \frac{1}{n}\sum_{j=1}^{k_\beta}\sum_{i=1}^n \bar{C}_n\mathbf{E}(g_{ni}g_{ni}^{(k_\alpha+j)'} + g_{ni}^{(k_\alpha+j)}g_{ni}')\bar{C}_n'e_{k_\beta j}'\bar{H}_{nW\beta}\bar{G}_{nb\alpha}\psi_{n\check{\alpha}},$$

with $\psi_{n\tilde{C}}$ *given in (iii). In* $\psi_{n\tilde{\Omega}_d}$*,* $\ddot{\psi}_{n\tilde{\Omega}_d}$ *involves terms due to the estimation of* α *and* $\bar{C}_n$*, but* $\dot{\psi}_{n\tilde{\Omega}_d}$ *does not.*

Proof. (i) By Assumption 2(iv) and Lemma H.1(i), $\bar{\gamma} = (\check{\alpha}', \tilde{\beta}')'$ is a consistent estimator of γ_0. Then $\|\Omega_n(\bar{\gamma}) - \bar{\Omega}_n\| \le \|\Omega_n(\bar{\gamma}) - \bar{\Omega}_n(\bar{\gamma})\| + \|\bar{\Omega}_n(\bar{\gamma}) - \bar{\Omega}_n\| = o_p(1)$ under Assumption 1(vi). Thus the result follows.

(ii) A first order Taylor expansion yields

$$\Omega_n(\bar{\gamma}) = \bar{\Omega}_n + \Big(\frac{1}{n}\sum_{i=1}^n g_{ni}g_{ni}' - \bar{\Omega}_n\Big) + \sum_{j=1}^{k_\alpha+k_\beta}\frac{1}{n}\sum_{i=1}^n (g_{ni}g_{ni}^{(j)'} + g_{ni}^{(j)}g_{ni}')(\bar{\gamma}_j - \gamma_{j0}) + O_p(\|\bar{\gamma} - \gamma_0\|^2). \quad \text{(H.4)}$$

Substituting the expressions for $(\check{\alpha} - \alpha_0)$ in Assumption 3(i) and $(\tilde{\beta} - \beta_0)$ in Lemma H.1(ii) into (H.4) and keeping only leading order terms, we obtain $\sqrt{n}[\Omega_n(\bar{\gamma}) - \bar{\Omega}_n] = \psi_{n\tilde{\Omega}} + O_p(n^{-1/2})$.

(iii) By the MVT, $C_n(\check{\alpha}, \tilde{\beta}) = C_n(\alpha_0, \beta_0) + \sum_{j=1}^{k_\alpha}\frac{\partial C_n(\ddot{\alpha}, \ddot{\beta})}{\partial \alpha_j}(\check{\alpha}_j - \alpha_{j0}) + \sum_{j=1}^{k_\beta}\frac{\partial C_n(\ddot{\alpha}, \ddot{\beta})}{\partial \beta_j}(\tilde{\beta}_j - \beta_{j0})$, where $\ddot{\alpha}$ lies between $\check{\alpha}$ and α_0, and $\ddot{\beta}$ lies between $\tilde{\beta}$ and β_0. Then under Assumption 3(ii), $C_n(\check{\alpha}, \tilde{\beta}) = \bar{C}_n + n^{-1/2}\psi_{nC} + n^{-1/2}\sum_{j=1}^{k_\alpha}\bar{C}_n^{(j)}\psi_{n\check{\alpha}j} + n^{-1/2}\sum_{j=1}^{k_\beta}\bar{C}_n^{(k_\alpha+j)}\psi_{n\tilde{\beta}j} + O_p(n^{-1})$.

(iv) With (ii) and (iii),

$$\Omega_{nd}(\bar{\gamma}) = [\bar{C}_n + n^{-1/2}\psi_{n\tilde{C}} + O_p(n^{-1})][\bar{\Omega}_n + n^{-1/2}\psi_{n\tilde{\Omega}} + O_p(n^{-1})][\bar{C}_n' + n^{-1/2}\psi_{n\tilde{C}}' + O_p(n^{-1})]$$
$$= \bar{C}_n\bar{\Omega}_n\bar{C}_n' + n^{-1/2}(\bar{C}_n\bar{\Omega}_n\psi_{n\tilde{C}}' + \psi_{n\tilde{C}}\bar{\Omega}_n\bar{C}_n' + \bar{C}_n\psi_{n\tilde{\Omega}}\bar{C}_n') + O_p(n^{-1}).$$

Hence the result holds by using the expression of $\psi_{n\tilde{\Omega}}$ in (ii). ∎

I Proofs of some theorems

I.1 Proof of Theorem D.2

(i) Let $v_{ni}(\alpha,\theta) = \mu'\tilde{C}_n g_{ni}(\alpha,\beta)$, where $\theta = (\beta',\mu')'$, $h_{ni}(\alpha,\theta) = \frac{\partial v_{ni}(\alpha,\theta)}{\partial\theta} = \begin{pmatrix} G'_{ni\beta}(\alpha,\beta)\tilde{C}'_n\mu \\ \tilde{C}_n g_{ni}(\alpha,\beta) \end{pmatrix}$,

$$m_{ni}(\alpha,\theta) = \rho_1(v_{ni}(\alpha,\theta))h_{ni}(\alpha,\theta),$$

and $\hat{m}_n(\alpha,\theta) = \frac{1}{n}\sum_{i=1}^n m_{ni}(\alpha,\theta)$. Then the first order condition for $\hat{\theta}_{\text{TGEL}3}$ is

$$\hat{m}_n(\check{\alpha},\hat{\theta}_{\text{TGEL}3}) = 0. \tag{I.1}$$

We shall derive a Nagar-type expansion by applying a Taylor's expansion to this condition. For that purpose, we compute the following derivatives:

$$\frac{\partial h_{ni}(\alpha,\theta)}{\partial\alpha'} = \begin{pmatrix} [G^{(1)'}_{ni\beta}(\alpha,\beta)\tilde{C}'_n\mu,\ldots,G^{(k_\alpha)'}_{ni\beta}(\alpha,\beta)\tilde{C}'_n\mu] \\ \tilde{C}_n G_{ni\alpha}(\alpha,\beta) \end{pmatrix},$$

$$\frac{\partial h_{ni}(\alpha,\theta)}{\partial\theta'} = \begin{pmatrix} [G^{(k_\alpha+1)'}_{ni\beta}(\alpha,\beta)\tilde{C}'_n\mu,\ldots,G^{(k_\alpha+k_\beta)'}_{ni\beta}(\alpha,\beta)\tilde{C}'_n\mu] & G'_{ni\beta}(\alpha,\beta)\tilde{C}'_n \\ \tilde{C}_n G_{ni\beta}(\alpha,\beta) & 0 \end{pmatrix},$$

$$\frac{\partial^2 h_{ni}(\alpha,\theta)}{\partial\alpha_j\partial\alpha'} = \begin{pmatrix} [G^{(1j)'}_{ni\beta}(\alpha,\beta)\tilde{C}'_n\mu,\ldots,G^{(k_\alpha j)'}_{ni\beta}(\alpha,\beta)\tilde{C}'_n\mu] \\ \tilde{C}_n G^{(j)}_{ni\alpha}(\alpha,\beta) \end{pmatrix},$$

$$\frac{\partial^2 h_{ni}(\alpha,\theta)}{\partial\theta_j\partial\alpha'} = \begin{pmatrix} [G^{(1,k_\alpha+j)'}_{ni\beta}(\alpha,\beta)\tilde{C}'_n\mu,\ldots,G^{(k_\alpha,k_\alpha+j)'}_{ni\beta}(\alpha,\beta)\tilde{C}'_n\mu] \\ \tilde{C}_n G^{(k_\alpha+j)}_{ni\alpha}(\alpha,\beta) \end{pmatrix} \text{ for } 1\le j\le k_\beta,$$

$$\frac{\partial^2 h_{ni}(\alpha,\theta)}{\partial\theta_j\partial\alpha'} = \begin{pmatrix} [G^{(1)'}_{ni\beta}(\alpha,\beta)\tilde{C}'_n e_{m_b,j-k_\beta},\ldots,G^{(k_\alpha)'}_{ni\beta}(\alpha,\beta)\tilde{C}'_n e_{m_b,j-k_\beta}] \\ 0 \end{pmatrix} \text{ for } k_\beta+1\le j\le k_\theta,$$

$$\frac{\partial^2 h_{ni}(\alpha,\theta)}{\partial\alpha_j\partial\theta'} = \begin{pmatrix} [G^{(k_\alpha+1,j)'}_{ni\beta}(\alpha,\beta)\tilde{C}'_n\mu,\ldots,G^{(k_\alpha+k_\beta,j)'}_{ni\beta}(\alpha,\beta)\tilde{C}'_n\mu] & G^{(j)'}_{ni\beta}(\alpha,\beta)\tilde{C}'_n \\ \tilde{C}_n G^{(j)}_{ni\beta}(\alpha,\beta) & 0 \end{pmatrix},$$

$$\frac{\partial^2 h_{ni}(\alpha,\theta)}{\partial\theta_j\partial\theta'} = \begin{pmatrix} [G^{(k_\alpha+1,k_\alpha+j)'}_{ni\beta}(\alpha,\beta)\tilde{C}'_n\mu,\ldots,G^{(k_\alpha+k_\beta,k_\alpha+j)'}_{ni\beta}(\alpha,\beta)\tilde{C}'_n\mu] & G^{(k_\alpha+j)'}_{ni\beta}(\alpha,\beta)\tilde{C}'_n \\ \tilde{C}_n G^{(k_\alpha+j)}_{ni\beta}(\alpha,\beta) & 0 \end{pmatrix} \text{ for } 1\le j\le k_\beta,$$

$$\frac{\partial^2 h_{ni}(\alpha,\theta)}{\partial\theta_j\partial\theta'} = \begin{pmatrix} [G^{(k_\alpha+1)'}_{ni\beta}(\alpha,\beta)\tilde{C}'_n e_{m_b,j-k_\beta},\ldots,G^{(k_\alpha+k_\beta)'}_{ni\beta}(\alpha,\beta)\tilde{C}'_n e_{m_b,j-k_\beta}] & 0 \\ 0 & 0 \end{pmatrix} \text{ for } k_\beta+1\le j\le k_\theta,$$

$$\frac{\partial^3 h_{ni}(\alpha,\theta)}{\partial\alpha_k\partial\alpha_j\partial\alpha'} = \begin{pmatrix} [G_{ni\beta}^{(1jk)'}(\alpha,\beta)\tilde{C}_n'\mu,\ldots,G_{ni\beta}^{(k_\alpha jk)'}(\alpha,\beta)\tilde{C}_n'\mu] \\ \tilde{C}_n G_{ni\alpha}^{(jk)}(\alpha,\beta) \end{pmatrix},$$

$$\frac{\partial^3 h_{ni}(\alpha,\theta)}{\partial\alpha_k\partial\alpha_j\partial\theta'} = \begin{pmatrix} [G_{ni\beta}^{(k_\alpha+1,jk)'}(\alpha,\beta)\tilde{C}_n'\mu,\ldots,G_{ni\beta}^{(k_\alpha+k_\beta,jk)'}(\alpha,\beta)\tilde{C}_n'\mu] & G_{ni\beta}^{(jk)'}(\alpha,\beta)\tilde{C}_n' \\ \tilde{C}_n G_{ni\beta}^{(jk)}(\alpha,\beta) & 0 \end{pmatrix},$$

$$\frac{\partial^3 h_{ni}(\alpha,\theta)}{\partial\alpha_k\partial\theta_j\partial\theta'} = \begin{pmatrix} [G_{ni\beta}^{(k_\alpha+1,k_\alpha+j,k)'}(\alpha,\beta)\tilde{C}_n'\mu,\ldots,G_{ni\beta}^{(k_\alpha+k_\beta,k_\alpha+j,k)'}(\alpha,\beta)\tilde{C}_n'\mu] & G_{ni\beta}^{(k_\alpha+j,k)'}(\alpha,\beta)\tilde{C}_n' \\ \tilde{C}_n G_{ni\beta}^{(k_\alpha+j,k)}(\alpha,\beta) & 0 \end{pmatrix} \text{ for } 1 \le j \le k_\beta,$$

$$\frac{\partial^3 h_{ni}(\alpha,\theta)}{\partial\alpha_k\partial\theta_j\partial\theta'} = \begin{pmatrix} [G_{ni\beta}^{(k_\alpha+1,k)'}(\alpha,\beta)\tilde{C}_n' e_{m_b,j-k_\beta},\ldots,G_{ni\beta}^{(k_\alpha+k_\beta,k)'}(\alpha,\beta)\tilde{C}_n' e_{m_b,j-k_\beta}] & 0 \\ 0 & 0 \end{pmatrix} \text{ for } k_\beta+1 \le j \le k_\theta,$$

$$\frac{\partial^3 h_{ni}(\alpha,\theta)}{\partial\theta_k\partial\theta_j\partial\theta'} = \begin{pmatrix} [G_{ni\beta}^{(k_\alpha+1,k_\alpha+j,k_\alpha+k)'}(\alpha,\beta)\tilde{C}_n'\mu,\ldots,G_{ni\beta}^{(k_\alpha+k_\beta,k_\alpha+j,k_\alpha+k)'}(\alpha,\beta)\tilde{C}_n'\mu] & G_{ni\beta}^{(k_\alpha+j,k_\alpha+k)'}(\alpha,\beta)\tilde{C}_n' \\ \tilde{C}_n G_{ni\beta}^{(k_\alpha+j,k_\alpha+k)}(\alpha,\beta) & 0 \end{pmatrix}$$

for $1 \le j \le k_\beta$ and $1 \le k \le k_\beta$,

$$\frac{\partial^3 h_{ni}(\alpha,\theta)}{\partial\theta_k\partial\theta_j\partial\theta'} = \begin{pmatrix} [G_{ni\beta}^{(k_\alpha+1,k_\alpha+j)'}(\alpha,\beta)\tilde{C}_n' e_{m_b,k-k_\beta},\ldots,G_{ni\beta}^{(k_\alpha+k_\beta,k_\alpha+j)'}(\alpha,\beta)\tilde{C}_n' e_{m_b,k-k_\beta}] & 0 \\ 0 & 0 \end{pmatrix} \text{ for } 1 \le j \le k_\beta \text{ and}$$

$k_\beta + 1 \le k \le k_\theta$,

$$\frac{\partial^3 h_{ni}(\alpha,\theta)}{\partial\theta_k\partial\theta_j\partial\theta'} = 0 \text{ for } k_\beta+1 \le j \le k_\theta \text{ and } k_\beta+1 \le k \le k_\theta,$$

where $k_\theta = k_\beta + m_b$. Hence, by the chain rule of differentiation, and then evaluated at the true parameters (with $\mu_0 = 0$),

$$\frac{\partial m_{ni}(\alpha_0,\theta_0)}{\partial\alpha'} = -\begin{pmatrix} 0 \\ \tilde{C}_n G_{ni\alpha} \end{pmatrix}, \quad \frac{\partial m_{ni}(\alpha_0,\theta_0)}{\partial\theta'} = -\begin{pmatrix} 0 & G_{ni\beta}'\tilde{C}_n' \\ \tilde{C}_n G_{ni\beta} & \tilde{C}_n g_{ni} g_{ni}' \tilde{C}_n' \end{pmatrix}, \quad \frac{\partial^2 m_{ni}(\alpha_0,\theta_0)}{\partial\alpha_j\partial\alpha'} = -\begin{pmatrix} 0 \\ \tilde{C}_n G_{ni\alpha}^{(j)} \end{pmatrix},$$

$$\frac{\partial^2 m_{ni}(\alpha_0,\theta_0)}{\partial\theta_j\partial\alpha'} = -\begin{pmatrix} 0 \\ \tilde{C}_n G_{ni\alpha}^{(k_\alpha+j)} \end{pmatrix} \text{ for } 1 \le j \le k_\beta,$$

$$\frac{\partial^2 m_{ni}(\alpha_0,\theta_0)}{\partial\theta_j\partial\alpha'} = -\begin{pmatrix} [G_{ni\beta}^{(1)'}\tilde{C}_n' e_{m_b,j-k_\beta},\ldots,G_{ni\beta}^{(k_\alpha)'}\tilde{C}_n' e_{m_b,j-k_\beta}] \\ e_{m_b,j-k_\beta}'\tilde{C}_n g_{ni}\tilde{C}_n G_{ni\alpha} + \tilde{C}_n g_{ni} e_{m_b,j-k_\beta}'\tilde{C}_n G_{ni\alpha} \end{pmatrix} \text{ for } k_\beta+1 \le j \le k_\theta,$$

$$\frac{\partial^2 m_{ni}(\alpha_0,\theta_0)}{\partial\alpha_j\partial\theta'} = -\begin{pmatrix} 0 & G_{ni\beta}^{(j)'}\tilde{C}_n' \\ \tilde{C}_n G_{ni\beta}^{(j)} & \tilde{C}_n g_{ni}^{(j)} g_{ni}'\tilde{C}_n' + \tilde{C}_n g_{ni} g_{ni}^{(j)'}\tilde{C}_n' \end{pmatrix},$$

$$\frac{\partial^2 m_{ni}(\alpha_0,\theta_0)}{\partial\theta_j\partial\theta'} = -\begin{pmatrix} 0 & G_{ni\beta}^{(k_\alpha+j)'}\tilde{C}_n' \\ \tilde{C}_n G_{ni\beta}^{(k_\alpha+j)} & \tilde{C}_n g_{ni}^{(k_\alpha+j)} g_{ni}'\tilde{C}_n' + \tilde{C}_n g_{ni} g_{ni}^{(k_\alpha+j)'}\tilde{C}_n' \end{pmatrix} \text{ for } 1 \le j \le k_\beta,$$

$$\frac{\partial^2 m_{ni}(\alpha_0,\theta_0)}{\partial\theta_j\partial\theta'} = -\begin{pmatrix} [G_{ni\beta}^{(k_\alpha+1)'}\tilde{C}_n' e_{m_b,j-k_\beta},\ldots,G_{ni\beta}^{(k_\alpha+k_\beta)'}\tilde{C}_n' e_{m_b,j-k_\beta}] & e_{m_b,j-k_\beta}'\tilde{C}_n g_{ni}G_{ni\beta}'\tilde{C}_n' + G_{ni\beta}'\tilde{C}_n' e_{m_b,j-k_\beta} g_{ni}'\tilde{C}_n' \\ e_{m_b,j-k_\beta}'\tilde{C}_n g_{ni}\tilde{C}_n G_{ni\beta} + \tilde{C}_n g_{ni} e_{m_b,j-k_\beta}'\tilde{C}_n G_{ni\beta} & -\rho_3(0) e_{m_b,j-k_\beta}'\tilde{C}_n g_{ni}\tilde{C}_n g_{ni} g_{ni}'\tilde{C}_n' \end{pmatrix}$$

for $k_\beta + 1 \le j \le k_\theta$.

Under Assumption 3, each term involving $h_{ni}(\alpha,\theta)$ and its third derivatives is bounded by cb_{ni}^k w.p.a.1 in a neighborhood of (α_0,θ_0) for some constant c and $1 \le k \le 4$. By the Lipschitz condition on $\rho_j(v)$, for some constant c, $|\rho_j(v_{ni}(\alpha,\theta)) - \rho_j(0)| \le c|v_{ni}(\alpha,\theta)| \le c\|\mu\|\cdot\|\tilde{C}_n\|\cdot\|g_{ni}(\alpha,\beta)\|$. Then each term involving $m_{ni}(\alpha,\theta)$ and its third derivatives is bounded by cb_{ni}^k w.p.a.1 in a neighborhood of (α_0,θ_0) for some constant c and $1 \le k \le 5$. Thus, by a second-order Taylor expansion of (I.1),

$$\begin{aligned} 0 = {} & \hat{m}_n(\alpha_0,\theta_0) + \frac{\partial\hat{m}_n(\alpha_0,\theta_0)}{\partial\alpha'}(\check{\alpha}-\alpha_0) + \frac{\partial\hat{m}_n(\alpha_0,\theta_0)}{\partial\theta'}(\hat{\theta}_{\text{TGEL3}}-\theta_0) \\ & + \frac{1}{2}\sum_{j=1}^{k_\alpha}(\check{\alpha}_j-\alpha_{j0})\frac{\partial^2\hat{m}_n(\alpha_0,\theta_0)}{\partial\alpha_j\partial\alpha'}(\check{\alpha}-\alpha_0) + \frac{1}{2}\sum_{j=1}^{k_\theta}(\hat{\theta}_{j,\text{TGEL3}}-\theta_{j0})\frac{\partial^2\hat{m}_n(\alpha_0,\theta_0)}{\partial\theta_j\partial\alpha'}(\check{\alpha}-\alpha_0) \\ & + \frac{1}{2}\sum_{j=1}^{k_\alpha}(\check{\alpha}_j-\alpha_{j0})\frac{\partial^2\hat{m}_n(\alpha_0,\theta_0)}{\partial\alpha_j\partial\theta'}(\hat{\theta}_{\text{TGEL3}}-\theta_0) + \frac{1}{2}\sum_{j=1}^{k_\theta}(\hat{\theta}_{j,\text{TGEL3}}-\theta_{j0})\frac{\partial^2\hat{m}_n(\alpha_0,\theta_0)}{\partial\theta_j\partial\theta'}(\hat{\theta}_{\text{TGEL3}}-\theta_0) \\ & + O_p(n^{-3/2}). \end{aligned}$$

