# Online supplementary material to "Second order bias reduction for nonlinear panel data models with fixed effects based on expected quantities "

Martin Schumann
Department of Quantitative Economics, School of Business and Economics, Maastricht University, 6211 LM Maastricht, The Netherlands.
m.schumann@maastrichtuniversity.nl

This online supplement contains additional tables and proofs for the results in the paper. Notation and symbols not defined here are defined in the paper. All sources cited in this supplement are listed in the bibliography of the paper.

## Appendix F. Additional simulation results

Table 4. Simulation results for static logit (columns 3–6) and static probit (columns 7–10) with $n = 100$ using the design in Schumann, Severini, and Tripathi (2021a).

| $T$ | Estimator | Bias | STD | MSE | Coverage | Bias | STD | MSE | Coverage |
|---|---|---|---|---|---|---|---|---|---|
| 3 | $\tilde{\theta}$ | 0.6256 | 0.3100 | 0.4874 | 0.2583 | 0.7653 | 0.2961 | 0.6733 | 0.0261 |
| | $\breve{\theta}_{avg}$ | 0.2738 | 0.2428 | 0.1339 | 0.7422 | 0.6545 | 0.2905 | 0.5127 | 0.0782 |
| | $\breve{\theta}_{exp}$ | 0.1714 | 0.2106 | 0.0737 | 0.8685 | 0.2607 | 0.1996 | 0.1078 | 0.6106 |
| | $\hat{\theta}$ | 0.1062 | 0.2220 | 0.0606 | 0.8982 | 0.0606 | 0.1468 | 0.0252 | 0.9228 |
| | $\tilde{\theta}_{CMLE}$ | 0.0189 | 0.1804 | 0.0329 | 0.9680 | – | – | – | – |
| 4 | $\tilde{\theta}$ | 0.4272 | 0.2360 | 0.2382 | 0.3989 | 0.5448 | 0.2194 | 0.3449 | 0.0729 |
| | $\breve{\theta}_{avg}$ | 0.1655 | 0.1900 | 0.0634 | 0.8400 | 0.4029 | 0.2087 | 0.2059 | 0.2542 |
| | $\breve{\theta}_{exp}$ | 0.1025 | 0.1725 | 0.0403 | 0.9090 | 0.1621 | 0.1497 | 0.0487 | 0.7590 |
| | $\hat{\theta}$ | 0.0463 | 0.1681 | 0.0304 | 0.9331 | 0.0371 | 0.1253 | 0.0171 | 0.9256 |
| | $\hat{\theta}^{FO}_{SPJ}$ | −0.2704 | 0.3646 | 0.2060 | 0.4965 | −0.1548 | 0.4697 | 0.2444 | 0.4284 |
| | $\tilde{\theta}_{CMLE}$ | 0.0127 | 0.1548 | 0.0241 | 0.9520 | – | – | – | – |
| 5 | $\tilde{\theta}$ | 0.3159 | 0.1875 | 0.1349 | 0.5045 | 0.5448 | 0.2194 | 0.3449 | 0.0729 |
| | $\breve{\theta}_{avg}$ | 0.1070 | 0.1547 | 0.0354 | 0.8960 | 0.4029 | 0.2087 | 0.2059 | 0.2542 |
| | $\breve{\theta}_{exp}$ | 0.0661 | 0.1445 | 0.0252 | 0.9355 | 0.1621 | 0.1497 | 0.0487 | 0.7590 |
| | $\hat{\theta}$ | 0.0251 | 0.1388 | 0.0199 | 0.9505 | 0.0371 | 0.1253 | 0.0171 | 0.9256 |
| | $\hat{\theta}^{FO}_{SPJ}$ | −0.2061 | 0.2343 | 0.0974 | 0.5720 | −0.1548 | 0.4697 | 0.2444 | 0.4284 |
| | $\tilde{\theta}_{CMLE}$ | 0.0087 | 0.1340 | 0.0180 | 0.9565 | – | – | – | – |
| 6 | $\tilde{\theta}$ | 0.2457 | 0.1597 | 0.0859 | 0.5820 | 0.3185 | 0.1409 | 0.1213 | 0.2135 |
| | $\breve{\theta}_{avg}$ | 0.0714 | 0.1342 | 0.0231 | 0.9240 | 0.1799 | 0.1258 | 0.0482 | 0.6143 |
| | $\breve{\theta}_{exp}$ | 0.0435 | 0.1284 | 0.0184 | 0.9445 | 0.0761 | 0.1048 | 0.0168 | 0.8854 |
| | $\hat{\theta}$ | 0.0139 | 0.1240 | 0.0156 | 0.9565 | 0.0119 | 0.0940 | 0.0090 | 0.9485 |
| | $\hat{\theta}^{FO}_{SPJ}$ | −0.1578 | 0.1785 | 0.0567 | 0.6470 | −0.2168 | 0.2429 | 0.1060 | 0.4013 |
| | $\hat{\theta}^{SO}_{SPJ}$ | −0.0669 | 0.4796 | 0.2344 | 0.4145 | −0.3317 | 0.7500 | 0.6722 | 0.2135 |
| | $\tilde{\theta}_{CMLE}$ | 0.0042 | 0.1216 | 0.0148 | 0.9600 | – | – | – | – |
| 10 | $\tilde{\theta}$ | 0.1317 | 0.1094 | 0.0293 | 0.7455 | 0.1650 | 0.0864 | 0.0347 | 0.4647 |
| | $\breve{\theta}_{avg}$ | 0.0255 | 0.0974 | 0.0101 | 0.9405 | 0.0651 | 0.0772 | 0.0102 | 0.8679 |
| | $\breve{\theta}_{exp}$ | 0.0155 | 0.0960 | 0.0095 | 0.9465 | 0.0280 | 0.0720 | 0.0060 | 0.9340 |
| | $\hat{\theta}$ | 0.0046 | 0.0946 | 0.0090 | 0.9515 | 0.0045 | 0.0692 | 0.0048 | 0.9510 |
| | $\hat{\theta}^{FO}_{SPJ}$ | −0.0528 | 0.1046 | 0.0137 | 0.8705 | −0.0719 | 0.0905 | 0.0134 | 0.7199 |
| | $\hat{\theta}^{SO}_{SPJ}$ | 0.0129 | 0.1890 | 0.0359 | 0.9515 | 0.0225 | 0.2255 | 0.0514 | 0.4997 |
| | $\tilde{\theta}_{CMLE}$ | 0.0020 | 0.0941 | 0.0089 | 0.9525 | – | – | – | – |