Let $\hat{\mathfrak{m}}_n(\alpha,\theta)$ be the term derived by replacing $\tilde{C}_n$ in $\hat{m}_n(\alpha,\theta)$ with $\bar{C}_n$. As $\tilde{C}_n = \bar{C}_n + n^{-1/2}\psi_{n\tilde{C}} + O_p(n^{-1})$ by Lemma H.5(*iii*), $\|\frac{\partial^2\hat{m}_n(\alpha_0,\theta_0)}{\partial\alpha_j\partial\alpha'} - \mathbf{E}\frac{\partial^2\hat{\mathfrak{m}}_n(\alpha_0,\theta_0)}{\partial\alpha_j\partial\alpha'}\| = O_p(n^{-1/2})$. Similar results hold for $\frac{\partial^2\hat{m}_n(\alpha_0,\theta_0)}{\partial\theta_j\partial\alpha'}$, $\frac{\partial^2\hat{m}_n(\alpha_0,\theta_0)}{\partial\alpha_j\partial\theta'}$ and $\frac{\partial^2\hat{m}_n(\alpha_0,\theta_0)}{\partial\theta_j\partial\theta'}$. Thus,

$$\begin{aligned} \sqrt{n}(\hat{\theta}_{\text{TGEL3}}-\theta_0) = {} & \sqrt{n}\bar{K}_{nd\beta}^{-1}\hat{m}_n(\alpha_0,\theta_0) + \bar{K}_{nd\beta}^{-1}\frac{\partial\hat{m}_n(\alpha_0,\theta_0)}{\partial\alpha'}\sqrt{n}(\check{\alpha}-\alpha_0) \\ & + \bar{K}_{nd\beta}^{-1}\Big(\frac{\partial\hat{m}_n(\alpha_0,\theta_0)}{\partial\theta'} + \bar{K}_{nd\beta}\Big)\sqrt{n}(\hat{\theta}_{\text{TGEL3}}-\theta_0) \\ & + \frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\alpha}(\check{\alpha}_j-\alpha_{j0})\Big(\mathbf{E}\frac{\partial^2\hat{\mathfrak{m}}_n(\alpha_0,\theta_0)}{\partial\alpha_j\partial\alpha'}\Big)\sqrt{n}(\check{\alpha}-\alpha_0) \\ & + \frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\theta}(\hat{\theta}_{j,\text{TGEL3}}-\theta_{j0})\Big(\mathbf{E}\frac{\partial^2\hat{\mathfrak{m}}_n(\alpha_0,\theta_0)}{\partial\theta_j\partial\alpha'}\Big)\sqrt{n}(\check{\alpha}-\alpha_0) \\ & + \frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\alpha}(\check{\alpha}_j-\alpha_{j0})\Big(\mathbf{E}\frac{\partial^2\hat{\mathfrak{m}}_n(\alpha_0,\theta_0)}{\partial\alpha_j\partial\theta'}\Big)\sqrt{n}(\hat{\theta}_{\text{TGEL3}}-\theta_0) \\ & + \frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\theta}(\hat{\theta}_{j,\text{TGEL3}}-\theta_{j0})\Big(\mathbf{E}\frac{\partial^2\hat{\mathfrak{m}}_n(\alpha_0,\theta_0)}{\partial\theta_j\partial\theta'}\Big)\sqrt{n}(\hat{\theta}_{\text{TGEL3}}-\theta_0) + O_p(n^{-1}), \end{aligned} \tag{I.2}$$

where $\bar{K}_{nd\beta} = \begin{pmatrix} 0 & \bar{G}'_{n\beta}\bar{C}'_n \\ \bar{C}_n\bar{G}_{n\beta} & \bar{\Omega}_{nd} \end{pmatrix}$. Using $\bar{C}_n\bar{G}_{n\alpha} = 0$ and expressions for derivatives of $m_{ni}(\alpha,\theta)$, we have $\sqrt{n}(\hat{\theta}_{\text{TGEL3}} - \theta_0) = \psi_{n\theta} + O_p(n^{-1/2})$, where $\psi_{n\theta} = -\bar{K}^{-1}_{nd\beta}\binom{0}{\sqrt{n}\bar{C}_n g_n} = -\binom{\bar{H}_{nd}}{\bar{P}_{nd}}\sqrt{n}\bar{C}_n g_n$, since $\bar{K}^{-1}_{nd\beta} = \begin{pmatrix} -\bar{\Sigma}_{nd} & \bar{H}_{nd} \\ \bar{H}'_{nd} & \bar{P}_{nd} \end{pmatrix}$. Substituting this expression into (I.2) yields $\sqrt{n}(\hat{\theta}_{\text{TGEL3}} - \theta_0) = \psi_{n\theta} + n^{-1/2}\varphi_{n\theta,\text{TGEL3}} + O_p(n^{-1})$, where

$$
\begin{aligned}
\varphi_{n\theta,\text{TGEL3}} = & -\bar{K}^{-1}_{nd\beta}\begin{pmatrix} 0 \\ \sqrt{n}\psi_{n\tilde{C}}g_n \end{pmatrix} - \bar{K}^{-1}_{nd\beta}\begin{pmatrix} 0 \\ \sqrt{n}\bar{C}_n(G_{n\alpha} - \bar{G}_{n\alpha}) + \psi_{n\tilde{C}}\bar{G}_{n\alpha} \end{pmatrix}\psi_{n\check{\alpha}} \\
& - \bar{K}^{-1}_{nd\beta}\begin{pmatrix} 0 & \sqrt{n}(G'_{n\beta} - \bar{G}'_{n\beta})\bar{C}'_n + \bar{G}'_{n\beta}\psi'_{n\tilde{C}} \\ \sqrt{n}\bar{C}_n(G_{n\beta} - \bar{G}_{n\beta}) + \psi_{n\tilde{C}}\bar{G}_{n\beta} & \psi_{n\tilde{C}}\bar{\Omega}_n\bar{C}'_n + \bar{C}_n\bar{\Omega}_n\psi'_{n\tilde{C}} + \sqrt{n}\bar{C}_n(\Omega_n - \bar{\Omega}_n)\bar{C}'_n \end{pmatrix}\psi_{n\theta} \\
& - \frac{1}{2}\bar{K}^{-1}_{nd\beta}\sum_{j=1}^{k_\alpha}\psi_{n\check{\alpha}j}\begin{pmatrix} 0 \\ \bar{C}_n\bar{G}^{(j)}_{n\alpha} \end{pmatrix}\psi_{n\check{\alpha}} - \frac{1}{2}\bar{K}^{-1}_{nd\beta}\sum_{j=1}^{k_\beta}\psi_{n\theta j}\begin{pmatrix} 0 \\ \bar{C}_n\bar{G}^{(k_\alpha+j)}_{n\alpha} \end{pmatrix}\psi_{n\check{\alpha}} \\
& - \frac{1}{2n}\bar{K}^{-1}_{nd\beta}\sum_{s=1}^{m_b}\psi_{n\theta,k_\beta+s}\sum_{i=1}^{n}\mathbf{E}\begin{pmatrix} [G^{(1)'}_{ni\beta}\bar{C}'_n e_{m_b s},\ldots,G^{(k_\alpha)'}_{ni\beta}\bar{C}'_n e_{m_b s}] \\ e'_{m_b s}\bar{C}_n g_{ni}\bar{C}_n G_{ni\alpha} + \bar{C}_n g_{ni}e'_{m_b s}\bar{C}_n G_{ni\alpha} \end{pmatrix}\psi_{n\check{\alpha}} \\
& - \frac{1}{2n}\bar{K}^{-1}_{nd\beta}\sum_{j=1}^{k_\alpha}\psi_{n\check{\alpha}j}\sum_{i=1}^{n}\mathbf{E}\begin{pmatrix} 0 & G^{(j)'}_{ni\beta}\bar{C}'_n \\ \bar{C}_n G^{(j)}_{ni\beta} & \bar{C}_n g^{(j)}_{ni}g'_{ni}\bar{C}'_n + \bar{C}_n g_{ni}g^{(j)'}_{ni}\bar{C}'_n \end{pmatrix}\psi_{n\theta} \\
& - \frac{1}{2n}\bar{K}^{-1}_{nd\beta}\sum_{j=1}^{k_\beta}\psi_{n\theta j}\sum_{i=1}^{n}\mathbf{E}\begin{pmatrix} 0 & G^{(k_\alpha+j)'}_{ni\beta}\bar{C}'_n \\ \bar{C}_n G^{(k_\alpha+j)}_{ni\beta} & \bar{C}_n g^{(k_\alpha+j)}_{ni}g'_{ni}\bar{C}'_n + \bar{C}_n g_{ni}g^{(k_\alpha+j)'}_{ni}\bar{C}'_n \end{pmatrix}\psi_{n\theta} \\
& - \frac{1}{2n}\bar{K}^{-1}_{nd\beta}\sum_{s=1}^{m_b}\psi_{n\theta,k_\beta+s} \\
& \cdot\sum_{i=1}^{n}\mathbf{E}\begin{pmatrix} [G^{(k_\alpha+1)'}_{ni\beta}\bar{C}'_n e_{m_b s},\ldots,G^{(k_\alpha+k_\beta)'}_{ni\beta}\bar{C}'_n e_{m_b s}] & e'_{m_b s}\bar{C}_n g_{ni}G'_{ni\beta}\bar{C}'_n + G'_{ni\beta}\bar{C}'_n e_{m_b s}g'_{ni}\bar{C}'_n \\ e'_{m_b s}\bar{C}_n g_{ni}\bar{C}_n G_{ni\beta} + \bar{C}_n g_{ni}e'_{m_b s}\bar{C}_n G_{ni\beta} & -\rho_3(0)e'_{m_b s}\bar{C}_n g_{ni}\bar{C}_n g_{ni}g'_{ni}\bar{C}'_n \end{pmatrix}\psi_{n\theta}.
\end{aligned}
$$

Denote

$$
\begin{aligned}
\dot{\varphi}_{n\theta,\text{TGEL3}} = & -\bar{K}^{-1}_{nd\beta}\begin{pmatrix} 0 & \sqrt{n}(G'_{n\beta} - \bar{G}'_{n\beta})\bar{C}'_n \\ \sqrt{n}\bar{C}_n(G_{n\beta} - \bar{G}_{n\beta}) & \sqrt{n}\bar{C}_n(\Omega_n - \bar{\Omega}_n)\bar{C}'_n \end{pmatrix}\psi_{n\theta} \\
& - \frac{1}{2n}\bar{K}^{-1}_{nd\beta}\sum_{j=1}^{k_\beta}\psi_{n\theta j}\sum_{i=1}^{n}\mathbf{E}\begin{pmatrix} 0 & \bar{G}^{(k_\alpha+j)'}_{i\beta}\bar{C}'_n \\ \bar{C}_n G^{(k_\alpha+j)}_{ni\beta} & \bar{C}_n g^{(k_\alpha+j)}_{ni}g'_{ni}\bar{C}'_n + \bar{C}_n g_{ni}g^{(k_\alpha+j)'}_{ni}\bar{C}'_n \end{pmatrix}\psi_{n\theta} \\
& - \frac{1}{2n}\bar{K}^{-1}_{nd\beta}\sum_{s=1}^{m_b}\psi_{n\theta,k_\beta+s}\sum_{i=1}^{n}\mathbf{E}\begin{pmatrix} [G^{(k_\alpha+1)'}_{ni\beta}\bar{C}'_n e_{m_b s},\ldots,G^{(k_\alpha+k_\beta)'}_{ni\beta}\bar{C}'_n e_{m_b s}] & e'_{m_b s}\bar{C}_n g_{ni}G'_{ni\beta}\bar{C}'_n + G'_{ni\beta}\bar{C}'_n e_{m_b s}g'_{ni}\bar{C}'_n \\ e'_{m_b s}\bar{C}_n g_{ni}\bar{C}_n G_{ni\beta} + \bar{C}_n g_{ni}e'_{m_b s}\bar{C}_n G_{ni\beta} & -\rho_3(0)e'_{m_b s}\bar{C}_n g_{ni}\bar{C}_n g_{ni}g'_{ni}\bar{C}'_n \end{pmatrix}\psi_{n\theta},
\end{aligned}
$$

and

$$\ddot{\varphi}_{n\theta,\text{TGEL}3} = -\bar{K}_{nd\beta}^{-1}\begin{pmatrix} 0 \\ \sqrt{n}\psi_{n\tilde{C}}g_n \end{pmatrix} - \bar{K}_{nd\beta}^{-1}\begin{pmatrix} 0 \\ \sqrt{n}\bar{C}_n(G_{n\alpha}-\bar{G}_{n\alpha}) + \psi_{n\tilde{C}}\bar{G}_{n\alpha} \end{pmatrix}\psi_{n\check{\alpha}}$$
$$- \bar{K}_{nd\beta}^{-1}\begin{pmatrix} 0 & \bar{G}'_{n\beta}\psi'_{n\tilde{C}} \\ \psi_{n\tilde{C}}\bar{G}_{n\beta} & \psi_{n\tilde{C}}\bar{\Omega}_n\bar{C}'_n + \bar{C}_n\bar{\Omega}_n\psi'_{n\tilde{C}} \end{pmatrix}\psi_{n\theta}$$
$$- \frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\alpha}\psi_{n\check{\alpha}j}\begin{pmatrix} 0 \\ \bar{C}_n\bar{G}_{n\alpha}^{(j)} \end{pmatrix}\psi_{n\check{\alpha}} - \frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\beta}\psi_{n\theta j}\begin{pmatrix} 0 \\ \bar{C}_n\bar{G}_{n\alpha}^{(k_\alpha+j)} \end{pmatrix}\psi_{n\check{\alpha}} \tag{I.3}$$
$$- \frac{1}{2n}\bar{K}_{nd\beta}^{-1}\sum_{s=1}^{m_b}\psi_{n\theta,k_\beta+s}\sum_{i=1}^{n}\mathbf{E}\begin{pmatrix} [G_{ni\beta}^{(1)'}\bar{C}'_n e_{m_b s},\dots,G_{ni\beta}^{(k_\alpha)'}\bar{C}'_n e_{m_b s}] \\ e'_{m_b s}\bar{C}_n g_{ni}\bar{C}_n G_{ni\alpha} + \bar{C}_n g_{ni} e'_{m_b s}\bar{C}_n G_{ni\alpha} \end{pmatrix}\psi_{n\check{\alpha}}$$
$$- \frac{1}{2n}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\alpha}\psi_{n\check{\alpha}j}\sum_{i=1}^{n}\mathbf{E}\begin{pmatrix} 0 & G_{ni\beta}^{(j)'}\bar{C}'_n \\ \bar{C}_n G_{ni\beta}^{(j)} & \bar{C}_n g_{ni}^{(j)} g'_{ni}\bar{C}'_n + \bar{C}_n g_{ni} g_{ni}^{(j)'}\bar{C}'_n \end{pmatrix}\psi_{n\theta}.$$

Then $\varphi_{n\theta,\text{TGEL}3} = \dot{\varphi}_{n\theta,\text{TGEL}3} + \ddot{\varphi}_{n\theta,\text{TGEL}3}$, where $\ddot{\varphi}_{n\theta,\text{TGEL}3}$ involves terms due to the estimation of $\bar{C}_n$ and α_0, but $\dot{\varphi}_{n\theta,\text{TGEL}3}$ does not. The higher order bias of $\hat{\beta}_{\text{TGEL}3}$ is computed as the first k_β elements of $\frac{1}{n}\mathbf{E}(\varphi_{n\theta,\text{TGEL}3}) = \frac{1}{n}\mathbf{E}(\dot{\varphi}_{n\theta,\text{TGEL}3}) + \frac{1}{n}\mathbf{E}(\ddot{\varphi}_{n\theta,\text{TGEL}3})$. Note that $\mathbf{E}(\psi_{n\theta}\psi'_{n\theta}) = \begin{pmatrix} \bar{\Sigma}_{nd} & 0 \\ 0 & \bar{P}_{nd} \end{pmatrix}$. Then,

$$\mathbf{E}(\dot{\varphi}_{n\theta,\text{TGEL}3})$$
$$= \bar{K}_{nd\beta}^{-1}\mathbf{E}\left[\begin{pmatrix} 0 & \sqrt{n}G'_{n\beta}\bar{C}'_n \\ \sqrt{n}\bar{C}_n G_{n\beta} & \sqrt{n}\bar{C}_n\Omega_n\bar{C}'_n \end{pmatrix}\begin{pmatrix} \bar{H}_{nd} \\ \bar{P}_{nd} \end{pmatrix}\sqrt{n}\bar{C}_n g_n\right]$$
$$- \frac{1}{2n}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\beta}\sum_{i=1}^{n}\mathbf{E}\begin{pmatrix} 0 & G_{ni\beta}^{(k_\alpha+j)'}\bar{C}'_n \\ \bar{C}_n G_{ni\beta}^{(k_\alpha+j)} & \bar{C}_n g_{ni}^{(k_\alpha+j)} g'_{ni}\bar{C}'_n + \bar{C}_n g_{ni} g_{ni}^{(k_\alpha+j)'}\bar{C}'_n \end{pmatrix}\begin{pmatrix} \bar{\Sigma}_{nd} & 0 \\ 0 & \bar{P}_{nd} \end{pmatrix} e_{k_\theta j}$$
$$- \frac{1}{2n}\bar{K}_{nd\beta}^{-1}\sum_{s=1}^{m_b}\sum_{i=1}^{n}\mathbf{E}\begin{pmatrix} [G_{ni\beta}^{(k_\alpha+1)'}\bar{C}'_n e_{m_b s},\dots,G_{ni\beta}^{(k_\alpha+k_\beta)'}\bar{C}'_n e_{m_b s}] & e'_{m_b s}\bar{C}_n g_{ni}G'_{ni\beta}\bar{C}'_n + G'_{ni\beta}\bar{C}'_n e_{m_b s} g'_{ni}\bar{C}'_n \\ e'_{m_b s}\bar{C}_n g_{ni}\bar{C}_n G_{ni\beta} + \bar{C}_n g_{ni} e'_{m_b s}\bar{C}_n G_{ni\beta} & -\rho_3(0) e'_{m_b s}\bar{C}_n g_{ni}\bar{C}_n g_{ni} g'_{ni}\bar{C}'_n \end{pmatrix}\begin{pmatrix} \bar{\Sigma}_{nd} & 0 \\ 0 & \bar{P}_{nd} \end{pmatrix} e_{k_\theta,k_\beta+s}.$$

Thus, the first k_β elements of $\frac{1}{n}\mathbf{E}(\dot{\varphi}_{n\theta,\text{TGEL}3})$ are

$$-\bar{\Sigma}_{nd}\mathbf{E}(G'_{n\beta}\bar{C}'_n\bar{P}_{nd}\bar{C}_n g_n) + \bar{H}_{nd}\mathbf{E}(\bar{C}_n G_{n\beta}\bar{H}_{nd}\bar{C}_n g_n) + \bar{H}_{nd}\mathbf{E}(\bar{C}_n\Omega_n\bar{C}'_n\bar{P}_{nd}\bar{C}_n g_n)$$
$$- \frac{1}{2n}\sum_{j=1}^{k_\beta}\bar{H}_{nd}\bar{C}_n\bar{G}_{n\beta}^{(k_\alpha+j)}\bar{\Sigma}_{nd}e_{k_\beta j} + \frac{1}{2n^2}\sum_{s=1}^{m_b}\sum_{i=1}^{n}\bar{\Sigma}_{nd}\mathbf{E}(e'_{m_b s}\bar{C}_n g_{ni}G'_{ni\beta}\bar{C}'_n + G'_{ni\beta}\bar{C}'_n e_{m_b s} g'_{ni}\bar{C}'_n)\bar{P}_{nd}e_{m_b s}$$
$$+ \frac{\rho_3(0)}{2n^2}\sum_{s=1}^{m_b}\sum_{i=1}^{n}\bar{H}_{nd}e'_{m_b s}\bar{C}_n\mathbf{E}(g_{ni}\bar{C}_n g_{ni} g'_{ni})\bar{C}'_n\bar{P}_{nd}e_{m_b s}. \tag{I.4}$$

Since $\sum_{s=1}^{m_b} e'_{m_b s}\bar{C}_n g_{ni}\cdot G'_{ni\beta}\bar{C}'_n\bar{P}_{nd}e_{m_b s} = \sum_{s=1}^{m_b} G'_{ni\beta}\bar{C}'_n\bar{P}_{nd}e_{m_b s}\cdot e'_{m_b s}\bar{C}_n g_{ni} = G'_{ni\beta}\bar{C}'_n\bar{P}_{nd}\bar{C}_n g_{ni}$,

$$\sum_{s=1}^{m_b} G'_{ni\beta}\bar{C}'_n e_{m_b s}\cdot g'_{ni}\bar{C}'_n\bar{P}_{nd}e_{m_b s} = \sum_{s=1}^{m_b} G'_{ni\beta}\bar{C}'_n e_{m_b s}\cdot e'_{m_b s}\bar{P}_{nd}\bar{C}_n g_{ni} = G'_{ni\beta}\bar{C}'_n\bar{P}_{nd}\bar{C}_n g_{ni},$$

and $\sum_{s=1}^{m_b} e'_{m_b s}\bar{C}_n g_{ni}\cdot\bar{C}_n g_{ni}g'_{ni}\bar{C}'_n\bar{P}_{nd}e_{m_b s} = \sum_{s=1}^{m_b}\bar{C}_n g_{ni}g'_{ni}\bar{C}'_n\bar{P}_{nd}e_{m_b s}\cdot e'_{m_b s}\bar{C}_n g_{ni} = \bar{C}_n g_{ni}g'_{ni}\bar{C}'_n\bar{P}_{nd}\bar{C}_n g_{ni}$, (I.4) becomes

$$\begin{aligned}&-\bar{\Sigma}_{nd}\,\mathbf{E}(G'_{n\beta}\bar{C}'_n\bar{P}_{nd}\bar{C}_n g_n) + \frac{1}{n^2}\bar{\Sigma}_{nd}\sum_{i=1}^{n}\mathbf{E}(G'_{ni\beta}\bar{C}'_n\bar{P}_{nd}\bar{C}_n g_{ni}) + \bar{H}_{nd}\,\mathbf{E}(\bar{C}_n G_{n\beta}\bar{H}_{nd}\bar{C}_n g_n)\\ &+\bar{H}_{nd}\,\mathbf{E}(\bar{C}_n\Omega_n\bar{C}'_n\bar{P}_{nd}\bar{C}_n g_n) + \frac{\rho_3(0)}{2n^2}\sum_{i=1}^{n}\bar{H}_{nd}\,\mathbf{E}(\bar{C}_n g_{ni}g'_{ni}\bar{C}'_n\bar{P}_{nd}\bar{C}_n g_{ni}) - \frac{1}{2n}\sum_{j=1}^{k_\beta}\bar{H}_{nd}\bar{C}_n\bar{G}_{n\beta}^{(k_\alpha+j)}\bar{\Sigma}_{nd}e_{k_\beta j}.\end{aligned} \tag{I.5}$$

To compute $\mathbf{E}(\ddot{\varphi}_{n\theta,\text{TGEL3}})$, denote $\psi_{n\theta} = (\psi'_{n\beta},\psi'_{n\mu})'$, where $\psi_{n\beta} = -\bar{H}_{nd}\sqrt{n}\bar{C}_n g_n$ and $\psi_{n\mu} = -\bar{P}_{nd}\sqrt{n}\bar{C}_n g_n$. It is straightforward to show that $\sum_{j=1}^{k_\beta}\bar{C}_n\bar{G}_{n\alpha}^{(k_\alpha+j)}\psi_{n\check{\alpha}}\psi_{n\theta j} = \sum_{j=1}^{k_\alpha}\bar{C}_n\bar{G}_{n\beta}^{(j)}\psi_{n\beta}\psi_{n\check{\alpha}j}$,

$$\sum_{s=1}^{m_b}[\bar{G}_{n\beta}^{(1)'}\bar{C}'_n e_{m_b s},\ldots,\bar{G}_{n\beta}^{(k_\alpha)'}\bar{C}'_n e_{m_b s}]\psi_{n\check{\alpha}}\psi_{n\theta,k_\beta+s} = \sum_{j=1}^{k_\alpha}\bar{G}_{n\beta}^{(j)'}\bar{C}'_n\psi_{n\mu}\psi_{n\check{\alpha}j},$$

and $\sum_{s=1}^{m_b}\mathbf{E}(e'_{m_b s}\bar{C}_n g_{ni}\bar{C}_n G_{ni\alpha}+\bar{C}_n g_{ni}e'_{m_b s}\bar{C}_n G_{ni\alpha})\psi_{n\check{\alpha}}\psi_{n\theta,k_\beta+s} = \sum_{j=1}^{k_\alpha}\bar{C}_n\,\mathbf{E}(g_{ni}^{(j)}g'_{ni}+g_{ni}g_{ni}^{(j)'})\bar{C}'_n\psi_{n\mu}\psi_{n\check{\alpha}j}$. Then the first k_β elements of $\frac{1}{n}\mathbf{E}(\ddot{\varphi}_{n\theta,\text{TGEL3}})$ form the vector

$$\begin{aligned}&\frac{1}{n}\mathbf{E}\Big\{\bar{\Sigma}_{nd}\bar{G}'_{n\beta}\psi'_{n\tilde{C}}\psi_{n\mu} + \bar{\Sigma}_{nd}\sum_{j=1}^{k_\alpha}\bar{G}_{n\beta}^{(j)'}\bar{C}'_n\psi_{n\mu}\psi_{n\check{\alpha}j} - \bar{H}_{nd}\sqrt{n}\psi_{n\tilde{C}}g_n - \bar{H}_{nd}\sqrt{n}\bar{C}_n(G_{n\alpha}-\bar{G}_{n\alpha})\psi_{n\check{\alpha}}\\ &-\bar{H}_{nd}\psi_{n\tilde{C}}\bar{G}_{n\alpha}\psi_{n\check{\alpha}} - \bar{H}_{nd}\psi_{n\tilde{C}}\bar{G}_{n\beta}\psi_{n\beta} - \bar{H}_{nd}(\psi_{n\tilde{C}}\bar{\Omega}_n\bar{C}'_n + \bar{C}_n\bar{\Omega}_n\psi'_{n\tilde{C}})\psi_{n\mu} - \frac{1}{2}\bar{H}_{nd}\sum_{j=1}^{k_\alpha}\bar{C}_n\bar{G}_{n\alpha}^{(j)}\psi_{n\check{\alpha}}\psi_{n\check{\alpha}j}\\ &-\bar{H}_{nd}\sum_{j=1}^{k_\alpha}\bar{C}_n\bar{G}_{n\beta}^{(j)}\psi_{n\beta}\psi_{n\check{\alpha}j} - \frac{1}{n}\bar{H}_{nd}\sum_{j=1}^{k_\alpha}\sum_{i=1}^{n}\bar{C}_n\,\mathbf{E}(g_{ni}^{(j)}g'_{ni} + g_{ni}g_{ni}^{(j)'})\bar{C}'_n\psi_{n\mu}\psi_{n\check{\alpha}j}\Big\}.\end{aligned} \tag{I.6}$$

By (I.5) and (I.6), the bias of $\hat{\beta}_{\text{TGEL3}}$ is $B_{nd}^I + (B_{nd}^\Omega + \frac{\rho_3(0)}{2}\tilde{B}_{nd}^\Omega) + (B_{nd}^G - \tilde{B}_{nd}^G) + B_{nd}^{C-g} + B_{nd}^{C-\Omega} + B_{nd}^{C-G} + B_{nd}^\alpha$.

(ii) For $\hat{\beta}_{\text{TGMM3}}$, by Lemma H.5$(i)$, $\tilde{\Omega}_n = \bar{\Omega}_n + o_p(1)$. Since $\bar{C}_n$ has full row rank and $\lim_{n\to\infty}\bar{\Omega}_n$ is nonsingular by Assumption 1(vi), $\lim_{n\to\infty}\bar{C}_n\bar{\Omega}_n\bar{C}'_n$ is nonsingular. It is shown in the proof of Lemma H.1 that $\sup_{\beta\in\mathcal{B}}\|g_n(\check{\alpha},\beta) - \bar{g}_n(\alpha_0,\beta)\| = o_p(1)$ and shown in the proof of Lemma H.2 that

$\tilde{C}_n = \bar{C}_n + o_p(1)$. It follows that $\sup_{\beta\in\mathcal{B}} \|\tilde{C}_n g_n(\check{\alpha},\beta) - \bar{C}_n \bar{g}_n(\alpha_0,\beta)\| = o_p(1)$. Then by the nonsingularity of $\lim_{n\to\infty} \bar{C}_n \bar{\Omega}_n \bar{C}_n'$, the uniform convergence $\sup_{\beta\in\mathcal{B}} \|\tilde{C}_n g_n(\check{\alpha},\beta) - \bar{C}_n \bar{g}_n(\alpha_0,\beta)\| = o_p(1)$, and the identification Assumption 2(v), the consistency of $\hat{\beta}_{\text{TGMM3}}$ to β_0 follows. Then, by the MVT, the first order condition for $\hat{\beta}_{\text{TGEL3}}$ becomes

$$
\begin{aligned}
0 &= G_{n\beta}'(\check{\alpha},\hat{\beta}_{\text{TGMM3}})\tilde{C}_n'(\tilde{C}_n\tilde{\Omega}_n\tilde{C}_n')^{-1}\tilde{C}_n g_n(\check{\alpha},\hat{\beta}_{\text{TGMM3}}) \\
&= G_{n\beta}'(\check{\alpha},\hat{\beta}_{\text{TGMM3}})\tilde{C}_n'(\tilde{C}_n\tilde{\Omega}_n\tilde{C}_n')^{-1}\tilde{C}_n\Big[g_n(\check{\alpha},\beta_0) + G_{n\beta}(\ddot{\alpha},\ddot{\beta})(\hat{\beta}_{\text{TGMM3}} - \beta_0) + G_{n\alpha}(\ddot{\alpha},\ddot{\beta})(\hat{\alpha}_{\text{TGMM3}} - \alpha_0)\Big],
\end{aligned}
$$

where $\ddot{\alpha}$ lies between α_0 and $\check{\alpha}$, and $\ddot{\beta}$ lies between β_0 and $\hat{\beta}_{\text{TGMM3}}$. Thus,

$$
\sqrt{n}(\hat{\beta}_{\text{TGMM3}} - \beta_0) = -(\bar{D}_{n\beta}'\bar{\Omega}_{nd}^{-1}\bar{D}_{n\beta})^{-1}\bar{D}_{n\beta}'\bar{\Omega}_{nd}^{-1}\bar{C}_n\sqrt{n}g_n(\alpha_0,\beta_0) + o_p(1) = O_p(1). \tag{I.7}
$$

Let $\hat{\mu}_{\text{TGMM3}} = -(\tilde{C}_n\tilde{\Omega}_n\tilde{C}_n')^{-1}\tilde{C}_n g_n(\check{\alpha},\hat{\beta}_{\text{TGMM3}}) = O_p(n^{-1/2})$. Using Lemma H.5($iv$), the first order condition for $\hat{\beta}_{\text{TGMM3}}$ can be written as

$$
\begin{aligned}
0 &= -\begin{pmatrix} G_{n\beta}'(\check{\alpha},\hat{\beta}_{\text{TGMM3}})\tilde{C}_n'\hat{\mu}_{\text{TGMM3}} \\ \tilde{C}_n g_n(\check{\alpha},\hat{\beta}_{\text{TGMM3}}) + \tilde{C}_n\tilde{\Omega}_n\tilde{C}_n'\hat{\mu}_{\text{TGMM3}} \end{pmatrix} \\
&= -\begin{pmatrix} G_{n\beta}'(\check{\alpha},\hat{\beta}_{\text{TGMM3}})\tilde{C}_n'\hat{\mu}_{\text{TGMM3}} \\ \tilde{C}_n g_n(\check{\alpha},\hat{\beta}_{\text{TGMM3}}) + (\bar{\Omega}_{nd} + n^{-1/2}\psi_{n\tilde{\Omega}_d})\hat{\mu}_{\text{TGMM3}} \end{pmatrix} + O_p(n^{-3/2}).
\end{aligned} \tag{I.8}
$$

Denote $\hat{\theta}_{\text{TGMM3}} = (\hat{\beta}_{\text{TGMM3}}', \hat{\mu}_{\text{TGMM3}}')'$. By a second order Taylor expansion,

$$
\begin{aligned}
0 = &-\begin{pmatrix} 0 \\ \tilde{C}_n g_n \end{pmatrix} - \begin{pmatrix} 0 \\ \tilde{C}_n G_{n\alpha} \end{pmatrix}(\check{\alpha} - \alpha_0) - \begin{pmatrix} 0 & G_{n\beta}'\tilde{C}_n' \\ \tilde{C}_n G_{n\beta} & \bar{\Omega}_{nd} + n^{-1/2}\psi_{n\tilde{\Omega}_d} \end{pmatrix}(\hat{\theta}_{\text{TGMM3}} - \theta_0) \\
&- \frac{1}{2}\sum_{j=1}^{k_\alpha}(\check{\alpha}_j - \alpha_{j0})\begin{pmatrix} 0 \\ \tilde{C}_n G_{n\alpha}^{(j)} \end{pmatrix}(\check{\alpha} - \alpha_0) - \frac{1}{2}\sum_{j=1}^{k_\beta}(\hat{\theta}_{j,\text{TGMM3}} - \theta_{j0})\begin{pmatrix} 0 \\ \tilde{C}_n G_{n\alpha}^{(k_\alpha+j)} \end{pmatrix}(\check{\alpha} - \alpha_0) \\
&- \frac{1}{2}\sum_{j=k_\beta+1}^{k_\theta}(\hat{\theta}_{j,\text{TGMM3}} - \theta_{j0})\begin{pmatrix} [G_{n\beta}^{(1)\prime}\tilde{C}_n' e_{m_b,j-k_\beta}, \ldots, G_{n\beta}^{(k_\alpha)\prime}\tilde{C}_n' e_{m_b,j-k_\beta}] \\ 0 \end{pmatrix}(\check{\alpha} - \alpha_0) \\
&- \frac{1}{2}\sum_{j=1}^{k_\alpha}(\check{\alpha}_j - \alpha_{j0})\begin{pmatrix} 0 & G_{n\beta}^{(j)\prime}\tilde{C}_n' \\ \tilde{C}_n G_{n\beta}^{(j)} & 0 \end{pmatrix}(\hat{\theta}_{\text{TGMM3}} - \theta_0) - \frac{1}{2}\sum_{j=1}^{k_\theta}(\hat{\theta}_{j,\text{TGMM3}} - \theta_{j0})\hat{K}_{nd\beta,j}(\hat{\theta}_{\text{TGMM3}} - \theta_0) \\
&+ O_p(n^{-3/2}),
\end{aligned}
$$

where $\hat{K}_{nd\beta,j} = \begin{pmatrix} 0 & G_{n\beta}^{(k_\alpha+j)\prime}\tilde{C}_n' \\ \tilde{C}_n G_{n\beta}^{(k_\alpha+j)} & 0 \end{pmatrix}$ for $1 \le j \le k_\beta$, and $\hat{K}_{nd\beta,j} = \begin{pmatrix} [G_{n\beta}^{(k_\alpha+1)\prime}\tilde{C}_n' e_{m_b,j-k_\beta}, \ldots, G_{n\beta}^{(k_\alpha+k_\beta)\prime}\tilde{C}_n' e_{m_b,j-k_\beta}] & 0 \\ 0 & 0 \end{pmatrix}$

for $k_\beta + 1 \le j \le k_\theta$. Then,

$$
\begin{aligned}
&\sqrt{n}(\hat{\theta}_{\text{TGMM3}} - \theta_0) \\
&= -\bar{K}_{nd\beta}^{-1}\begin{pmatrix} 0 \\ \sqrt{n}\tilde{C}_n g_n \end{pmatrix} - \bar{K}_{nd\beta}^{-1}\begin{pmatrix} 0 \\ \tilde{C}_n G_{n\alpha} \end{pmatrix}\sqrt{n}(\check{\alpha} - \alpha_0) \\
&\quad - \bar{K}_{nd\beta}^{-1}\left[\begin{pmatrix} 0 & G_{n\beta}'\tilde{C}_n' \\ \tilde{C}_n G_{n\beta} & \bar{\Omega}_{nd} + n^{-1/2}\psi_{n\tilde{\Omega}_d} \end{pmatrix} - \bar{K}_{nd\beta}\right]\sqrt{n}(\hat{\theta}_{\text{TGMM3}} - \theta_0) \\
&\quad - \frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\alpha}(\check{\alpha}_j - \alpha_{j0})\begin{pmatrix} 0 \\ \tilde{C}_n G_{n\alpha}^{(j)} \end{pmatrix}\sqrt{n}(\check{\alpha} - \alpha_0) - \frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\beta}(\hat{\theta}_{j,\text{TGMM3}} - \theta_{j0})\begin{pmatrix} 0 \\ \tilde{C}_n G_{n\alpha}^{(k_\alpha+j)} \end{pmatrix}\sqrt{n}(\check{\alpha} - \alpha_0) \\
&\quad - \frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=k_\beta+1}^{k_\theta}(\hat{\theta}_{j,\text{TGMM3}} - \theta_{j0})\begin{pmatrix} [G_{n\beta}^{(1)'}\tilde{C}_n' e_{m_b,j-k_\beta}, \ldots, G_{n\beta}^{(k_\alpha)'}\tilde{C}_n' e_{m_b,j-k_\beta}] \\ 0 \end{pmatrix}\sqrt{n}(\check{\alpha} - \alpha_0) \\
&\quad - \frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\alpha}(\check{\alpha}_j - \alpha_{j0})\begin{pmatrix} 0 & G_{n\beta}^{(j)'}\tilde{C}_n' \\ \tilde{C}_n G_{n\beta}^{(j)} & 0 \end{pmatrix}\sqrt{n}(\hat{\theta}_{\text{TGMM3}} - \theta_0) \\
&\quad - \frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\theta}(\hat{\theta}_{j,\text{TGMM3}} - \theta_{j0})\hat{K}_{nd\beta,j}\sqrt{n}(\hat{\theta}_{\text{TGMM3}} - \theta_0) + O_p(n^{-1}).
\end{aligned} \tag{I.9}
$$

Since $\sqrt{n}(\hat{\theta}_{\text{TGMM3}}-\theta_0) = O_p(1)$, $\sqrt{n}(\check{\alpha}-\alpha_0) = O_p(1)$, $\tilde{C}_n G_n - \bar{C}_n\bar{G}_n = o_p(1)$ and $\bar{C}_n\bar{G}_{n\alpha} = 0$, $\sqrt{n}(\hat{\theta}_{\text{TGMM3}} - \theta_0) = \psi_{n\theta} + o_p(1)$, where $\psi_{n\theta} = -\bar{K}_{nd\beta}^{-1}\begin{pmatrix} 0 \\ \sqrt{n}\bar{C}_n g_n \end{pmatrix} = -\begin{pmatrix} \bar{H}_{nd} \\ \bar{P}_{nd} \end{pmatrix}\sqrt{n}\bar{C}_n g_n$. Substituting $\sqrt{n}(\hat{\theta}_{\text{TGMM3}} - \theta_0) = \psi_{n\theta} + o_p(1)$ into (I.9) yields $\sqrt{n}(\hat{\theta}_{\text{TGMM3}} - \theta_0) = \psi_{n\theta} + n^{-1/2}\varphi_{n\theta,\text{TGMM3}} + O_p(n^{-1})$, where

$$
\begin{aligned}
\varphi_{n\theta,\text{TGMM3}} &= -\bar{K}_{nd\beta}^{-1}\begin{pmatrix} 0 \\ \sqrt{n}\psi_{n\tilde{C}} g_n \end{pmatrix} - \bar{K}_{nd\beta}^{-1}\begin{pmatrix} 0 \\ \sqrt{n}\bar{C}_n(G_{n\alpha} - \bar{G}_{n\alpha}) + \psi_{n\tilde{C}}\bar{G}_{n\alpha} \end{pmatrix}\psi_{n\check{\alpha}} \\
&\quad - \bar{K}_{nd\beta}^{-1}\begin{pmatrix} 0 & \sqrt{n}(G_{n\beta}' - \bar{G}_{n\beta}')\bar{C}_n' + \bar{G}_{n\beta}'\psi_{n\tilde{C}}' \\ \sqrt{n}\bar{C}_n(G_{n\beta} - \bar{G}_{n\beta}) + \psi_{n\tilde{C}}\bar{G}_{n\beta} & \psi_{n\tilde{\Omega}_d} \end{pmatrix}\psi_{n\theta} \\
&\quad - \frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\alpha}\psi_{n\check{\alpha}j}\begin{pmatrix} 0 \\ \bar{C}_n\bar{G}_{n\alpha}^{(j)} \end{pmatrix}\psi_{n\check{\alpha}} - \frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\beta}\psi_{n\theta j}\begin{pmatrix} 0 \\ \bar{C}_n\bar{G}_{n\alpha}^{(k_\alpha+j)} \end{pmatrix}\psi_{n\check{\alpha}} \\
&\quad - \frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=k_\beta+1}^{k_\theta}\psi_{n\theta j}\begin{pmatrix} [\bar{G}_{n\beta}^{(1)'}\bar{C}_n' e_{m_b,j-k_\beta}, \ldots, \bar{G}_{n\beta}^{(k_\alpha)'}\bar{C}_n' e_{m_b,j-k_\beta}] \\ 0 \end{pmatrix}\psi_{n\check{\alpha}} \\
&\quad - \frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\alpha}\psi_{n\check{\alpha}j}\begin{pmatrix} 0 & \bar{G}_{n\beta}^{(j)'}\bar{C}_n' \\ \bar{C}_n\bar{G}_{n\beta}^{(j)} & 0 \end{pmatrix}\psi_{n\theta} - \frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\theta}\psi_{n\theta j}\bar{K}_{nd\beta,j}\psi_{n\theta},
\end{aligned}
$$

with $\bar{K}_{nd\beta,j} = \begin{pmatrix} 0 & \bar{G}_{n\beta}^{(k_\alpha+j)'}\bar{C}_n' \\ \bar{C}_n\bar{G}_{n\beta}^{(k_\alpha+j)} & 0 \end{pmatrix}$ for $1 \le j \le k_\beta$, and $\bar{K}_{nd\beta,j} = \begin{pmatrix} [\bar{G}_{n\beta}^{(k_\alpha+1)'}\bar{C}_n' e_{m_b,j-k_\beta}, \ldots, \bar{G}_{n\beta}^{(k_\alpha+k_\beta)'}\bar{C}_n' e_{m_b,j-k_\beta}] & 0 \\ 0 & 0 \end{pmatrix}$

for $k_\beta+1 \le j \le k_\theta$. Using the expression of $\psi_{n\tilde{\Omega}_d}$ in Lemma H.5(iv), $\varphi_{n\theta,\text{TGMM3}} = \dot{\varphi}_{n\theta,\text{TGMM3}} + \ddot{\varphi}_{n\theta,\text{TGMM3}}$, where

$$\dot{\varphi}_{n\theta,\text{TGMM3}} = -\bar{K}_{nd\beta}^{-1}\begin{pmatrix} 0 & \sqrt{n}(G'_{n\beta}-\bar{G}'_{n\beta})\bar{C}'_n \\ \sqrt{n}\bar{C}_n(G_{n\beta}-\bar{G}_{n\beta}) & \dot{\psi}_{n\tilde{\Omega}_d}\end{pmatrix}\psi_{n\theta} - \frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\theta}\psi_{n\theta j}\bar{K}_{nd\beta,j}\psi_{n\theta},$$

and

$$\begin{aligned}\ddot{\varphi}_{n\theta,\text{TGMM3}} = &-\bar{K}_{nd\beta}^{-1}\begin{pmatrix}0\\ \sqrt{n}\psi_{n\tilde{C}}g_n\end{pmatrix} - \bar{K}_{nd\beta}^{-1}\begin{pmatrix}0\\ \sqrt{n}\bar{C}_n(G_{n\alpha}-\bar{G}_{n\alpha}) + \psi_{n\tilde{C}}\bar{G}_{n\alpha}\end{pmatrix}\psi_{n\check{\alpha}}\\
&-\bar{K}_{nd\beta}^{-1}\begin{pmatrix}0 & \bar{G}'_{n\beta}\psi'_{n\tilde{C}}\\ \psi_{n\tilde{C}}\bar{G}_{n\beta} & \ddot{\psi}_{n\tilde{\Omega}_d}\end{pmatrix}\psi_{n\theta}\\
&-\frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\alpha}\psi_{n\check{\alpha}j}\begin{pmatrix}0\\ \bar{C}_n\bar{G}_{n\alpha}^{(j)}\end{pmatrix}\psi_{n\check{\alpha}} - \frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\beta}\psi_{n\theta j}\begin{pmatrix}0\\ \bar{C}_n\bar{G}_{n\alpha}^{(k_\alpha+j)}\end{pmatrix}\psi_{n\check{\alpha}}\\
&-\frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=k_\beta+1}^{k_\theta}\psi_{n\theta j}\begin{pmatrix}[\bar{G}_{n\beta}^{(1)'}\bar{C}'_n e_{m_b,j-k_\beta},\ldots,\bar{G}_{n\beta}^{(k_\alpha)'}\bar{C}'_n e_{m_b,j-k_\beta}]\\ 0\end{pmatrix}\psi_{n\check{\alpha}}\\
&-\frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\alpha}\psi_{n\check{\alpha}j}\begin{pmatrix}0 & \bar{G}_{n\beta}^{(j)'}\bar{C}'_n\\ \bar{C}_n\bar{G}_{n\beta}^{(j)} & 0\end{pmatrix}\psi_{n\theta}.\end{aligned} \tag{I.10}$$

$\ddot{\varphi}_{n\theta,\text{TGMM3}}$ contains terms due to the estimation of α_0 and $\bar{C}_n$, but $\dot{\varphi}_{n\theta,\text{TGMM3}}$ does not. Then,

$$\begin{aligned}\mathbf{E}(\dot{\varphi}_{n\theta,\text{TGMM3}}) = &\bar{K}_{nd\beta}^{-1}\mathbf{E}\left[\begin{pmatrix}0 & \sqrt{n}G'_{n\beta}\bar{C}'_n\\ \sqrt{n}\bar{C}_nG_{n\beta} & \dot{\psi}_{n\tilde{\Omega}_d}\end{pmatrix}\begin{pmatrix}\bar{H}_{nd}\\ \bar{P}_{nd}\end{pmatrix}\sqrt{n}\bar{C}_ng_n\right]\\
&-\frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\beta}\begin{pmatrix}0 & \bar{G}_{n\beta}^{(k_\alpha+j)'}\bar{C}'_n\\ \bar{C}_n\bar{G}_{n\beta}^{(k_\alpha+j)} & 0\end{pmatrix}\begin{pmatrix}\bar{\Sigma}_{nd} & 0\\ 0 & \bar{P}_{nd}\end{pmatrix}e_{k_\theta,j}\\
&-\frac{1}{2}\bar{K}_{nd\beta}^{-1}\sum_{j=k_\beta+1}^{k_\theta}\begin{pmatrix}[\bar{G}_{n\beta}^{(k_\alpha+1)'}\bar{C}'_n e_{m_b,j-k_\beta},\ldots,\bar{G}_{n\beta}^{(k_\alpha+k_\beta)'}\bar{C}'_n e_{m_b,j-k_\beta}] & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\bar{\Sigma}_{nd} & 0\\ 0 & \bar{P}_{nd}\end{pmatrix}e_{k_\theta,j}.\end{aligned}$$

Using $\bar{K}_{nd\beta}^{-1} = \begin{pmatrix}-\bar{\Sigma}_{nd} & \bar{H}_{nd}\\ \bar{H}'_{nd} & \bar{P}_{nd}\end{pmatrix}$ and the expression of $\dot{\psi}_{n\tilde{\Omega}_d}$ in Lemma H.5(iv), the first k_β elements of $\frac{1}{n}\mathbf{E}(\dot{\varphi}_{n\theta,\text{TGMM3}})$ form the vector

$$\begin{aligned}&-\bar{\Sigma}_{nd}\mathbf{E}(G'_{n\beta}\bar{C}'_n\bar{P}_{nd}\bar{C}_ng_n) + \bar{H}_{nd}\mathbf{E}(\bar{C}_nG_{n\beta}\bar{H}_{nd}\bar{C}_ng_n) + \bar{H}_{nd}\bar{C}_n\mathbf{E}(\Omega_n\bar{C}'_n\bar{P}_{nd}\bar{C}_ng_n)\\
&-\frac{1}{n^2}\bar{H}_{nd}\bar{C}_n\sum_{j=1}^{k_\beta}\sum_{i=1}^{n}\mathbf{E}(g_{ni}g_{ni}^{(k_\alpha+j)'} + g_{ni}^{(k_\alpha+j)}g'_{ni})\bar{C}'_n\bar{P}_{nd}\bar{C}_n\bar{\Omega}_{nb}\bar{H}'_{nW\beta}e_{k_\beta j} - \frac{1}{2n}\bar{H}_{nd}\sum_{j=1}^{k_\beta}\bar{C}_n\bar{G}_{n\beta}^{(k_\alpha+j)}\bar{\Sigma}_{nd}e_{k_\beta j}.\end{aligned} \tag{I.11}$$

To compute the first k_β elements of $\frac{1}{n}\mathbf{E}(\ddot{\varphi}_{n\theta,\text{TGMM3}})$, we compare the expressions for $\ddot{\varphi}_{n\theta,\text{TGMM3}}$ and $\ddot{\varphi}_{n\theta,\text{TGEL3}}$. By (I.3) and (I.10),

$$
\begin{aligned}
\ddot{\varphi}_{n\theta,\text{TGMM3}} - \ddot{\varphi}_{n\theta,\text{TGEL3}} &= -\bar{K}_{nd\beta}^{-1}\begin{pmatrix} 0 & 0 \\ 0 & B_n \end{pmatrix}\psi_{n\theta} \\
&\quad + \frac{1}{2n}\bar{K}_{nd\beta}^{-1}\sum_{s=1}^{m_b}\sum_{i=1}^{n}\mathbf{E}\begin{pmatrix} 0 \\ e'_{m_b s}\bar{C}_n g_{ni}\bar{C}_n G_{ni\alpha} + \bar{C}_n g_{ni} e'_{m_b s}\bar{C}_n G_{ni\alpha} \end{pmatrix}\psi_{n\check{\alpha}}\psi_{n\theta,k_\beta+s} \\
&\quad + \frac{1}{2n}\bar{K}_{nd\beta}^{-1}\sum_{j=1}^{k_\alpha}\sum_{i=1}^{n}\mathbf{E}\begin{pmatrix} 0 & 0 \\ 0 & \bar{C}_n g_{ni}^{(j)} g'_{ni}\bar{C}'_n + \bar{C}_n g_{ni} g_{ni}^{(j)'}\bar{C}'_n \end{pmatrix}\psi_{n\theta}\psi_{n\check{\alpha}j} \\
&= \bar{K}_{nd\beta}^{-1}\begin{pmatrix} 0 \\ \frac{1}{n}\bar{C}_n\sum_{j=1}^{k_\beta}\sum_{i=1}^{n}\mathbf{E}(g_{ni}g_{ni}^{(k_\alpha+j)'} + g_{ni}^{(k_\alpha+j)}g'_{ni})\bar{C}'_n e'_{k_\beta j}\bar{H}_{nW\beta}\bar{G}_{nb\alpha}\psi_{n\check{\alpha}} \end{pmatrix}\psi_{n\mu},
\end{aligned}
$$

where

$$
\begin{aligned}
B_n &= \frac{1}{n}\bar{C}_n\sum_{j=1}^{k_\alpha}\sum_{i=1}^{n}\mathbf{E}(g_{ni}g_{ni}^{(j)'} + g_{ni}^{(j)}g'_{ni})\psi_{n\check{\alpha}j}\bar{C}'_n \\
&\quad - \frac{1}{n}\bar{C}_n\sum_{j=1}^{k_\beta}\sum_{i=1}^{n}\mathbf{E}(g_{ni}g_{ni}^{(k_\alpha+j)'} + g_{ni}^{(k_\alpha+j)}g'_{ni})\bar{C}'_n e'_{k_\beta j}\bar{H}_{nW\beta}\bar{G}_{nb\alpha}\psi_{n\check{\alpha}}
\end{aligned}
$$

and $\psi_{n\mu} = -\bar{P}_{nd}\sqrt{n}\bar{C}_n g_n$. This term arises from the use of $\check{\alpha}$ in deriving $\tilde{\beta}$. The first k_β elements of $\frac{1}{n}\mathbf{E}(\ddot{\varphi}_{n\theta,\text{TGMM3}} - \ddot{\varphi}_{n\theta,\text{TGEL3}})$ are

$$
B_{nd}^{W\alpha} = -\frac{1}{n^2}\bar{H}_{nd}\bar{C}_n\sum_{j=1}^{k_\beta}\sum_{i=1}^{n}\mathbf{E}(g_{ni}g_{ni}^{(k_\alpha+j)'} + g_{ni}^{(k_\alpha+j)}g'_{ni})\bar{C}'_n\bar{P}_{nd}\bar{C}_n\mathbf{E}(\sqrt{n}g_n\psi'_{n\check{\alpha}})\bar{G}'_{nb\alpha}\bar{H}'_{nW\beta}e_{k_\beta j}. \tag{I.12}
$$

Therefore, by (I.6), (I.11) and (I.12), the bias of $\hat{\beta}_{\text{TGMM3}}$ is $B_{nd}^{I} + B_{nd}^{\Omega} + B_{nd}^{G} + B_{nd}^{W} + B_{nd}^{C-g} + B_{nd}^{C-\Omega} + B_{nd}^{C-G} + B_{nd}^{\alpha} + B_{nd}^{W\alpha}$.

(iii) We shall use some part of the proof of (i), so we use the following similar notations: $v_{ni}(\alpha,\theta) = \mu'\check{C}_n g_{ni}(\alpha,\beta)$, $h_{ni}(\alpha,\theta) = \begin{pmatrix} G'_{ni\beta}(\alpha,\beta)\check{C}'_n\mu \\ \check{C}_n g_{ni}(\alpha,\beta) \end{pmatrix}$, $m_{ni}(\alpha,\theta) = \rho_1(v_{ni}(\alpha,\theta))h_{ni}(\alpha,\theta)$, and $\hat{m}_n(\alpha,\theta) = \frac{1}{n}\sum_{i=1}^{n} m_{ni}(\alpha,\theta)$. In addition, let $\dot{\alpha}_{n\beta}(\beta) = \frac{\partial\dot{\alpha}_n(\beta)}{\partial\beta'}$ and $\hat{\Delta}_{ni} = \dot{\alpha}'_{n\beta}(\hat{\beta}_{\text{TGEL4}})G'_{ni\alpha}(\dot{\alpha}_n(\hat{\beta}_{\text{TGEL4}}),\hat{\beta}_{\text{TGEL4}})\check{C}'_n\hat{\mu}_{\text{TGEL4}}$. The first order condition for $\hat{\theta}_{\text{TGEL4}}$ is

$$
\hat{m}_n(\dot{\alpha}_n(\hat{\beta}_{\text{TGEL4}}),\hat{\theta}_{\text{TGEL4}}) + \frac{1}{n}\sum_{i=1}^{n}\rho_1(v_{ni}(\dot{\alpha}_n(\hat{\beta}_{\text{TGEL4}}),\hat{\theta}_{\text{TGEL4}}))\begin{pmatrix} \hat{\Delta}_{ni} \\ 0 \end{pmatrix} = 0. \tag{I.13}
$$

By Theorem 1, $\sqrt{n}(\hat{\theta}_{\text{TGEL4}}-\theta_0)=O_p(1)$. Thus, by a second order Taylor expansion of $\rho_1(\cdot)$,

$$\frac{1}{n}\sum_{i=1}^{n}\rho_1(v_{ni}(\dot{\alpha}_n(\hat{\beta}_{\text{TGEL4}}),\hat{\theta}_{\text{TGEL4}}))\hat{\Delta}_{ni}=\hat{A}+O_p(n^{-3/2}),$$

where $\hat{A}=\frac{1}{n}\sum_{i=1}^{n}\rho_1(0)\hat{\Delta}_{ni}+\frac{1}{n}\sum_{i=1}^{n}\rho_2(0)v_{ni}(\dot{\alpha}_n(\hat{\beta}_{\text{TGEL4}}),\hat{\theta}_{\text{TGEL4}})\hat{\Delta}_{ni}$. Comparing the first order conditions (I.13) for $\hat{\theta}_{\text{TGEL4}}$ and (I.1) for $\hat{\theta}_{\text{TGEL}}$, we have

$$\sqrt{n}(\hat{\theta}_{\text{TGEL4}}-\theta_0)=\hat{B}+\sqrt{n}\bar{K}_{nd\beta}^{-1}\begin{pmatrix}\hat{A}\\0\end{pmatrix}+O_p(n^{-1}),\tag{I.14}$$

where $\bar{K}_{nd\beta}=\begin{pmatrix}0 & \bar{G}_{n\beta}'\bar{C}_n'\\ \bar{C}_n\bar{G}_{n\beta} & \bar{\Omega}_{nd}\end{pmatrix}$ and $\hat{B}$ is the r.h.s. of (I.2) but with $\tilde{C}_n$, $\hat{\theta}_{\text{TGEL3}}$ and $\check{\alpha}$ replaced by, respectively, $\check{C}_n$, $\hat{\theta}_{\text{TGEL4}}$ and $\dot{\alpha}_n(\hat{\beta}_{\text{TGEL4}})$. Then as in the above proof of (i), $\sqrt{n}(\hat{\theta}_{\text{TGEL4}}-\theta_0)=\binom{\psi_{n\beta}}{\psi_{n\mu}}+O_p(n^{-1/2})$, where $\psi_{n\beta}=-\bar{H}_{nd}\sqrt{n}\bar{C}_n g_n$ and $\psi_{n\mu}=-\bar{P}_{nd}\sqrt{n}\bar{C}_n g_n$. By first order Taylor expansions, $g_n(\dot{\alpha}_n(\hat{\beta}_{\text{TGEL4}}),\hat{\beta}_{\text{TGEL4}})=g_n+G_{n\alpha}(\alpha_0,\beta_0)(\dot{\alpha}_n(\hat{\beta}_{\text{TGEL4}})-\alpha_0)+G_{n\beta}(\alpha_0,\beta_0)(\hat{\beta}_{\text{TGEL4}}-\beta_0)+O_p(n^{-1})=g_n+n^{-1/2}\bar{G}_{n\alpha}\psi_{n\dot{\alpha}}+n^{-1/2}\bar{G}_{n\beta}\psi_{n\beta}+O_p(n^{-1})$, where $\psi_{n\dot{\alpha}}=\sqrt{n}(\dot{\alpha}_n(\beta_0)-\alpha_0)+\frac{\partial\alpha_n(\beta_0)}{\partial\beta'}\psi_{n\beta}$ by Lemma H.1(iii), $\frac{1}{n}\sum_{i=1}^{n}g_{ni}(\dot{\alpha}_n(\hat{\beta}_{\text{TGEL4}}),\hat{\beta}_{\text{TGEL4}})g_{ni}'(\dot{\alpha}_n(\hat{\beta}_{\text{TGEL4}}),\hat{\beta}_{\text{TGEL4}})=\bar{\Omega}_n+O_p(n^{-1/2})$, and

$$G_{n\alpha}(\dot{\alpha}_n(\hat{\beta}_{\text{TGEL4}}),\hat{\beta}_{\text{TGEL4}})=\bar{G}_{n\alpha}+n^{-1/2}\psi_{nG_\alpha}+O_p(n^{-1}),$$

where

$$\psi_{nG_\alpha}=\sqrt{n}(G_{n\alpha}-\bar{G}_{n\alpha})+\sum_{j=1}^{k_\alpha}\bar{G}_{n\alpha}^{(j)}\psi_{n\dot{\alpha}j}+\sum_{j=1}^{k_\beta}\bar{G}_{n\alpha}^{(k_\alpha+j)}\psi_{n\beta j}\tag{I.15}$$

with $\psi_{n\dot{\alpha}j}$ and $\psi_{n\beta j}$ being, respectively, the jth elements of $\psi_{n\dot{\alpha}}$ and $\psi_{n\beta}$. It follows that $\hat{A}=A+O_p(n^{-3/2})$, where $A=-\frac{1}{n}\alpha_{n\beta}'(\psi_{nG_\alpha}'\bar{C}_n'+\bar{G}_{n\alpha}'\psi_{n\check{C}}')\psi_{n\mu}-\frac{1}{n^2}\sum_{i=1}^{n}\psi_{n\mu}'\bar{C}_n\mathbf{E}(g_{ni}\alpha_{n\beta}'G_{ni\alpha}')\bar{C}_n'\psi_{n\mu}=O_p(n^{-1})$. Since $\sqrt{n}\bar{K}_{nd\beta}^{-1}\binom{\hat{A}}{0}=\begin{pmatrix}-\sqrt{n}\bar{\Sigma}_{nd}\hat{A}\\ \sqrt{n}\bar{H}_{nd}'\hat{A}\end{pmatrix}$, the contribution of $\sqrt{n}\bar{K}_{nd\beta}^{-1}\binom{\hat{A}}{0}$ in (I.14) for the higher order bias of $\hat{\beta}_{\text{TGEL4}}$ is

$$-\bar{\Sigma}_{nd}\,\mathbf{E}(A)=\frac{1}{n}\bar{\Sigma}_{nd}\,\mathbf{E}[\alpha_{n\beta}'(\psi_{nG_\alpha}'\bar{C}_n'+\bar{G}_{n\alpha}'\psi_{n\check{C}}')\psi_{n\mu}]+\frac{1}{n^2}\bar{\Sigma}_{nd}\sum_{i=1}^{n}\mathbf{E}[\psi_{n\mu}'\bar{C}_n\mathbf{E}(g_{ni}\alpha_{n\beta}'G_{ni\alpha}')\bar{C}_n'\psi_{n\mu}].$$

We may show that $\frac{1}{n^2}\bar{\Sigma}_{nd}\sum_{i=1}^{n}\mathbf{E}[\psi_{n\mu}'\bar{C}_n\mathbf{E}(g_{ni}\alpha_{n\beta}'G_{ni\alpha}')\bar{C}_n'\psi_{n\mu}]=\frac{1}{n^2}\bar{\Sigma}_{nd}\sum_{i=1}^{n}\mathbf{E}(\alpha_{n\beta}'G_{ni\alpha}'\bar{C}_n'\bar{P}_{nd}\bar{C}_n g_{ni})$. Then by the above proof of (i), the higher order bias of $\hat{\beta}_{\text{TGEL4}}$ is equal to the sum of $-\bar{\Sigma}_{nd}\,\mathbf{E}(A)$ and the higher order bias of $\hat{\beta}_{\text{TGEL3}}$ (with $\tilde{C}_n$ and $\check{\alpha}$ replaced by, respectively, $\check{C}_n$ and $\dot{\alpha}_n(\hat{\beta}_{\text{TGEL4}})$). As $\psi_{n\beta}$ is uncorrelated with $\psi_{n\mu}$, the result follows by (i).

(iv) For $\hat{\beta}_{\text{TGMM4}}$, by Lemma H.1, $\check{\beta}=\beta_0+o_p(1)$. By Lemma H.5, $\check{C}_n\check{\Omega}_n\check{C}_n=\bar{\Omega}_{nd}+o_p(1)$. Together with $\sup_{\beta\in\mathcal{B}}\|\check{C}_n g_n(\dot{\alpha}_n(\beta),\beta)-\bar{C}_n\bar{g}_n(\alpha_n(\beta),\beta)\|=o_p(1)$ and the identification condition in

Assumption 2(v), the consistency of $\hat{\beta}_{\text{TGMM4}}$ follows. Then, by the MVT, the first order condition for $\hat{\beta}_{\text{TGMM4}}$ becomes

$$\begin{aligned} 0 &= A_n'(\hat{\beta}_{\text{TGMM4}})(\breve{C}_n\breve{\Omega}_n\breve{C}_n')^{-1}\breve{C}_n g_n(\dot{\alpha}_n(\hat{\beta}_{\text{TGMM4}}),\hat{\beta}_{\text{TGMM4}}) \\ &= A_n'(\hat{\beta}_{\text{TGMM4}})(\breve{C}_n\breve{\Omega}_n\breve{C}_n')^{-1}\Big[\breve{C}_n g_n(\dot{\alpha}_n(\beta_0),\beta_0) + A_n(\ddot{\beta})(\hat{\beta}_{\text{TGMM4}} - \beta_0)\Big], \end{aligned}$$

where $A_n(\beta) = \breve{C}_n \frac{\partial g_n(\dot{\alpha}_n(\beta),\beta)}{\partial \alpha'}\frac{\partial \dot{\alpha}_n(\beta)}{\partial \beta'} + \breve{C}_n \frac{\partial g_n(\dot{\alpha}_n(\beta),\beta)}{\partial \beta'}$, $\breve{C}_n g_n(\dot{\alpha}_n(\beta_0),\beta_0) = \breve{C}_n g_n(\alpha_0,\beta_0) + \breve{C}_n \frac{\partial g_n(\ddot{\alpha},\beta_0)}{\partial \alpha'}(\dot{\alpha}_n(\beta_0) - \alpha_0)$, $\ddot{\beta}$ lies between β_0 and $\hat{\beta}_{\text{TGMM4}}$, and $\ddot{\alpha}$ lies between α_0 and $\dot{\alpha}_n(\beta_0)$. Thus,

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{\text{TGMM4}} - \beta_0) &= -\Big[A_n'(\hat{\beta}_{\text{TGMM4}})(\breve{C}_n\breve{\Omega}_n\breve{C}_n')^{-1}A_n(\ddot{\beta})\Big]^{-1} \\ &\quad \cdot A_n'(\hat{\beta}_{\text{TGMM4}})(\breve{C}_n\breve{\Omega}_n\breve{C}_n')^{-1}\Big[\sqrt{n}\breve{C}_n g_n(\alpha_0,\beta_0) + \breve{C}_n \frac{\partial g_n(\ddot{\alpha},\beta_0)}{\partial \alpha'}\sqrt{n}(\dot{\alpha}_n(\beta_0) - \alpha_0)\Big]. \end{aligned}$$

Since $\dot{\alpha}_n(\ddot{\beta}) = \alpha_n(\ddot{\beta}) + o_p(1) = \alpha_0 + o_p(1)$, $\breve{C}_n \frac{\partial g_n(\dot{\alpha}_n(\ddot{\beta}),\ddot{\beta})}{\partial \alpha'} = \bar{C}_n\bar{G}_{n\alpha} + o_p(1) = o_p(1)$ and $\frac{\partial g_n(\dot{\alpha}_n(\ddot{\beta}),\ddot{\beta})}{\partial \beta'} = \bar{G}_{n\beta} + o_p(1)$. In addition, $\frac{\partial \dot{\alpha}_n(\ddot{\beta})}{\partial \beta'} = \frac{\partial \alpha_n(\ddot{\beta})}{\partial \beta'} + o_p(1) = \frac{\partial \alpha_n(\beta_0)}{\partial \beta'} + o_p(1)$. Then,

$$\sqrt{n}(\hat{\beta}_{\text{TGMM4}} - \beta_0) = -(\bar{G}_{n\beta}'\bar{C}_n'\bar{\Omega}_{nd}^{-1}\bar{C}_n\bar{G}_{n\beta})^{-1}\bar{G}_{n\beta}'\bar{C}_n'\bar{\Omega}_{nd}^{-1}\bar{C}_n\sqrt{n}g_n + o_p(1) = O_p(1).$$

Let $\hat{\mu}_{\text{TGMM4}} = -(\breve{C}_n\breve{\Omega}_n\breve{C}_n')^{-1}\breve{C}_n g_n(\dot{\alpha}_n(\hat{\beta}_{\text{TGMM4}}),\hat{\beta}_{\text{TGMM4}}) = O_p(n^{-1/2})$. Then the first order condition for $\hat{\beta}_{\text{TGMM4}}$ is

$$\begin{aligned} 0 = &-\begin{pmatrix} G_{n\beta}'(\dot{\alpha}_n(\hat{\beta}_{\text{TGMM4}}),\hat{\beta}_{\text{TGMM4}})\breve{C}_n'\hat{\mu}_{\text{TGMM4}} \\ \breve{C}_n g_n(\dot{\alpha}_n(\hat{\beta}_{\text{TGMM4}}),\hat{\beta}_{\text{TGMM4}}) + \breve{C}_n\Omega_n(\dot{\alpha}_n(\hat{\beta}_{\text{TGMM4}}),\tilde{\beta})\breve{C}_n'\hat{\mu}_{\text{TGMM4}} \end{pmatrix} \\ &-\begin{pmatrix} \hat{\Xi}_n \\ \breve{C}_n[\Omega_n(\dot{\alpha}_n(\breve{\beta}),\breve{\beta}) - \Omega_n(\dot{\alpha}_n(\hat{\beta}_{\text{TGMM4}}),\tilde{\beta})]\breve{C}_n'\hat{\mu}_{\text{TGMM4}} \end{pmatrix}, \end{aligned} \tag{I.16}$$

where $\hat{\Xi}_n = \dot{\alpha}_{n\beta}'(\hat{\beta}_{\text{TGMM4}})G_{n\alpha}'(\dot{\alpha}_n(\hat{\beta}_{\text{TGMM4}}),\hat{\beta}_{\text{TGMM4}})\breve{C}_n'\hat{\mu}_{\text{TGMM4}} = O_p(n^{-1})$. Comparing the first order conditions (I.16) for $\hat{\theta}_{\text{TGMM4}}$ and (I.8) for $\hat{\theta}_{\text{TGMM3}}$, we have

$$\sqrt{n}(\hat{\theta}_{\text{TGMM4}} - \theta_0) = \hat{F} - \bar{K}_{nd\beta}^{-1}\begin{pmatrix} \sqrt{n}\hat{\Xi}_n \\ \sqrt{n}\breve{C}_n[\Omega_n(\dot{\alpha}_n(\breve{\beta}),\breve{\beta}) - \Omega_n(\dot{\alpha}_n(\hat{\beta}_{\text{TGMM4}}),\tilde{\beta})]\breve{C}_n'\hat{\mu}_{\text{TGMM4}} \end{pmatrix} + O_p(n^{-1}), \tag{I.17}$$

where $\hat{F}$ is the r.h.s. of (I.9) but with $\tilde{C}_n$, $\hat{\beta}_{\text{TGMM3}}$ and $\check{\alpha}$ replaced by, respectively, $\breve{C}_n$, $\hat{\beta}_{\text{TGMM4}}$ and $\dot{\alpha}_n(\hat{\beta}_{\text{TGMM4}})$. By arguments similar to those for (iii), $\hat{\Xi}_n = \Xi_n + O_p(n^{-3/2})$, where $\Xi_n = \frac{1}{n}\alpha_{n\beta}'(\psi_{nG_\alpha}'\bar{C}_n' + \bar{G}_{n\alpha}'\psi_{n\breve{C}}')\psi_{n\mu} = O_p(n^{-1})$. By the above proof of (ii), the higher order bias of $\hat{\beta}_{\text{TGMM4}}$ is equal to the sum of the higher order bias of $\hat{\beta}_{\text{TGMM3}}$ (with $\tilde{C}_n$ and $\check{\alpha}$ replaced by, respectively,

$\check{C}_n$ and $\dot{\alpha}_n(\hat{\beta}_{\text{TGMM2}})$) and

$$\frac{1}{n}\bar{\Sigma}_{nd}\,\mathbf{E}[\alpha'_{n\beta}(\psi'_{nG_\alpha}\bar{C}'_n+\bar{G}'_{n\alpha}\psi'_{n\check{C}})\psi_{n\mu}]-\frac{1}{n^2}\bar{H}_{nd}\bar{C}_n\sum_{j=1}^{k_\alpha}\sum_{i=1}^{n}\mathbf{E}(g_{ni}g_{ni}^{(j)'}+g_{ni}^{(j)}g'_{ni})\,\mathbf{E}[(\psi_{n\check{\alpha}j}-\psi_{n\dot{\alpha}j})\bar{C}'_n\psi_{n\mu}]$$
$$-\frac{1}{n^2}\bar{H}_{nd}\bar{C}_n\sum_{j=1}^{k_\beta}\sum_{i=1}^{n}\mathbf{E}(g_{ni}g_{ni}^{(k_\alpha+j)'}+g_{ni}^{(k_\alpha+j)}g'_{ni})\,\mathbf{E}[(\psi_{n\check{\beta}j}-\psi_{n\dot{\beta}j})\bar{C}'_n\psi_{n\mu}].$$

Hence, the result follows by the above proof of (ii).

I.2 Proof of Theorem E.1

(i) For $\hat{\beta}_{\text{TGMM}}$, we first prove its consistency. By Lemma H.1(i), $\tilde{\beta}=\beta_0+o_p(1)$. Then under Assumption 2(ii), $C_n(\check{\alpha},\tilde{\beta})=\bar{C}_n(\alpha_0,\tilde{\beta})+o_p(1)=\bar{C}_n(\alpha_0,\beta_0)+o_p(1)=\bar{C}_n+o_p(1)$. By Lemma H.5$(i)$, $\tilde{\Omega}_n=\bar{\Omega}_n+o_p(1)$. Since $\bar{C}_n$ has full row rank and $\lim_{n\to\infty}\bar{\Omega}_n$ is nonsingular by Assumption 1(v), $\lim_{n\to\infty}\bar{C}_n\bar{\Omega}_n\bar{C}'_n$ is nonsingular. Then by the uniform convergence $\sup_{\beta\in\mathcal{B}}\|C_n(\check{\alpha},\beta)g_n(\check{\alpha},\beta)-\bar{C}_n(\alpha_0,\beta)\bar{g}_n(\alpha_0,\beta)\|=o_p(1)$ shown in the proof of Theorem 1, the uniform equicontinuity of $\bar{C}_n(\alpha_0,\beta)\bar{g}_n(\alpha_0,\beta)$, the nonsingularity of $\lim_{n\to\infty}\bar{C}_n\bar{\Omega}_n\bar{C}'_n$, and the identification Assumption 2(v), the consistency of $\hat{\beta}_{\text{TGMM}}$ to β_0 follows.

By the MVT, the first order condition for $\hat{\beta}_{\text{TGMM}}$ becomes

$$\begin{aligned}0&=\frac{\partial[C_n(\check{\alpha},\hat{\beta}_{\text{TGMM}})g_n(\check{\alpha},\hat{\beta}_{\text{TGMM}})]'}{\partial\beta}(\tilde{C}_n\tilde{\Omega}_n\tilde{C}'_n)^{-1}C_n(\check{\alpha},\hat{\beta}_{\text{TGMM}})g_n(\check{\alpha},\hat{\beta}_{\text{TGMM}})\\&=\frac{\partial[C_n(\check{\alpha},\hat{\beta}_{\text{TGMM}})g_n(\check{\alpha},\hat{\beta}_{\text{TGMM}})]'}{\partial\beta}(\tilde{C}_n\tilde{\Omega}_n\tilde{C}'_n)^{-1}\Big[C_n(\check{\alpha},\beta_0)g_n(\check{\alpha},\beta_0)+\frac{\partial[C_n(\check{\alpha},\ddot{\beta})g_n(\check{\alpha},\ddot{\beta})]}{\partial\beta'}(\hat{\beta}_{\text{TGMM}}-\beta_0)\Big],\end{aligned}$$

where $\frac{\partial[C_n(\alpha,\beta)g_n(\alpha,\beta)]}{\partial\beta'}=C_n(\alpha,\beta)G_{n\beta}(\alpha,\beta)+[C_n^{(k_\alpha+1)}(\alpha,\beta)g_n(\alpha,\beta),\dots,C_n^{(k_\alpha+k_\beta)}(\alpha,\beta)g_n(\alpha,\beta)]$, and $\ddot{\beta}$ lies between β_0 and $\hat{\beta}_{\text{TGMM}}$. As $g_n(\check{\alpha},\ddot{\beta})=o_p(1)$, the term $[C_n^{(k_\alpha+1)}(\alpha,\beta)g_n(\alpha,\beta),\dots,C_n^{(k_\alpha+k_\beta)}(\alpha,\beta)g_n(\alpha,\beta)]$ in the derivative $\frac{\partial[C_n(\alpha,\beta)g_n(\alpha,\beta)]}{\partial\beta'}$ does not affect the asymptotic distribution of $\hat{\theta}_{\text{TGMM}}$. Thus, as in the proof of Theorem 1, $\sqrt{n}(\hat{\theta}_{\text{TGMM}}-\theta_0)=\sqrt{n}(\hat{\theta}_{\text{TGEL}}-\theta_0)+o_p(1)=[\psi'_{n\beta},\psi'_{n\mu}]'+o_p(1)$.

Let $\check{C}_n=C_n(\check{\alpha},\hat{\beta}_{\text{TGMM}})$ and $\hat{\mu}_{\text{TGMM}}=-(\tilde{C}_n\tilde{\Omega}_n\tilde{C}'_n)^{-1}\check{C}_ng_n(\check{\alpha},\hat{\beta}_{\text{TGMM}})=O_p(n^{-1/2})$. Then the first order condition for $\hat{\beta}_{\text{TGMM}}$ is

$$0=-\begin{pmatrix}G'_{n\beta}(\check{\alpha},\hat{\beta}_{\text{TGMM}})\check{C}'_n\hat{\mu}_{\text{TGMM}}\\ \check{C}_ng_n(\check{\alpha},\hat{\beta}_{\text{TGMM}})+\check{C}_n\tilde{\Omega}_n\check{C}'_n\hat{\mu}_{\text{TGMM}}\end{pmatrix}-\begin{pmatrix}\hat{\Xi}_n\\(\tilde{C}_n\tilde{\Omega}_n\tilde{C}'_n-\check{C}_n\tilde{\Omega}_n\check{C}'_n)\hat{\mu}_{\text{TGMM}}\end{pmatrix},\qquad\text{(I.18)}$$

where $\hat{\Xi}_n = [\hat{\mu}'_{\text{TGMM}} \check{C}_n^{(k_\alpha+1)}(\check{\alpha}, \hat{\beta}_{\text{TGMM}}) g_n(\check{\alpha}, \hat{\beta}_{\text{TGMM}}), \ldots, \hat{\mu}'_{\text{TGMM}} \check{C}_n^{(k_\alpha+k_\beta)}(\check{\alpha}, \hat{\beta}_{\text{TGMM}}) g_n(\check{\alpha}, \hat{\beta}_{\text{TGMM}})]' = O_p(n^{-1})$. By Lemma H.5($ii$)–($iii$), $\tilde{\Omega}_n = \bar{\Omega}_n + n^{-1/2}\psi_{n\tilde{\Omega}} + O_p(n^{-1})$, $\tilde{C}_n = \bar{C}_n + n^{-1/2}\psi_{n\tilde{C}} + O_p(n^{-1})$ and $\check{C}_n = \bar{C}_n + n^{-1/2}\psi_{n\check{C}} + O_p(n^{-1})$, where $\psi_{n\check{C}} = \psi_{nC} + \sum_{j=1}^{k_\alpha} \bar{C}_n^{(j)} \psi_{n\check{\alpha}j} + \sum_{j=1}^{k_\beta} \bar{C}_n^{(k_\alpha+j)} \psi_{n\beta j}$. Thus, $\tilde{C}_n\tilde{\Omega}_n\tilde{C}'_n - \check{C}_n\tilde{\Omega}_n\check{C}'_n = [\bar{C}_n + n^{-1/2}\psi_{n\tilde{C}} + O_p(n^{-1})][\bar{\Omega}_n + n^{-1/2}\psi_{n\tilde{\Omega}} + O_p(n^{-1})][\bar{C}'_n + n^{-1/2}\psi'_{n\tilde{C}} + O_p(n^{-1})] - [\bar{C}_n + n^{-1/2}\psi_{n\check{C}} + O_p(n^{-1})][\bar{\Omega}_n + n^{-1/2}\psi_{n\tilde{\Omega}} + O_p(n^{-1})][\bar{C}'_n + n^{-1/2}\psi'_{n\check{C}} + O_p(n^{-1})] = n^{-1/2}(\psi_{n\tilde{C}} - \psi_{n\check{C}})\bar{\Omega}_n\bar{C}'_n + n^{-1/2}\bar{C}_n\bar{\Omega}_n(\psi'_{n\tilde{C}} - \psi'_{n\check{C}}) + O_p(n^{-1})$. Comparing the first order conditions (I.18) for $\hat{\theta}_{\text{TGMM}}$ and (I.8) for $\hat{\theta}_{\text{TGMM}}$, we have

$$\sqrt{n}(\hat{\theta}_{\text{TGMM}} - \theta_0) = \hat{F} - \bar{K}_{nd\beta}^{-1}\begin{pmatrix} \sqrt{n}\hat{\Xi}_n \\ n^{-1/2}[(\psi_{n\tilde{C}} - \psi_{n\check{C}})\bar{\Omega}_n\bar{C}'_n + \bar{C}_n\bar{\Omega}_n(\psi'_{n\tilde{C}} - \psi'_{n\check{C}})]\psi_{n\mu} \end{pmatrix} + O_p(n^{-1}), \quad \text{(I.19)}$$

where $\hat{F}$ is the r.h.s. of (I.9) but with $\tilde{C}_n$ and $\hat{\beta}_{\text{TGMM3}}$ replaced by, respectively, $\check{C}_n$ and $\hat{\beta}_{\text{TGMM}}$. By arguments similar to those for (i), $\hat{\Xi}_n = \Xi_n + O_p(n^{-3/2})$, where $\Xi_n = \frac{1}{n}[\psi'_{n\mu}\bar{C}_n^{(k_\alpha+1)}(\sqrt{n}g_n + \bar{G}_{n\alpha}\psi_{n\check{\alpha}} + \bar{G}_{n\beta}\psi_{n\beta}), \ldots, \psi'_{n\mu}\bar{C}_n^{(k_\alpha+k_\beta)}(\sqrt{n}g_n + \bar{G}_{n\alpha}\psi_{n\check{\alpha}} + \bar{G}_{n\beta}\psi_{n\beta})]' = O_p(n^{-1})$. The first k_β elements of the second term on the r.h.s. of (I.19) form the vector $\bar{\Sigma}_{nd}\sqrt{n}\hat{\Xi}_n - n^{-1/2}\bar{H}_{nd}[(\psi_{n\tilde{C}} - \psi_{n\check{C}})\bar{\Omega}_n\bar{C}'_n + \bar{C}_n\bar{\Omega}_n(\psi'_{n\tilde{C}} - \psi'_{n\check{C}})]\psi_{n\mu}$. By the proof of Theorem 2(ii), the higher order bias of $\hat{\beta}_{\text{TGMM}}$ is equal to the sum of $\bar{\Sigma}_{nd}\,\mathbf{E}(\Xi_n) - \frac{1}{n}\bar{H}_{nd}\,\mathbf{E}[(\psi_{n\tilde{C}} - \psi_{n\check{C}})\bar{\Omega}_n\bar{C}'_n\psi_{n\mu} + \bar{C}_n\bar{\Omega}_n(\psi'_{n\tilde{C}} - \psi'_{n\check{C}})\psi_{n\mu}]$ and the higher order bias of $\hat{\beta}_{\text{TGMM3}}$ (with $\tilde{C}_n$ replaced by $\check{C}_n$), where

$$\mathbf{E}(\Xi_n) = \frac{1}{n}\bar{\Sigma}_{nd}\Big[\text{tr}[\bar{C}_n^{(k_\alpha+1)}(-\bar{\Omega}_n\bar{C}'_n\bar{P}_{nd} + \bar{G}_{n\alpha}\,\mathbf{E}(\psi_{n\check{\alpha}}\psi'_{n\mu}))], \ldots, \text{tr}[\bar{C}_n^{(k_\alpha+k_\beta)}(-\bar{\Omega}_n\bar{C}'_n\bar{P}_{nd} + \bar{G}_{n\alpha}\,\mathbf{E}(\psi_{n\check{\alpha}}\psi'_{n\mu}))]\Big]'.$$

Hence, the result follows by Theorem D.2(ii).

(ii) For $\hat{\beta}_{\text{TGMM2}}$, as $\breve{\beta} = \beta_0 + o_p(1)$ by Lemma H.1(i), $\dot{\alpha}_n(\breve{\beta}) = \alpha_0 + o_p(1)$. Then $\Omega_n(\dot{\alpha}_n(\breve{\beta}), \breve{\beta}) = \bar{\Omega}_n + o_p(1)$ by Lemma H.5(i). With the nonsingularity of $\lim_{n\to\infty}\bar{C}_n\bar{\Omega}_n\bar{C}'_n$, $\sup_{\beta\in\mathcal{B}}\|C_n(\dot{\alpha}_n(\beta), \beta)g_n(\dot{\alpha}_n(\beta), \beta) - \bar{C}_n(\alpha_n(\beta), \beta)\bar{g}_n(\alpha_n(\beta), \beta)\| = o_p(1)$ shown in the proof of Theorem 1 and the identification condition that $\lim_{n\to\infty}\bar{C}_n(\alpha_n(\beta), \beta)\bar{g}_n(\alpha_n(\beta), \beta)$ is uniquely zero at $\beta = \beta_0$, the consistency of $\hat{\beta}_{\text{TGMM2}}$ follows. The first order condition of $\hat{\beta}_{\text{TGMM2}}$, compared with that of $\hat{\beta}_{\text{TGMM4}}$, has an additional term due to the presence of $\dot{\alpha}_n(\beta)$ in $C_n(\dot{\alpha}_n(\beta), \beta)$, which does not affect the asymptotic distribution of $\hat{\beta}_{\text{TGMM2}}$, by arguments similar to those for $\hat{\beta}_{\text{TGEL2}}$. Thus, $\sqrt{n}(\hat{\theta}_{\text{TGMM2}} - \theta_0) = \sqrt{n}(\hat{\theta}_{\text{TGMM2}} - \theta_0) + o_p(1) = [\psi'_{n\beta}, \psi'_{n\mu}]' + o_p(1)$.

Denote $\hat{\beta} = \hat{\beta}_{\text{TGMM2}}$, $C_n = C_n(\dot{\alpha}_n(\hat{\beta}), \hat{\beta})$ and $\hat{\mu} = -(\check{C}_n\breve{\Omega}_n\check{C}'_n)^{-1}C_n g_n(\dot{\alpha}_n(\hat{\beta}), \hat{\beta}) = O_p(n^{-1/2})$. Then the first order condition for $\hat{\beta}$ is

$$0 = -\begin{pmatrix} G'_{n\beta}(\dot{\alpha}_n(\hat{\beta}), \hat{\beta})C'_n\hat{\mu} + \dot{\alpha}'_{n\beta}(\hat{\beta})G'_{n\alpha}(\dot{\alpha}_n(\hat{\beta}), \hat{\beta})C'_n\hat{\mu} \\ C_n g_n(\dot{\alpha}_n(\hat{\beta}), \hat{\beta}) + C_n\breve{\Omega}_n C'_n\hat{\mu} \end{pmatrix} - \begin{pmatrix} \hat{\Xi}_n \\ (\breve{C}_n\breve{\Omega}_n\breve{C}'_n - C_n\breve{\Omega}_n C'_n)\hat{\mu} \end{pmatrix}, \tag{I.20}$$

where

$$\begin{aligned} \hat{\Xi}_n = \bigg[& g'_n(\dot{\alpha}_n(\hat{\beta}), \hat{\beta})\Big[\sum_{k=1}^{k_\alpha} C_n^{(k)}(\dot{\alpha}_n(\hat{\beta}), \hat{\beta})\check{\alpha}_{nk}^{(1)}(\hat{\beta}) + C_n^{(k_\alpha+1)}(\dot{\alpha}_n(\hat{\beta}), \hat{\beta})\Big]'\hat{\mu}, \\ & \ldots, g'_n(\dot{\alpha}_n(\hat{\beta}), \hat{\beta})\Big[\sum_{k=1}^{k_\alpha} C_n^{(k)}(\dot{\alpha}_n(\hat{\beta}), \hat{\beta})\check{\alpha}_{nk}^{(k_\beta)}(\hat{\beta}) + C_n^{(k_\alpha+k_\beta)}(\dot{\alpha}_n(\hat{\beta}), \hat{\beta})\Big]'\hat{\mu}\bigg]' = O_p(n^{-1}). \end{aligned}$$

The first term on the r.h.s. of (I.20) has a similar form as the first order condition for $\hat{\beta}_{\text{TGMM4}}$. By arguments similar to those for (ii) above, $\hat{\Xi}_n = \Xi_n + O_p(n^{-3/2})$, where

$$\begin{aligned} \Xi_n = \frac{1}{n}\bigg[& (\sqrt{n}g_n + \bar{G}_{n\alpha}\psi_{n\dot{\alpha}} + \bar{G}_{n\beta}\psi_{n\beta})'\Big(\sum_{k=1}^{k_\alpha} \bar{C}_n^{(k)}\alpha_{nk}^{(1)} + \bar{C}_n^{(k_\alpha+1)}\Big)'\psi_{n\mu}, \\ & \ldots, (\sqrt{n}g_n + \bar{G}_{n\alpha}\psi_{n\dot{\alpha}} + \bar{G}_{n\beta}\psi_{n\beta})'\Big(\sum_{k=1}^{k_\alpha} \bar{C}_n^{(k)}\alpha_{nk}^{(k_\beta)} + \bar{C}_n^{(k_\alpha+k_\beta)}\Big)'\psi_{n\mu}\bigg]'. \end{aligned} \tag{I.21}$$

By the proof of Theorem D.2(iv) above for $\hat{\beta}_{\text{TGMM4}}$, the higher order bias of $\hat{\beta}_{\text{TGMM2}}$ is the sum of $\hat{\beta}_{\text{TGMM4}}$'s higher order bias (with $\psi_{n\tilde{C}}$ replaced by $\psi_{n\dot{C}}$) and

$$\begin{aligned} & \bar{\Sigma}_{nd}\,\mathbf{E}(\Xi_n) - \frac{1}{n}\bar{H}_{nd}\,\mathbf{E}[(\psi_{n\tilde{C}} - \psi_{n\dot{C}})\bar{\Omega}_n\bar{C}'_n\psi_{n\mu} + \bar{C}_n\bar{\Omega}_n(\psi'_{n\tilde{C}} - \psi'_{n\dot{C}})\psi_{n\mu}] \\ &= \frac{1}{n}\bar{\Sigma}_{nd}\bigg[\mathrm{tr}\Big[\Big(\sum_{k=1}^{k_\alpha} \bar{C}_n^{(k)}\alpha_{nk}^{(1)} + \bar{C}_n^{(k_\alpha+1)}\Big)'(-\bar{P}_{nd}\bar{C}_n\bar{\Omega}_n + \mathbf{E}(\psi_{n\mu}\psi'_{n\dot{\alpha}})\bar{G}'_{n\alpha})\Big], \\ &\qquad \ldots, \mathrm{tr}\Big[\Big(\sum_{k=1}^{k_\alpha} \bar{C}_n^{(k)}\alpha_{nk}^{(k_\beta)} + \bar{C}_n^{(k_\alpha+k_\beta)}\Big)'(-\bar{P}_{nd}\bar{C}_n\bar{\Omega}_n + \mathbf{E}(\psi_{n\mu}\psi'_{n\dot{\alpha}})\bar{G}'_{n\alpha})\Big]\bigg] \\ &\quad - \frac{1}{n}\bar{H}_{nd}\,\mathbf{E}[(\psi_{n\tilde{C}} - \psi_{n\dot{C}})\bar{\Omega}_n\bar{C}'_n\psi_{n\mu} + \bar{C}_n\bar{\Omega}_n(\psi'_{n\tilde{C}} - \psi'_{n\dot{C}})\psi_{n\mu}]. \end{aligned}$$

Hence, the result follows.

I.3 Proof of Theorem G.1

To derive the asymptotic distribution of $\mathcal{R}_{\text{TGMM}}$, we first investigate $\hat{\beta}_{\text{rTGMM}}$. Under Assumption 4, the consistency $\hat{\beta}_{\text{rTGMM}} = \beta_0 + o_p(1)$ can be proved as in the proof of Theorem 1. In the following proof, $\hat{\beta}$ denotes $\hat{\beta}_{\text{RTGMM}}$. The Lagrangian for $\hat{\beta}$ is $d_n'(\check{\alpha},\beta)\Omega_{nd}^{-1}(\check{\alpha},\tilde{\beta})d_n(\check{\alpha},\beta) - r'(\beta)a$, where a is a $k_r \times 1$ vector of Lagrangian multipliers. The first order conditions are

$$2[G_{n\beta}'(\check{\alpha},\hat{\beta})\hat{C}_n' + [\hat{C}_n^{(k_\alpha+1)}g_n(\check{\alpha},\hat{\beta}),\ldots,\hat{C}_n^{(k_\alpha+k_\beta)}g_n(\check{\alpha},\hat{\beta})]']\Omega_{nd}^{-1}(\check{\alpha},\tilde{\beta})d_n(\check{\alpha},\hat{\beta}) - R'(\hat{\beta})a = 0$$

and $r(\hat{\beta}) = 0$, where $\hat{C}_n = C_n(\check{\alpha},\hat{\beta})$ and $\hat{C}_n^{(k_\alpha+j)} = \frac{\partial C_n(\check{\alpha},\hat{\beta})}{\partial \beta_j}$. By the MVT, $\sqrt{n}d_n(\check{\alpha},\hat{\beta}) = \sqrt{n}d_n(\check{\alpha},\beta_0) + \hat{C}_n G_{n\beta}(\check{\alpha},\ddot{\beta})\sqrt{n}(\hat{\beta}-\beta_0)$ and $\sqrt{n}\cdot r(\hat{\beta}) = \sqrt{n}\cdot r(\beta_0) + R(\ddot{\beta})\sqrt{n}(\hat{\beta}-\beta_0) = R(\ddot{\beta})\sqrt{n}(\hat{\beta}-\beta_0)$, where $\ddot{\beta}$ lies between $\hat{\beta}$ and β_0. Then we have

$$\begin{pmatrix} 2[G_{n\beta}'(\check{\alpha},\hat{\beta})\hat{C}_n' + [\hat{C}_n^{(k_\alpha+1)}g_n(\check{\alpha},\hat{\beta}),\ldots,\hat{C}_n^{(k_\alpha+k_\beta)}g_n(\check{\alpha},\hat{\beta})]']\Omega_{nd}^{-1}(\check{\alpha},\tilde{\beta})\sqrt{n}d_n(\check{\alpha},\beta_0) \\ 0 \end{pmatrix}$$
$$+\begin{pmatrix} 2[G_{n\beta}'(\check{\alpha},\hat{\beta})\hat{C}_n' + [\hat{C}_n^{(k_\alpha+1)}g_n(\check{\alpha},\hat{\beta}),\ldots,\hat{C}_n^{(k_\alpha+k_\beta)}g_n(\check{\alpha},\hat{\beta})]']\Omega_{nd}^{-1}(\check{\alpha},\tilde{\beta})\hat{C}_n G_{n\beta}(\check{\alpha},\ddot{\beta}) & -R'(\hat{\beta}) \\ -R(\ddot{\beta}) & 0 \end{pmatrix}\begin{pmatrix} \sqrt{n}(\hat{\beta}-\beta_0) \\ \sqrt{n}a \end{pmatrix} = 0. \tag{I.22}$$

We may show that $\Omega_{nd}(\check{\alpha},\tilde{\beta}) = \bar{\Omega}_{nd} + o_p(1)$, $\hat{C}_n G_{n\beta}(\check{\alpha},\ddot{\beta}) = \bar{C}_n\bar{G}_{n\beta} + o_p(1)$, $g_n(\check{\alpha},\hat{\beta}) = o_p(1)$ and $\sqrt{n}d_n(\check{\alpha},\beta_0) = \sqrt{n}d_n(\alpha_0,\beta_0) + o_p(1)$. Thus, by (I.22) and the block matrix inverse formula,

$$\sqrt{n}(\hat{\beta}-\beta_0) = -[\bar{\Sigma}_{nd} - \bar{\Sigma}_{nd}R'(R\bar{\Sigma}_{nd}R')^{-1}R\bar{\Sigma}_{nd}]\bar{D}_{n\beta}'\bar{\Omega}_{nd}^{-1}\sqrt{n}d_n(\gamma_0) + o_p(1).$$

Hence, by the MVT,

$$nd_n'(\check{\alpha},\hat{\beta}_{\text{RTGMM}})\Omega_{nd}^{-1}(\check{\alpha},\hat{\beta}_{\text{RTGMM}})d_n(\check{\alpha},\hat{\beta}_{\text{RTGMM}})$$
$$= \sqrt{n}d_n'(\gamma_0)[\bar{\Omega}_{nd}^{-1} - \bar{\Omega}_{nd}^{-1}\bar{D}_{n\beta}\bar{\Sigma}_{nd}\bar{D}_{n\beta}'\bar{\Omega}_{nd}^{-1} + \bar{\Omega}_{nd}^{-1}\bar{D}_{n\beta}\bar{\Sigma}_{nd}R'(R\bar{\Sigma}_{nd}R')^{-1}R\bar{\Sigma}_{nd}\bar{D}_{n\beta}'\bar{\Omega}_{nd}^{-1}]\sqrt{n}d_n(\gamma_0) + o_p(1).$$

Similarly, $nd_n'(\check{\alpha},\hat{\beta}_{\text{TGMM}})\Omega_{nd}^{-1}(\check{\alpha},\hat{\beta}_{\text{TGMM}})d_n(\check{\alpha},\hat{\beta}_{\text{TGMM}}) = \sqrt{n}d_n'(\gamma_0)[\bar{\Omega}_{nd}^{-1} - \bar{\Omega}_{nd}^{-1}\bar{D}_{n\beta}\bar{\Sigma}_{nd}\bar{D}_{n\beta}'\bar{\Omega}_{nd}^{-1}]\sqrt{n}d_n(\gamma_0) + o_p(1)$. Therefore, $\mathcal{R}_{\text{TGMM}} = \sqrt{n}d_n'(\alpha_0,\beta_0)\bar{\Omega}_{nd}^{-1}\bar{D}_{n\beta}\bar{\Sigma}_{nd}R'(R\bar{\Sigma}_{nd}R')^{-1}R\bar{\Sigma}_{nd}\bar{D}_{n\beta}'\bar{\Omega}_{nd}^{-1}\sqrt{n}d_n(\alpha_0,\beta_0) + o_p(1)$. We may prove that $\mathcal{W}_{\text{TGMM}} = \mathcal{R}_{\text{TGMM}} + o_p(1)$ similarly as $\mathcal{W}_{\text{TGEL}} = \mathcal{R}_{\text{TGEL}} + o_p(1)$ in the proof of Theorem 3.

For $\mathcal{G}_{\text{TGMM}}$, by the MVT,

$$D_{n\beta}'(\check{\alpha},\hat{\beta}_{\text{RTGMM}})\Omega_{nd}^{-1}(\check{\alpha},\hat{\beta}_{\text{RTGMM}})\hat{C}_n\sqrt{n}g_n(\check{\alpha},\hat{\beta}_{\text{RTGMM}})$$
$$= D_{n\beta}'(\check{\alpha},\hat{\beta}_{\text{RTGMM}})\Omega_{nd}^{-1}(\check{\alpha},\hat{\beta}_{\text{RTGMM}})[\hat{C}_n\sqrt{n}g_n(\gamma_0) + \hat{C}_n G_{n\alpha}(\ddot{\gamma})\sqrt{n}(\check{\alpha}-\alpha_0) + \hat{C}_n G_{n\beta}(\ddot{\gamma})\sqrt{n}(\hat{\beta}_{\text{RTGMM}}-\beta_0)]$$
$$= R'(R\bar{\Sigma}_{nd}R')^{-1}R\bar{\Sigma}_{nd}\bar{D}_{n\beta}'\bar{\Omega}_{nd}^{-1}\bar{C}_n\sqrt{n}g_n(\alpha_0,\beta_0) + o_p(1).$$

where $\ddot{\gamma}$ lies between γ_0 and $(\check{\alpha}', \hat{\beta}'_{\text{rTGMM}})'$, and $\hat{C}_n \bar{G}_{n\alpha}(\ddot{\gamma}) = \bar{C}_n \bar{G}_{n\alpha} + o_p(1) = o_p(1)$ is used. It follows that

$$\mathcal{G}_{\text{TGMM}} = \sqrt{n} d_n'(\alpha_0, \beta_0) \bar{\Omega}_{nd}^{-1} \bar{D}_{n\beta} \bar{\Sigma}_{nd} R'(R\bar{\Sigma}_{nd} R')^{-1} R\bar{\Sigma}_{nd} \bar{D}_{n\beta}' \bar{\Omega}_{nd}^{-1} \sqrt{n} d_n(\alpha_0, \beta_0) + o_p(1) = \mathcal{R}_{\text{TGMM}} + o_p(1).$$

For $\mathcal{OPG}_{\text{T}}$, as $R\sqrt{n}(\hat{\beta}_r - \beta_0) = o_p(1)$, applying the MVT yields $\sqrt{n}\Psi_n(\check{\alpha}, \check{\beta}_r) = \sqrt{n}\Psi_n(\alpha_0, \beta_0) + o_p(1) = R\bar{\Sigma}_{nd} \bar{D}_{n\beta}' \bar{\Omega}_{nd}^{-1} \sqrt{n} d_n(\gamma_0) + o_p(1)$. As $\frac{1}{n}\sum_{i=1}^n \Psi_{ni}(\check{\alpha}, \check{\beta}_r)\Psi_{ni}'(\check{\alpha}, \check{\beta}_r) = R\bar{\Sigma}_{nd} R' + o_p(1)$, $\mathcal{OPG}_{\text{T}} = \mathcal{R}_{\text{TGMM}} + o_p(1)$.

I.4 Proof of Theorem 2

(i) Let $C_n = C_n(\check{\alpha}, \hat{\beta}_{\text{TGEL}})$, $v_{ni}(\alpha, \theta) = \mu' C_n g_{ni}(\alpha, \beta)$, where $\theta = (\beta', \mu')'$, $h_{ni}(\alpha, \theta) = \frac{\partial v_{ni}(\alpha, \theta)}{\partial \theta} = \begin{pmatrix} G_{ni\beta}'(\alpha,\beta) C_n' \mu \\ C_n g_{ni}(\alpha,\beta) \end{pmatrix}$, $m_{ni}(\alpha, \theta) = \rho_1(v_{ni}(\alpha, \theta)) h_{ni}(\alpha, \theta)$, $\hat{m}_n(\alpha, \theta) = \frac{1}{n}\sum_{i=1}^n m_{ni}(\alpha, \theta)$, and

$$\hat{\Delta}_{ni} = [\hat{\mu}_{\text{TGEL}}' C_n^{(k_\alpha+1)}(\check{\alpha}, \hat{\beta}_{\text{TGEL}}) g_{ni}(\check{\alpha}, \hat{\beta}_{\text{TGEL}}), \ldots, \hat{\mu}_{\text{TGEL}}' C_n^{(k_\alpha+k_\beta)}(\check{\alpha}, \hat{\beta}_{\text{TGEL}}) g_{ni}(\check{\alpha}, \hat{\beta}_{\text{TGEL}})]'.$$

Then the first order condition for $\hat{\theta}_{\text{TGEL}}$ is

$$\hat{m}_n(\check{\alpha}, \hat{\theta}_{\text{TGEL}}) + \frac{1}{n}\sum_{i=1}^n \rho_1(v_{ni}(\check{\alpha}, \hat{\theta}_{\text{TGEL}})) \begin{pmatrix} \hat{\Delta}_{ni} \\ 0 \end{pmatrix} = 0. \tag{I.23}$$

By Theorem 1, $\sqrt{n}(\hat{\theta}_{\text{TGEL}} - \theta_0) = O_p(1)$. Thus, by a second order Taylor expansion of $\rho_1(\cdot)$ in (I.23), $\frac{1}{n}\sum_{i=1}^n \rho_1(v_{ni}(\check{\alpha}, \hat{\theta}_{\text{TGEL}}))\hat{\Delta}_{ni} = \hat{A} + O_p(n^{-3/2})$, where

$$\hat{A} = \frac{1}{n}\sum_{i=1}^n \rho_1(0)\hat{\Delta}_{ni} + \frac{1}{n}\sum_{i=1}^n \rho_2(0) v_{ni}(\check{\alpha}, \hat{\theta}_{\text{TGEL}})\hat{\Delta}_{ni}.$$

Comparing the first order conditions (I.23) for $\hat{\theta}_{\text{TGEL}}$ and (I.1) for $\hat{\theta}_{\text{TGEL}}$, we have

$$\sqrt{n}(\hat{\theta}_{\text{TGEL}} - \theta_0) = \hat{B} + \sqrt{n}\bar{K}_{nd\beta}^{-1} \begin{pmatrix} \hat{A} \\ 0 \end{pmatrix} + O_p(n^{-1}), \tag{I.24}$$

where $\hat{B}$ is the r.h.s. of (I.2) but with $\tilde{C}_n$ and $\hat{\theta}_{\text{TGEL}}$ replaced by, respectively, C_n and $\hat{\theta}_{\text{TGEL}}$. Then as in the proof of Theorem D.2, $\sqrt{n}(\hat{\theta}_{\text{TGEL}} - \theta_0) = \binom{\psi_{n\beta}}{\psi_{n\mu}} + O_p(n^{-1/2})$, where $\psi_{n\beta} = -\bar{H}_{nd}\sqrt{n}\bar{C}_n g_n$ and $\psi_{n\mu} = -\bar{P}_{nd}\sqrt{n}\bar{C}_n g_n$. By first order Taylor expansions, $C_n^{(k_\alpha+j)}(\check{\alpha}, \hat{\beta}_{\text{TGEL}}) = C_n^{(k_\alpha+j)}(\alpha_0, \beta_0) + O_p(n^{-1/2}) = \bar{C}_n^{(k_\alpha+j)}(\alpha_0, \beta_0) + O_p(n^{-1/2})$, $g_n(\check{\alpha}, \hat{\beta}_{\text{TGEL}}) = g_n + G_{n\alpha}(\alpha_0, \beta_0)(\check{\alpha} - \alpha_0) + G_{n\beta}(\alpha_0, \beta_0)(\hat{\beta}_{\text{TGEL}} - \beta_0) + O_p(n^{-1}) = g_n + n^{-1/2}\bar{G}_{n\alpha}\psi_{n\check{\alpha}} + n^{-1/2}\bar{G}_{n\beta}\psi_{n\beta} + O_p(n^{-1})$, and

$$\frac{1}{n}\sum_{i=1}^n g_{ni}(\check{\alpha}, \hat{\beta}_{\text{TGEL}}) g_{ni}'(\check{\alpha}, \hat{\beta}_{\text{TGEL}}) = \bar{\Omega}_n + O_p(n^{-1/2}).$$

It follows that $\hat{A} = A + O_p(n^{-3/2})$, where

$$A = -\frac{1}{n}[\psi'_{n\mu}\bar{C}_n^{(k_\alpha+1)}(\sqrt{n}g_n + \bar{G}_{n\alpha}\psi_{n\check{\alpha}} + \bar{G}_{n\beta}\psi_{n\beta} + \bar{\Omega}_n\bar{C}'_n\psi_{n\mu}),$$
$$\ldots, \psi'_{n\mu}\bar{C}_n^{(k_\alpha+k_\beta)}(\sqrt{n}g_n + \bar{G}_{n\alpha}\psi_{n\check{\alpha}} + \bar{G}_{n\beta}\psi_{n\beta} + \bar{\Omega}_n\bar{C}'_n\psi_{n\mu})]'$$
$$= O_p(n^{-1}).$$

Since $\sqrt{n}\bar{K}_{nd\beta}^{-1}\binom{\hat{A}}{0} = \begin{pmatrix} -\sqrt{n}\bar{\Sigma}_{nd}\hat{A} \\ \sqrt{n}\bar{H}'_{nd}\hat{A} \end{pmatrix}$, the contribution of $\sqrt{n}\bar{K}_{nd\beta}^{-1}\binom{\hat{A}}{0}$ in (I.24) for the higher order bias of $\hat{\beta}_{\text{TGEL}}$ is

$$-\bar{\Sigma}_{nd}\,\mathbf{E}(A) = \frac{1}{n}\bar{\Sigma}_{nd}\Big[\mathrm{tr}[\bar{C}_n^{(k_\alpha+1)}\bar{G}_{n\alpha}\,\mathbf{E}(\psi_{n\check{\alpha}}\psi'_{n\mu})],\ldots,\mathrm{tr}[\bar{C}_n^{(k_\alpha+k_\beta)}\bar{G}_{n\alpha}\,\mathbf{E}(\psi_{n\check{\alpha}}\psi'_{n\mu})]\Big]'.$$

By Lemma H.5(iii), $C_n = \bar{C}_n + n^{-1/2}\psi_{n\check{C}} + O_p(n^{-1/2})$, where $\psi_{n\check{C}} = \psi_{nC} + \sum_{j=1}^{k_\alpha}\bar{C}_n^{(j)}\psi_{n\check{\alpha}j} + \sum_{j=1}^{k_\beta}\bar{C}_n^{(k_\alpha+j)}\psi_{n\beta j}$. Then by the proof of Theorem D.2(i), the higher order bias of $\hat{\beta}_{\text{TGEL}}$ is equal to the sum of $-\bar{\Sigma}_{nd}\,\mathbf{E}(A)$ and the higher order bias of $\hat{\beta}_{\text{TGEL3}}$ (with $\tilde{C}_n$ replaced by C_n). Hence the result follows by Theorem 2(i).

(ii) For the higher order bias of $\hat{\beta}_{\text{TGEL2}}$, compared with that of $\hat{\beta}_{\text{TGEL4}}$, we need to take into account the additional bias term due to the derivative of $C_n(\dot{\alpha}_n(\beta),\beta)$ with respect β, in addition to the difference between $C_n(\dot{\alpha}_n(\hat{\beta}_{\text{TGEL2}}),\hat{\beta}_{\text{TGEL2}})$ and $\check{C}_n$. Let $\dot{\alpha}_{n\beta}(\beta) = \frac{\partial\dot{\alpha}_n(\beta)}{\partial\beta'}$,

$$v_{ni}(\theta) = \mu' C_n(\dot{\alpha}_n(\beta),\beta)g_{ni}(\dot{\alpha}_n(\beta),\beta),$$

where $\theta = (\beta',\mu')'$, $h_{ni}(\theta) = \begin{pmatrix} G'_{ni\beta}(\alpha,\beta)C'_n(\dot{\alpha}_n(\beta),\beta)\mu + \dot{\alpha}'_{n\beta}(\beta)G'_{ni\alpha}(\dot{\alpha}_n(\beta),\beta)C'_n(\dot{\alpha}_n(\beta),\beta)\mu \\ C_n(\dot{\alpha}_n(\beta),\beta)g_{ni}(\dot{\alpha}_n(\beta),\beta) \end{pmatrix}$, $m_{ni}(\theta) = \rho_1(v_{ni}(\theta))h_{ni}(\theta)$, $\hat{m}_n(\theta) = \frac{1}{n}\sum_{i=1}^n m_{ni}(\theta)$, and

$$\Delta_{ni}(\theta) = \Big[\sum_{k=1}^{k_\alpha}\mu' C_n^{(k)}(\dot{\alpha}_n(\beta),\beta)\check{\alpha}_{nk}^{(1)}(\beta)g_{ni}(\dot{\alpha}_n(\beta),\beta) + \mu' C_n^{(k_\alpha+1)}(\dot{\alpha}_n(\beta),\beta)g_{ni}(\dot{\alpha}_n(\beta),\beta),$$
$$\ldots, \sum_{k=1}^{k_\alpha}\mu' C_n^{(k)}(\dot{\alpha}_n(\beta),\beta)\check{\alpha}_{nk}^{(k_\beta)}(\beta)g_{ni}(\dot{\alpha}_n(\beta),\beta) + \mu' C_n^{(k_\alpha+k_\beta)}(\dot{\alpha}_n(\beta),\beta)g_{ni}(\dot{\alpha}_n(\beta),\beta)\Big]'.$$

Then the first order condition for $\hat{\theta}_{\text{TGEL2}}$ is

$$\hat{m}_n(\hat{\theta}_{\text{TGEL2}}) + \frac{1}{n}\sum_{i=1}^n \rho_1(v_{ni}(\hat{\theta}_{\text{TGEL2}}))\begin{pmatrix}\Delta_{ni}(\hat{\theta}_{\text{TGEL2}}) \\ 0\end{pmatrix} = 0. \tag{I.25}$$

As for $\hat{\beta}_{\text{TGEL4}}$, the contribution of $\Delta_{ni}(\theta)$ for the higher order bias of $\hat{\beta}_{\text{TGEL2}}$ is from the leading order term $-\bar{\Sigma}_{nd}A$ of $-\bar{\Sigma}_{nd}[\frac{1}{n}\sum_{i=1}^n\rho_1(0)\Delta_{ni}(\hat{\theta}_{\text{TGEL2}}) + \frac{1}{n}\sum_{i=1}^n\rho_2(0)v_{ni}(\hat{\theta}_{\text{TGEL2}})\Delta_{ni}(\hat{\theta}_{\text{TGEL2}})] = -\bar{\Sigma}_{nd}A +$

$O_p(n^{-1})$, which is

$$-\bar{\Sigma}_{nd}\,\mathbf{E}(A) = \frac{1}{n}\bar{\Sigma}_{nd}\,\mathbf{E}\Big[\psi'_{n\mu}\Big(\sum_{k=1}^{k_\alpha}\bar{C}_n^{(k)}\alpha_{nk}^{(1)} + \bar{C}_n^{(k_\alpha+1)}\Big)(\sqrt{n}g_n + \bar{G}_{n\alpha}\psi_{n\check{\alpha}} + \bar{G}_{n\beta}\psi_{n\beta} + \bar{\Omega}_n\bar{C}_n\psi_{n\mu}),$$

$$\ldots,\psi'_{n\mu}\Big(\sum_{k=1}^{k_\alpha}\bar{C}_n^{(k)}\alpha_{nk}^{(k_\beta)} + \bar{C}_n^{(k_\alpha+k_\beta)}\Big)(\sqrt{n}g_n + \bar{G}_{n\alpha}\psi_{n\check{\alpha}} + \bar{G}_{n\beta}\psi_{n\beta} + \bar{\Omega}_n\bar{C}_n\psi_{n\mu})\Big]'$$

$$= \frac{1}{n}\bar{\Sigma}_{nd}\,\mathbf{E}\Big[\psi'_{n\mu}\Big(\sum_{k=1}^{k_\alpha}\bar{C}_n^{(k)}\alpha_{nk}^{(1)} + \bar{C}_n^{(k_\alpha+1)}\Big)\bar{G}_{n\alpha}\psi_{n\check{\alpha}},\ldots,\psi'_{n\mu}\Big(\sum_{k=1}^{k_\alpha}\bar{C}_n^{(k)}\alpha_{nk}^{(k_\beta)} + \bar{C}_n^{(k_\alpha+k_\beta)}\Big)\bar{G}_{n\alpha}\psi_{n\check{\alpha}}\Big]'.$$

By Lemma H.5(iii), $C_n(\dot{\alpha}_n(\hat{\beta}_{\text{TGEL2}}),\hat{\beta}_{\text{TGEL2}}) = \bar{C}_n + n^{-1/2}\psi_{n\dot{C}} + O_p(n^{-1/2})$, where $\psi_{n\dot{C}} = \psi_{nC} + \sum_{j=1}^{k_\alpha}\bar{C}_n^{(j)}\psi_{n\dot{\alpha}j} + \sum_{j=1}^{k_\beta}\bar{C}_n^{(k_\alpha+j)}\psi_{n\beta j}$. Then by the proof of (i) above, the higher order bias of $\hat{\beta}_{\text{TGEL2}}$ is equal to the sum of $-\bar{\Sigma}_{nd}\,\mathbf{E}(A)$ and the higher order bias of $\hat{\beta}_{\text{TGEL4}}$ (with $\psi_{n\check{C}}$ replaced by $\psi_{n\dot{C}}$). Hence the result follows.

I.5 Some additional proof for Theorem 3

We derive the influence function for the restricted TGEL estimator in this section, which is used in the proof of Theorem 3 of the main paper. With a continuous function $r(\beta)$, the set $\{\beta \in \mathcal{B} : r(\beta) = 0\}$ is compact. Since $r(\beta_0) = 0$, the consistency that $\hat{\beta}_{\text{rTGEL}} = \beta_0 + o_p(1)$ and $\hat{\mu}_{\text{rTGEL}} = o_p(1)$ follow as in Theorem 1. In the following, we omit the subscripts of $\hat{\beta}_{\text{rTGEL}}$ and $\hat{\mu}_{\text{rTGEL}}$ for simplicity. For given β, let $\mu(\beta)$ be the GEL estimate of μ. Then $\mu(\beta)$ satisfies $\sum_{i=1}^n \rho_1(\mu'(\beta)d_{ni}(\check{\alpha},\beta))d_{ni}(\check{\alpha},\beta) = 0$. With $\mu(\beta)$, the Lagrangian for $\hat{\beta}$ is $\frac{1}{n}\sum_{i=1}^n \rho(\mu'(\beta)d_{ni}(\check{\alpha},\beta)) - r'(\beta)a$, where a is a $k_r \times 1$ vector of Lagrangian multipliers. The first order condition for β is

$$\frac{1}{n}\sum_{i=1}^n \rho_1(\hat{\mu}'d_{ni}(\check{\alpha},\hat{\beta}))[G'_{ni\beta}(\check{\alpha},\hat{\beta})\hat{C}'_n\hat{\mu} + [\hat{C}_n^{(k_\alpha+1)}g_{ni}(\check{\alpha},\hat{\beta}),\ldots,\hat{C}_n^{(k_\alpha+k_\beta)}g_{ni}(\check{\alpha},\hat{\beta})]'\hat{\mu}] - R'(\hat{\beta})a = 0, \quad \text{(I.26)}$$

where $R(\beta) = \frac{\partial r(\beta)}{\partial \beta'}$, $\hat{C}_n = C_n(\check{\alpha},\hat{\beta})$ and $\hat{C}_n^{(k_\alpha+j)} = \frac{\partial C_n(\check{\alpha},\hat{\beta})}{\partial \beta_j}$. Treating $\sum_{i=1}^n \rho_1(\mu'd_{ni}(\check{\alpha},\beta))d_{ni}(\check{\alpha},\beta) = \sum_{i=1}^n \rho_1(\mu'\hat{C}_n g_{ni}(\check{\alpha},\beta))\hat{C}_n g_{ni}(\check{\alpha},\beta)$ as a function of β for only β in the second $g_{ni}(\check{\alpha},\beta)$, an MVT expansion of the first order condition $\sum_{i=1}^n \rho_1(\hat{\mu}'d_{ni}(\check{\alpha},\hat{\beta}))d_{ni}(\check{\alpha},\hat{\beta}) = 0$ at $(\beta',\mu')' = (\beta_0',0)'$ yields

$$-d_n(\check{\alpha},\beta_0) + \frac{1}{n}\sum_{i=1}^n \rho_2(\bar{\mu}'d_{ni}(\check{\alpha},\hat{\beta}))d_{ni}(\check{\alpha},\hat{\beta})d'_{ni}(\check{\alpha},\hat{\beta})\hat{\mu} + \frac{1}{n}\sum_{i=1}^n \rho_1(\bar{\mu}'d_{ni}(\check{\alpha},\hat{\beta}))\hat{C}_n G_{ni\beta}(\check{\alpha},\bar{\beta})(\hat{\beta}-\beta_0) = 0, \quad \text{(I.27)}$$

where $\bar{\beta}$ lies between β_0 and $\hat{\beta}$, and $\bar{\mu}$ lies between 0 and $\hat{\mu}$. By the MVT,

$$0 = r(\hat{\beta}) = r(\beta_0) + R(\bar{\beta})(\hat{\beta}-\beta_0) = R(\bar{\beta})(\hat{\beta}-\beta_0). \quad \text{(I.28)}$$

As in the proof of Theorem 1, $\frac{1}{n}\sum_{i=1}^{n}\rho_2(\bar{\mu}'d_{ni}(\check{\alpha},\hat{\beta}))d_{ni}(\check{\alpha},\hat{\beta})d'_{ni}(\check{\alpha},\hat{\beta}) = -\bar{\Omega}_{nd} + o_p(1)$ and

$$\frac{1}{n}\sum_{i=1}^{n}\rho_1(\bar{\mu}'d_{ni}(\check{\alpha},\hat{\beta}))\hat{C}_n G_{ni\beta}(\check{\alpha},\bar{\beta}) = -\bar{D}_{n\beta} + o_p(1).$$

Then by (I.27),

$$\sqrt{n}\hat{\mu} = -\bar{\Omega}_{nd}^{-1}\sqrt{n}d_n(\check{\alpha},\beta_0) - \bar{\Omega}_{nd}^{-1}\bar{D}_{n\beta}\sqrt{n}(\hat{\beta}-\beta_0) + o_p(1). \tag{I.29}$$

Substituting (I.29) into (I.26) yields

$$\sqrt{n}(\hat{\beta}-\beta_0) = \bar{\Sigma}_{nd}R'\sqrt{n}a - \bar{\Sigma}_{nd}\bar{D}'_{n\beta}\bar{\Omega}_{nd}^{-1}\sqrt{n}d_n(\check{\alpha},\beta_0) + o_p(1). \tag{I.30}$$

Substituting (I.30) into (I.28) yields $\sqrt{n}a = (R\bar{\Sigma}_{nd}R')^{-1}R\bar{\Sigma}_{nd}\bar{D}'_{n\beta}\bar{\Omega}_{nd}^{-1}\sqrt{n}d_n(\check{\alpha},\beta_0) + o_p(1)$, where $R = R(\beta_0)$. This equation and (I.30) imply that

$$\sqrt{n}(\hat{\beta}-\beta_0) = -[\bar{\Sigma}_{nd} - \bar{\Sigma}_{nd}R'(R\bar{\Sigma}_{nd}R')^{-1}R\bar{\Sigma}_{nd}]\bar{D}'_{n\beta}\bar{\Omega}_{nd}^{-1}\sqrt{n}d_n(\check{\alpha},\beta_0) + o_p(1). \tag{I.31}$$

Then (I.29) implies that

$$\sqrt{n}\hat{\mu} = [-\bar{\Omega}_{nd}^{-1} + \bar{\Omega}_{nd}^{-1}\bar{D}_{n\beta}\bar{\Sigma}_{nd}\bar{D}'_{n\beta}\bar{\Omega}_{nd}^{-1} - \bar{\Omega}_{nd}^{-1}\bar{D}_{n\beta}\bar{\Sigma}_{nd}R'(R\bar{\Sigma}_{nd}R')^{-1}R\bar{\Sigma}_{nd}\bar{D}'_{n\beta}\bar{\Omega}_{nd}^{-1}]\sqrt{n}d_n(\check{\alpha},\beta_0) + o_p(1). \tag{I.32}$$

J Assumptions for the estimation of the probit model using simple moments

In this section, we verify the assumptions for the estimation of the probit model (2.1) using simple moments.

Assumption 1(i): With $g_b(\gamma) = x[y_1 - \Phi(x_1'\beta_1 + x_2'\alpha_2\beta_2)]$ and $g_a(\gamma) = x(y_2 - x'\alpha)$, where $\alpha = [\alpha_1', \alpha_2']'$ and $\gamma = [\alpha', \beta_1', \beta_2]'$, we have $\mathbf{E}[g_b(\gamma)] = \mathbf{E}\{x[\Phi(x_1'\beta_{10} + x_2'\alpha_{20}\beta_{20}) - \Phi(x_1'\beta_1 + x_2'\alpha_2\beta_2)]\}$ and $\mathbf{E}[g_a(\gamma)] = \mathbf{E}(xx')(\alpha_0 - \alpha)$, where $\alpha = [\alpha_1', \alpha_2']'$. Since $\mathbf{E}(xx') = I_{k_x}$, where k_x is the number of variables in x, by $\mathbf{E}[g_a(\gamma)] = 0$, $\mathbf{E}[g(\gamma)] = 0$ implies that $\alpha = \alpha_0$. With $\alpha = \alpha_0$, $\mathbf{E}[g_b(\alpha_0,\beta)] = \mathbf{E}\{x[\Phi(x_1'\beta_{10} + x_2'\alpha_{20}\beta_{20}) - \Phi(x_1'\beta_1 + x_2'\alpha_{20}\beta_2)]\} = 0$ implies that $\mathbf{E}[z\Phi(z'\beta_0) - z\Phi(z'\beta)] = 0$, where $z = [x_1', x_2'\alpha_{20}]'$. By the mean value theorem, $\mathbf{E}[z\Phi(z'\beta_0) - z\Phi(z'\beta)] = \mathbf{E}[\phi(z'\bar{\beta})zz'](\beta_0 - \beta)$, where $\bar{\beta}$ lies between β_0 and β. Then $0 = (\beta_0-\beta)'\mathbf{E}[z\Phi(z'\beta_0) - z\Phi(z'\beta)] = \mathbf{E}[\phi(z'\bar{\beta})\cdot(\beta_0-\beta)'z\cdot z'(\beta_0-\beta)]$.

As $\phi(z'\bar{\beta}) > 0$ for any $\bar{\beta}$, we must have $z'(\beta_0 - \beta) = 0$ with probability one. Since the elements of $[x_1', x_2']'$ are independently and normally distributed and the elements of α_{20} are equal and nonzero, we must have $\beta = \beta_0$.

Assumption 1(ii) on the compactness of the parameter space is a usual regularity condition for extremum estimation. Assumption 1(iii) is obviously satisfied. Assumption 1(iv) is satisfied by Lemma 2.4 in Newey and McFadden (1994), since $g_i(\gamma)$'s are i.i.d., Γ is compact, $\mathbf{E}\sup_{\gamma\in\Gamma}\|g_b(\gamma)\|^\eta \le 2^\eta\,\mathbf{E}\|x\|^\eta < \infty$ and $\mathbf{E}\sup_{\gamma\in\Gamma}\|g_a(\gamma)\|^\eta = \mathbf{E}\sup_{\gamma\in\Gamma}\|x[x'(\alpha_0-\alpha)+\epsilon]\|^\eta \le \mathbf{E}\sup_{\gamma\in\Gamma}\|x\|^\eta\cdot(\|x\|\cdot\|\alpha_0-\alpha\|+\|\epsilon\|)^\eta \le 2^{\eta-1}\,\mathbf{E}(\|x\|^{2\eta}\sup_{\gamma\in\Gamma}\|\alpha_0-\alpha\|^\eta + \|x\|^\eta\|\epsilon\|^\eta) < \infty$ for any $\eta \ge 1$, where the last "$\le$" uses the C_r-inequality. Assumption 1(v) is satisfied by the above proof for Assumption 1(iv).

Assumption 1(vi): The conditions except the last one are satisfied as Assumption 1(iv). Now we compute $\bar{\Omega}_n$. Let $\xi = [y_1 - \Phi(z'\beta_0), \epsilon]'$. Then $\bar{\Omega}_n = \mathbf{E}[(\xi\xi')\otimes(xx')] = \mathbf{E}[\mathbf{E}(\xi\xi'|x)\otimes(xx')]$. We first consider $\mathbf{E}[\xi\xi'|x] = \begin{pmatrix}\Phi(z'\beta_0)[1-\Phi(z'\beta_0)] & w\\ w & \sigma_\epsilon^2\end{pmatrix}$, where $w = \mathbf{E}(y_1\epsilon|x)$. Let $v = u - \mathrm{cov}(u,\epsilon)\,\mathrm{var}^{-1}(\epsilon)\epsilon = u - \rho_0\sigma_u\epsilon/\sigma_\epsilon$, which is independent of ϵ. Then $w = \mathbf{E}[\epsilon\cdot I((x'\alpha_0+\epsilon)\tau_0 + u > 0)|x] = \mathbf{E}[\epsilon\cdot I(x'\alpha_0\tau_0 + v + (\tau_0+\rho_0\sigma_u/\sigma_\epsilon)\epsilon > 0)|x] = \sigma_\epsilon\,\mathbf{E}\{[2\cdot I(\tau_0\sigma_\epsilon+\rho_0\sigma_u > 0) - 1]\phi(\frac{x'\alpha_0\tau_0+v}{\tau_0\sigma_\epsilon+\rho_0\sigma_u})|x\}$. Note that $\mathbf{E}[\xi\xi'|x]$ is invertible if $\Phi(z'\beta_0)[1-\Phi(z'\beta_0)] - w^2/\sigma_\epsilon^2 \neq 0$. For given z, Since $\Phi(z'\beta_0)[1-\Phi(z'\beta_0)] - w^2/\sigma_\epsilon^2$ is a nonzero and nonlinear function of z, it is nonzero with a positive probability. Then $\bar{\Omega}_n$ is nonsingular.

Assumption 1(vii) normalizes $\rho_1(0)$ and $\rho_2(0)$ to be -1, which loses no generality as long as $\rho_1(0) \neq 0$ and $\rho_2(0) \neq 0$ (Newey and Smith, 2004). Assumption 1($viii$) is a usual assumption for deriving the asymptotic distributions of extremum estimators. Assumption 1(ix) can be verified as for Assumption 1(iv).

Assumption 1(x): Note that $\bar{G}_n = -\mathbf{E}\begin{pmatrix} x[0,x_2'\beta_{20}]\phi(z'\beta_0) & xz'\phi(z'\beta_0)\\ xx' & 0\end{pmatrix}$. Consider

$$\bar{G}_n\begin{pmatrix}\delta_1\\ \delta_2\end{pmatrix} = -\mathbf{E}\begin{pmatrix} x[0,x_2'\beta_{20}]\delta_1\phi(z'\beta_0) + xz'\delta_2\phi(z'\beta_0)\\ xx'\delta_1\end{pmatrix},$$

where δ_1 is a $k_x\times 1$ vector and δ_2 is a $(k_{x_1}+1)\times 1$ vector. Since $\mathbf{E}(xx') = I_{k_x}$, the second row block of $\bar{G}_n\binom{\delta_1}{\delta_2} = 0$ implies that $\delta_1 = 0$. Then the first row block implies that $\mathbf{E}[xz'\phi(z'\beta_0)]\delta_2 = 0$, where $z = \begin{pmatrix} I_{k_{x_1}} & \\ & \alpha_{20}'\end{pmatrix}x$. Since $\phi(z'\beta_0) > 0$ and xx' is nonnegative definite with x being normally distributed, $\mathbf{E}[xx'\phi(z'\beta_0)] \neq 0$. Then $\delta_2 = 0$ as $\alpha_{20} \neq 0$. Hence, $\bar{G}_n$ has full column rank.

Assumption 1(xi) is satisfied by the Lindeberg-Lévy central limit theorem.

Assumption 2(i): As $m_a = k_\alpha$ in our Monte Carlo experiments, $\bar{C}_{1n} = \mathbf{E}(\frac{\partial g_b(\gamma_0)}{\partial \alpha'})(\mathbf{E}\frac{\partial g_a(\gamma_0)}{\partial \alpha'})^{-1} = \mathbf{E}[x[0, x_2'\alpha_{20}]\phi(z'\beta_0)][\mathbf{E}(xx')]^{-1}$.

Assumption 2(ii): We take $C_{1n}(\gamma) = [\sum_{i=1}^n x_i[0, x_{2i}'\alpha_2]\phi(x_{1i}'\beta_1 + x_{2i}'\alpha_2\beta_2)](\sum_{i=1}^n x_i x_i')^{-1}$ and $\bar{C}_{1n}(\gamma) = \mathbf{E}[C_{1n}(\gamma)]$. Then as for Assumption 1(iv), $\sup_{\gamma\in\Gamma}\|C_n(\gamma) - \bar{C}_n(\gamma)\| = o_p(1)$ and $\bar{C}_n(\gamma)$ is continuous on Γ uniformly in n.

Assumption 2(iii): $\check{\alpha} = (\sum_{i=1}^n x_i x_i')^{-1}\sum_{i=1}^n x_i y_{2i}$ has a closed form and needs not to be bounded, but $\check{\alpha}$ falls within a bounded parameter space w.p.a.1. Assumption 2(iii) can be relaxed to allow for this case and the current Assumption 2(iii) is maintained for convenience.

Assumption 2(iv) is obviously satisfied since $\check{\alpha}_2$ is an OLS estimator.

Assumption 2(v): In this model, $\mathbf{E}[g_a(\alpha_0, \beta)] = 0$, so $\bar{C}_n\mathbf{E}[g(\alpha_0, \beta)] = \mathbf{E}[g_b(\alpha_0, \beta)]$, which is shown to be uniquely zero at $\beta = \beta_0$ when verifying Assumption 1(i).

Assumption 2(vi) is obviously satisfied.

Assumption 2(vii): Since $\mathbf{E}(\frac{\partial g_a(\alpha_0,\beta_0)}{\partial \beta'}) = 0$, $\bar{C}_n\bar{G}_{n\beta} = \mathbf{E}(\frac{\partial g_b(\alpha_0,\beta_0)}{\partial \beta'}) = -\mathbf{E}[xz'\phi(z'\beta_0)]$, which is shown to be nonzero when verifying Assumption 1(x). Thus, $\bar{C}_n\bar{G}_{n\beta}$ has full column rank.

K Computational details for Monte Carlo experiments

For GEL and TGEL estimators, we use a double optimization method. For given model parameters, the auxiliary parameter vector is computed using the "minFunc" function written by Schmidt (2005), which is a function for unconstrained optimization using line-search methods and requires fewer function evaluations to converge than the "fminunc" function in Matlab on many problems. We use the default algorithm "lbfgs", which calls a quasi-Newton strategy with limited memory BFGS updating. The starting value is a zero vector and the first order derivative of the objective function is provided. Note that the function can fail to converge simply because the zero vector is not in the convex hull of $g_{ni}(\theta)$'s. In such a case, we follow the convention of setting the function value to infinity for subsequent optimization. Although this problem can be solved by various methods, e.g., see Tsao and Wu (2014) and references therein, we do not consider the problem in this paper. With the value of the auxiliary parameter, model parameters are then computed either by grid search or by an optimization algorithm. Grid search is used for two-step estimators of the coefficient β_2 on the endogenous regressor in the probit model using simple moments, where the endogenous regressor is the only regressor for the equation on y_1^*, and for two-step estimators of the spatial dependent parameter β in the SAR model. The "minFunc" function is used for other estimators. In grid search, for the probit model, β is searched over the interval $[-25, 25]$ with a grid size 0.01; for the SAR model, β is searched over the interval $[-0.99, 0.99]$ with a grid size 0.01. If an optimization algorithm is used, any estimate of β greater than 25 is set to 25, and any estimate smaller than -25 is set to -25. We input the first order derivatives, which are derived by the implicit function theorem for GEL and TGEL. The starting value is the GMM estimate $\tilde{\beta} = \arg\min_{\beta} g'_{nb}(\check{\alpha}, \beta) g_{nb}(\check{\alpha}, \beta)$, where $\check{\alpha}$ is the OLS estimate by regressing y_2 on x_2, the starting value for $\tilde{\beta}$ is a zero vector and "minFunc" is used for $\tilde{\beta}$. For the estimation of the probit model using efficient moments, since even two-step estimators have 2 unknown parameters, we do not use grid search for any estimator. As the variance parameter σ_ϵ^2 is non-negative and the correlation coefficient ρ is smaller than 1 in absolute value, we use the "fmincon" function and restrict σ_ϵ^2 to be non-smaller than 10^{-7} and restrict ρ to be in the interval $[-0.999, 0.999]$. The first order derivatives are provided and the trust-region-reflective algorithm is used. For the SAR model, "fmincon" is used to search for model parameters when grid search is not used, and the spatial dependence parameter β is restricted to be in the interval $[-0.99, 0.99]$ so that it is consistent with grid search.[4]

L More Monte Carlo results

L.1 CU estimator for the probit model

Table S.1 compares the performance of CU, two-step CU (TCU) and other estimates of β_2 in the probit model (2.1) with $R^2 = 0.7$. When there are 5 variables in x_2, all TPs are zero. CU and TCU estimates have similar MBs, MADs, IDRs, biases, SDs and RMSEs, like other GEL and TGEL estimates. In terms of MBs and biases, CU performs similarly as ET and EL, and TCU performs similarly as TET and TEL. However, CU and TCU have the largest MADs, IDRs, SDs and RMSEs. When there are 20 variables in x_2, all TPs are zero except CU. CU is observed

[4]See footnote 19 on page 30 of the main text for an explanation of this parameter space.

to have the largest MB, MAD IDR, bias, SD and RMSE among CU, ET, EL and GMM. TCU is also observed to have the largest MB, MAD, IDR, bias, SD and RMSE among TGEL and TGMM estimates. Thus, CU and TCU have worse performance than other GEL and TEL estimates in our Monte Carlo experiments.

Table S.1: Performance of CU, TCU and other estimates of β_2 in the probit model with $R^2 = 0.7$

	$k_2 = 5$							$k_2 = 20$						
	MB	MAD	IDR	Bias	SD	RMSE	TP	MB	MAD	IDR	Bias	SD	RMSE	TP
CU	-0.002	0.064	0.245	-0.005	0.097	0.098	0	0.010	0.116	0.470	0.021	0.596	0.596	0.001
ET	-0.002	0.061	0.233	-0.005	0.091	0.092	0	0.004	0.072	0.284	0.001	0.115	0.115	0
EL	-0.002	0.057	0.224	-0.005	0.088	0.088	0	0.002	0.058	0.231	0.002	0.092	0.092	0
GMM	0.000	0.061	0.240	-0.003	0.094	0.094	0	0.034	0.080	0.300	0.032	0.119	0.123	0
TCU	-0.003	0.064	0.245	-0.005	0.097	0.097	0	0.007	0.115	0.462	0.014	0.249	0.249	0
TET	-0.002	0.061	0.232	-0.005	0.091	0.092	0	0.003	0.071	0.282	0.001	0.115	0.115	0
TEL	-0.002	0.057	0.224	-0.005	0.088	0.088	0	0.002	0.058	0.231	0.002	0.092	0.092	0
TGMM	0.007	0.059	0.234	0.003	0.092	0.092	0	0.046	0.075	0.286	0.046	0.114	0.122	0
TCU_c	0.006	0.061	0.239	0.004	0.095	0.095	0	0.055	0.106	0.427	0.064	0.191	0.202	0
TET_c	0.006	0.057	0.228	0.003	0.089	0.090	0	0.041	0.068	0.259	0.042	0.108	0.116	0
TEL_c	0.006	0.056	0.220	0.003	0.086	0.086	0	0.037	0.056	0.216	0.036	0.088	0.095	0

(i) MB: median bias; MAD: median absolute deviation; IDR: interdecile range; SD: standard deviation; RMSE: root mean squared error; TP: tail probability.

(ii) TCU_c, TET_c and TEL_c are, respectively, TCU, TET and TEL estimators with $\hat{C}_n(\check{\alpha}, \tilde{\beta})$.

(iii) k_2 is the number of variables in x_2, the true values of τ is 0, and the sample size n is 100.

L.2 Estimation of the probit model using efficient moments

In this section, we consider estimation of the probit model (2.1) using efficient moments. With the normalization $\sigma_u^2 - \omega^2\sigma_\epsilon^2 = 1$, where $\omega = \rho\sigma_u/\sigma_\epsilon$, the log likelihood function for model (2.1) is (Rivers and Vuong, 1988):

$$\ln L(\gamma) = -\frac{n}{2}\log(2\pi\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{n}(y_{2i} - x_{1i}'\alpha_1 - x_{2i}'\alpha_2)^2 + \sum_{i=1}^{n} y_{1i}\ln\Phi\Big(x_{1i}'\kappa + (y_{2i} - x_{1i}'\alpha_1 - x_{2i}'\alpha_2)\omega + y_{2i}\tau\Big)$$
$$+ \sum_{i=1}^{n}(1 - y_{1i})\ln\Big[1 - \Phi\Big(x_{1i}'\kappa + (y_{2i} - x_{1i}'\alpha_1 - x_{2i}'\alpha_2)\omega + y_{2i}\tau\Big)\Big],$$

where $\gamma = (\alpha', \beta')'$ with $\alpha = (\alpha_1', \alpha_2', \sigma_\epsilon^2)'$ and $\beta = (\kappa', \omega, \tau)'$. Compared with the estimation using simple moments in the main text, now α includes the additional parameter σ_ϵ^2, and β includes $(\kappa', \tau)'$ and the additional parameter ω. Let $z_i(\alpha_1, \alpha_2) = [x_{1i}', y_{2i} - x_{1i}'\alpha_1 - x_{2i}'\alpha_2, y_{2i}]'$. Then the score vector over n is $g_n(\gamma) = \big(g'_{nb}(\gamma), g'_{na}(\gamma)\big)'$, where

$$g_{nb}(\gamma) = \frac{1}{n}\frac{\partial \ln L(\gamma)}{\partial \beta} = \frac{1}{n}\sum_{i=1}^{n} \frac{[y_{1i} - \Phi(z_i'(\alpha_1,\alpha_2)\beta)]\phi(z_i'(\alpha_1,\alpha_2)\beta)}{[1-\Phi(z_i'(\alpha_1,\alpha_2)\beta)]\Phi(z_i'(\alpha_1,\alpha_2)\beta)} z_i(\alpha_1,\alpha_2),$$

and

$$g_{na}(\gamma) = \frac{1}{n}\frac{\partial \ln L(\gamma)}{\partial \alpha} = \begin{pmatrix} \frac{1}{n\sigma_\epsilon^2}\sum_{i=1}^n x_i(y_{2i} - x_{1i}'\alpha_1 - x_{2i}'\alpha_2) - \frac{1}{n}\sum_{i=1}^n \frac{[y_{1i}-\Phi(z_i'(\alpha_1,\alpha_2)\beta)]\phi(z_i'(\alpha_2)\beta)}{[1-\Phi(z_i'(\alpha_1,\alpha_2)\beta)]\Phi(z_i'(\alpha_1,\alpha_2)\beta)} x_i\omega \\ -\frac{1}{2\sigma_\epsilon^2} + \frac{1}{2n\sigma_\epsilon^4}\sum_{i=1}^n (y_{2i} - x_{1i}'\alpha_1 - x_{2i}'\alpha_2)^2 \end{pmatrix},$$

with $\phi(\cdot)$ being the standard normal probability density function. We may implement estimation with this efficient moment vector $g_n(\gamma)$. An initial simple consistent estimate $[\check{\alpha}_1', \check{\alpha}_2']'$ of $[\alpha_1', \alpha_2']'$ can be the OLS estimate from the regression of y_{2i} on x_{1i} and x_{2i}, and an estimate $\check{\sigma}_\epsilon^2$ of σ_ϵ^2 can be $\frac{1}{n}\sum_{i=1}^n (y_{2i} - x_{1i}'\check{\alpha}_1 - x_{2i}'\check{\alpha}_2)^2$. This $\check{\alpha} = (\check{\alpha}_1', \check{\alpha}_2', \check{\sigma}_\epsilon^2)'$ is not a first stage estimate via the moment $g_{na}(\gamma)$, which is a rather complex subvector of $g_n(\gamma)$. With $\check{\alpha}$, an initial GMM estimator of β_0 can be derived from $\min_{\beta\in\mathcal{B}} g_{nb}'(\check{\alpha},\beta) g_{nb}(\check{\alpha},\beta)$.[5] With the score vector $g_n(\gamma)$, γ_0 is just identified.[6] Then the GEL estimates are numerically identical to the GMM estimates, and the TGEL estimates are numerically identical to the TGMM estimates. Thus we only report GMM and TGMM estimates of β. Note that TGMM involves two unknown parameters and GMM has more, so we do not use grid search.

Table S.2 reports results on the estimates of β_2. All TPs are zero. TGMM tends to have larger bias and MB than GMM, but it generally has smaller MAD, IDR and SD, especially for cases with $R^2 = 0.01$. In terms of RMSE, TGMM has similar performance as GMM when $R^2 = 0.7$, and it outperforms GMM when $R^2 = 0.01$.

The empirical sizes of various tests are reported in Table S.3.[7] With $n = 100$, the size distortions for cases with $\rho_0 = 0$ are within 3 percentage points, but they can be very large for cases with $\rho_0 = 0.8$, especially when $k_2 = 20$. With $n = 400$, the size distortions are smaller, and the empirical sizes for $\rho_0 = 0.8$ are much closer to the nominal 5%. Table S.4 reports the empirical

[5] As $g_{nb}(\check{\alpha},\beta)$ is just identifiable for β, the estimation is equivalent to solving the equation $g_{nb}(\check{\alpha},\beta) = 0$, which is also the probit ML with generated regressors $z_i(\check{\alpha}_1, \check{\alpha}_2)$.

[6] Since the identification of β_0 needs at least k_β moments and the two-step approaches reduce the number of moments by k_α, β_0 is just identified in the two-step approaches.

[7] For $\mathcal{OPG}$, $\mathcal{G}_{\text{ET}}$ and $\mathcal{G}_{\text{EL}}$, as any restricted estimator can be used, we use the GMM estimator based on the simpler moment vector $[g_{nb}'(\gamma), g_{na}^{*\prime}(\gamma)]'$, where $g_{na}^*(\gamma) = \frac{1}{n}\sum_{i=1}^n x_i(y_{2i} - x_{1i}'\alpha_1 - x_{2i}'\alpha_2)$.

Table S.2: Performance of various estimates of β_2 with efficient moments for the probit model

		$R^2 = 0.7$							$R^2 = 0.01$						
		MB	MAD	IDR	Bias	SD	RMSE	TP	MB	MAD	IDR	Bias	SD	RMSE	TP
$k_2 = 5$, $\rho_0 = 0$	GMM	0.001	0.116	0.442	-0.001	0.173	0.173	0.000	0.288	0.190	1.152	0.150	0.447	0.472	0.000
	TGMM	0.002	0.115	0.440	-0.001	0.172	0.172	0.000	0.292	0.107	0.774	0.164	0.336	0.374	0.000
$k_2 = 5$, $\rho_0 = 0.8$	GMM	0.075	0.124	0.336	0.056	0.137	0.148	0.000	-0.413	0.352	1.424	-0.478	0.528	0.712	0.000
	TGMM	0.096	0.103	0.325	0.068	0.133	0.149	0.000	-0.540	0.303	0.971	-0.554	0.397	0.681	0.000
$k_2 = 20$, $\rho_0 = 0$	GMM	0.004	0.064	0.247	0.004	0.098	0.098	0.000	-0.011	0.420	1.498	-0.005	0.536	0.536	0.000
	TGMM	0.004	0.063	0.248	0.004	0.099	0.099	0.000	-0.011	0.225	0.883	-0.005	0.339	0.339	0.000
$k_2 = 20$, $\rho_0 = 0.8$	GMM	0.033	0.056	0.213	0.033	0.084	0.090	0.000	0.789	0.381	1.549	0.853	0.559	1.020	0.000
	TGMM	0.046	0.053	0.198	0.046	0.077	0.090	0.000	0.842	0.221	0.794	0.859	0.316	0.915	0.000

(i) MB: median bias; MAD: median absolute deviation; IDR: interdecile range; SD: standard deviation; RMSE: root mean squared error; TP: tail probability.

(ii) k_2 is the number of variables in x_2. The sample size n is 100.

powers of various tests of $\tau_0 = 0$. The powers of different tests are generally close and they increase as τ_0 increases.

REFERENCES

Antoine, B., H. Bonnail and E. Renault (2007) On the efficient use of the informational content of estimating equations: Implied probabilities and Euclidean empirical likelihood. *Journal of Econometrics* 138(2), 461–487.

Davidson, R. and J. G. MacKinnon (2004) *Econometric Theory and Methods*. New York: Oxford University Press.

Lee, L.-F. and J. Yu (2012) The $C(\alpha)$-type gradient test for spatial dependence in spatial autoregressive models. *Letters in Spatial and Resource Sciences* 5(3), 119–135.

Newey, W. K. and D. McFadden (1994) Large sample estimation and hypothesis testing. vol. 4 of *Handbook of Econometrics*, chap. 36, Elsevier, 2111–2245.

Newey, W. K. and R. J. Smith (2004) Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72(1), 219–255.

Rivers, D. and Q. H. Vuong (1988) Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of Econometrics* 39(3), 347–366.

Schmidt, M. (2005) minFunc: unconstrained differentiable multivariate optimization in Matlab. `http://www.cs.ubc.ca/~schmidtm/Software/minFunc.html`.

Tsao, M. and F. Wu (2014) Extended empirical likelihood for estimating equations. *Biometrika* 101(3), 703–710.

White, H. (1994) *Estimation, Inference and Specification Analysis*. New York: Cambridge University Press.

Table S.3: Empirical sizes of various tests of $\beta_{20} = 0$ for the probit model with efficient moments

	$k_2 = 5, n = 100$		$k_2 = 5, n = 400$		$k_2 = 20, n = 100$		$k_2 = 20, n = 400$	
	$\rho_0 = 0$	$\rho_0 = 0.8$	$\rho_0 = 0$	$\rho_0 = 0.8$	$\rho_0 = 0$	$\rho_0 = 0.8$	$\rho_0 = 0$	$\rho_0 = 0.8$
$\mathcal{R}_{\text{ET}}$	0.050	0.110	0.046	0.049	0.048	0.330	0.056	0.057
$\mathcal{R}_{\text{EL}}$	0.050	0.107	0.046	0.049	0.048	0.319	0.055	0.056
$\mathcal{R}_{\text{GMM}}$	0.053	0.213	0.046	0.050	0.054	0.450	0.056	0.057
$\mathcal{W}_{\text{GMM}}$	0.046	0.054	0.045	0.048	0.043	0.077	0.053	0.048
$\mathcal{S}_{\text{ET}}$	0.047	0.112	0.045	0.053	0.039	0.286	0.052	0.062
$\mathcal{S}_{\text{EL}}$	0.052	0.102	0.046	0.054	0.047	0.259	0.055	0.062
$\mathcal{G}_{\text{GMM}}$	0.053	0.213	0.046	0.050	0.054	0.450	0.056	0.057
$\mathcal{OPG}$	0.043	0.058	0.045	0.045	0.039	0.124	0.052	0.059
$\mathcal{G}_{\text{ET}}$	0.048	0.060	0.046	0.046	0.042	0.132	0.056	0.059
$\mathcal{G}_{\text{EL}}$	0.049	0.057	0.045	0.045	0.043	0.128	0.055	0.058
$\mathcal{R}_{\text{TET}}$	0.075	0.197	0.052	0.052	0.060	0.782	0.056	0.222
$\mathcal{R}_{\text{TEL}}$	0.077	0.183	0.046	0.053	0.057	0.764	0.055	0.236
$\mathcal{R}_{\text{TGMM}}$	0.061	0.145	0.045	0.045	0.076	0.650	0.054	0.110
$\mathcal{W}_{\text{TGMM}}$	0.045	0.046	0.045	0.042	0.035	0.063	0.052	0.028
$\mathcal{S}_{\text{TET}}$	0.060	0.213	0.050	0.028	0.048	0.730	0.051	0.118
$\mathcal{S}_{\text{TEL}}$	0.070	0.179	0.045	0.027	0.049	0.703	0.052	0.123
$\mathcal{G}_{\text{TGMM}}$	0.061	0.145	0.045	0.045	0.076	0.650	0.054	0.110
$\mathcal{OPG}_{\text{T}}$	0.042	0.073	0.045	0.044	0.037	0.262	0.052	0.061
$\mathcal{G}_{\text{TET}}$	0.044	0.076	0.046	0.045	0.040	0.265	0.053	0.061
$\mathcal{G}_{\text{TEL}}$	0.045	0.075	0.046	0.044	0.038	0.252	0.053	0.061

(i) k_2 is the number of variables in x_2. The nominal size is 5%.

(ii) $\mathcal{R}_{\text{ET}}$: ET ratio test; $\mathcal{R}_{\text{EL}}$: EL ratio test; $\mathcal{R}_{\text{GMM}}$: GMM distance difference test; $\mathcal{W}_{\text{GMM}}$: GMM Wald test; $\mathcal{S}_{\text{ET}}$: score-type test in the ET framework; $\mathcal{S}_{\text{EL}}$: score-type test in the EL framework; $\mathcal{G}_{\text{GMM}}$: GMM gradient test; $\mathcal{OPG}$: OPG test; $\mathcal{G}_{\text{ET}}$: ET gradient test; $\mathcal{G}_{\text{EL}}$: EL gradient test.

(iii) $\mathcal{R}_{\text{TET}}$: TET ratio test; $\mathcal{R}_{\text{TEL}}$: TEL ratio test; $\mathcal{R}_{\text{TGMM}}$: TGMM distance difference test; $\mathcal{W}_{\text{TGMM}}$: TGMM Wald test; $\mathcal{S}_{\text{TET}}$: score-type test in the TET framework; $\mathcal{S}_{\text{TEL}}$: score-type test in the TEL framework; $\mathcal{G}_{\text{TGMM}}$: TGMM gradient test; $\mathcal{OPG}_{\text{T}}$: OPG test in the two-step framework; $\mathcal{G}_{\text{TET}}$: TET gradient test; $\mathcal{G}_{\text{TEL}}$: TEL gradient test.

Table S.4: Empirical powers of various tests of $\beta_{20} = 0$ for the probit model with efficient moments

		$\rho_0 = 0$				$\rho_0 = 0.8$			
		$\tau_0 = 0.1$	$\tau_0 = 0.2$	$\tau_0 = 0.3$	$\tau_0 = 0.4$	$\tau_0 = 0.1$	$\tau_0 = 0.2$	$\tau_0 = 0.3$	$\tau_0 = 0.4$
$k_2 = 5$	$\mathcal{R}_{\text{ET}}$	0.217	0.649	0.906	0.989	0.344	0.749	0.931	0.989
	$\mathcal{R}_{\text{EL}}$	0.216	0.647	0.907	0.989	0.332	0.738	0.925	0.988
	$\mathcal{R}_{\text{GMM}}$	0.222	0.649	0.906	0.989	0.534	0.879	0.982	0.996
	$\mathcal{W}_{\text{GMM}}$	0.208	0.643	0.911	0.990	0.283	0.690	0.909	0.984
	$\mathcal{S}_{\text{ET}}$	0.224	0.682	0.924	0.993	0.385	0.781	0.943	0.986
	$\mathcal{S}_{\text{EL}}$	0.238	0.694	0.930	0.994	0.366	0.766	0.936	0.988
	$\mathcal{G}_{\text{GMM}}$	0.222	0.649	0.906	0.989	0.534	0.879	0.982	0.996
	$\mathcal{OPG}$	0.200	0.622	0.893	0.985	0.281	0.694	0.913	0.985
	$\mathcal{G}_{\text{ET}}$	0.210	0.639	0.898	0.987	0.287	0.700	0.920	0.985
	$\mathcal{G}_{\text{EL}}$	0.205	0.631	0.896	0.986	0.274	0.685	0.911	0.983
	$\mathcal{R}_{\text{TET}}$	0.214	0.623	0.888	0.980	0.405	0.762	0.926	0.983
	$\mathcal{R}_{\text{TEL}}$	0.213	0.619	0.885	0.977	0.380	0.737	0.920	0.981
	$\mathcal{R}_{\text{TGMM}}$	0.207	0.617	0.881	0.978	0.302	0.685	0.888	0.975
	$\mathcal{W}_{\text{TGMM}}$	0.207	0.643	0.911	0.989	0.256	0.642	0.881	0.975
	$\mathcal{S}_{\text{TET}}$	0.165	0.424	0.591	0.746	0.470	0.798	0.942	0.981
	$\mathcal{S}_{\text{TEL}}$	0.174	0.420	0.596	0.760	0.428	0.751	0.920	0.966
	$\mathcal{G}_{\text{TGMM}}$	0.207	0.617	0.881	0.978	0.302	0.685	0.888	0.975
	$\mathcal{OPG}_{\text{T}}$	0.206	0.635	0.904	0.987	0.251	0.650	0.888	0.976
	$\mathcal{G}_{\text{TET}}$	0.218	0.651	0.911	0.990	0.259	0.661	0.894	0.977
	$\mathcal{G}_{\text{TEL}}$	0.215	0.647	0.910	0.989	0.253	0.656	0.892	0.976
$k_2 = 20$	$\mathcal{R}_{\text{ET}}$	0.196	0.574	0.875	0.975	0.644	0.907	0.991	0.999
	$\mathcal{R}_{\text{EL}}$	0.194	0.572	0.872	0.975	0.633	0.899	0.991	0.999
	$\mathcal{R}_{\text{GMM}}$	0.220	0.594	0.881	0.975	0.760	0.958	0.998	0.999
	$\mathcal{W}_{\text{GMM}}$	0.191	0.575	0.883	0.978	0.408	0.800	0.967	0.995
	$\mathcal{S}_{\text{ET}}$	0.202	0.599	0.900	0.975	0.625	0.856	0.897	0.801
	$\mathcal{S}_{\text{EL}}$	0.217	0.618	0.912	0.985	0.611	0.899	0.989	0.998
	$\mathcal{G}_{\text{GMM}}$	0.220	0.594	0.881	0.975	0.760	0.958	0.998	0.999
	$\mathcal{OPG}$	0.179	0.553	0.868	0.974	0.517	0.867	0.981	0.998
	$\mathcal{G}_{\text{ET}}$	0.191	0.570	0.875	0.976	0.526	0.871	0.982	0.998
	$\mathcal{G}_{\text{EL}}$	0.187	0.567	0.871	0.976	0.522	0.867	0.982	0.998
	$\mathcal{R}_{\text{TET}}$	0.192	0.559	0.846	0.955	0.887	0.972	0.993	1.000
	$\mathcal{R}_{\text{TEL}}$	0.190	0.554	0.843	0.952	0.875	0.963	0.993	1.000
	$\mathcal{R}_{\text{TGMM}}$	0.195	0.547	0.825	0.936	0.818	0.945	0.989	1.000
	$\mathcal{W}_{\text{TGMM}}$	0.178	0.566	0.877	0.974	0.324	0.678	0.819	0.875
	$\mathcal{S}_{\text{TET}}$	0.149	0.417	0.648	0.840	0.871	0.961	0.979	0.994
	$\mathcal{S}_{\text{TEL}}$	0.153	0.432	0.654	0.848	0.843	0.947	0.973	0.987
	$\mathcal{G}_{\text{TGMM}}$	0.195	0.547	0.825	0.936	0.818	0.945	0.989	1.000
	$\mathcal{OPG}_{\text{T}}$	0.180	0.546	0.856	0.961	0.363	0.676	0.868	0.943
	$\mathcal{G}_{\text{TET}}$	0.189	0.556	0.865	0.965	0.378	0.681	0.871	0.946
	$\mathcal{G}_{\text{TEL}}$	0.185	0.552	0.859	0.961	0.367	0.677	0.869	0.944

(i) k_2 is the number of variables in x_2. The nominal size is 5%. The sample size n is 100.

(ii) $\mathcal{R}_{\text{ET}}$: ET ratio test; $\mathcal{R}_{\text{EL}}$: EL ratio test; $\mathcal{R}_{\text{GMM}}$: GMM distance difference test; $\mathcal{W}_{\text{GMM}}$: GMM Wald test; $\mathcal{S}_{\text{ET}}$: score-type test in the ET framework; $\mathcal{S}_{\text{EL}}$: score-type test in the EL framework; $\mathcal{G}_{\text{GMM}}$: GMM gradient test; $\mathcal{OPG}$: OPG test; $\mathcal{G}_{\text{ET}}$: ET gradient test; $\mathcal{G}_{\text{EL}}$: EL gradient test.

(iii) $\mathcal{R}_{\text{TET}}$: TET ratio test; $\mathcal{R}_{\text{TEL}}$: TEL ratio test; $\mathcal{R}_{\text{TGMM}}$: TGMM distance difference test; $\mathcal{W}_{\text{TGMM}}$: TGMM Wald test; $\mathcal{S}_{\text{TET}}$: score-type test in the TET framework; $\mathcal{S}_{\text{TEL}}$: score-type test in the TEL framework; $\mathcal{G}_{\text{TGMM}}$: TGMM gradient test; $\mathcal{OPG}_{\text{T}}$: OPG test in the two-step framework; $\mathcal{G}_{\text{TET}}$: TET gradient test; $\mathcal{G}_{\text{TEL}}$: TEL gradient test.