## Appendix G. Derivation of the derivatives in Appendix A

First, we display the derivatives of the FOB of the profile likelihood that are necessary to find an approximation of the FOB that is unbiased up to an error of order $O_{\mathrm{p}}(T^{-3})$. First, recall that

$$\mathcal{B}_i^{(1)}(\theta,\beta_i) = -\frac{\mathbb{E}_{\tau_i}[l_{i01}^2(\theta,\alpha_i;\tau)]}{2T\lambda_{i02}(\theta,\alpha_i;\tau_i)} = -\frac{T^{-1}\sum_{t=1}^T \mathbb{E}_{\tau_i}[l_{it01}^2(\theta,\alpha,\tau)]}{2T^{-1}\sum_{t=1}^T \mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)]}.$$

We start by taking the derivative w.r.t. $\alpha$ and evaluating at $\beta_{i0}$ to obtain

$$\partial_\alpha \mathcal{B}_i^{(1)}(\theta,\beta_i)|_{\beta_i=\beta_{i0}} = -\frac{\mathbb{E}_{\tau_0}[l_{i01}(\theta)l_{i02}(\theta)]}{T\lambda_{i02}(\theta)} + \frac{\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)]\lambda_{i03}(\theta)}{2T\lambda_{i02}^2(\theta)}.$$

Similarly for the derivative w.r.t. $\gamma$,

$$\partial_\gamma \mathcal{B}_i^{(1)}(\theta,\beta_i)|_{\beta_i=\beta_{i0}} = -\frac{\sqrt{T}\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)l_{i10}]}{2T\lambda_{i02}(\theta)} + \frac{\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)]\mathbb{E}_{\tau_0}[l_{i02}(\theta)l_{i10}]}{2T\lambda_{i02}^2(\theta)}.$$

Next, we compute $\partial^2_{\alpha_i\alpha_i}\mathcal{B}_i^{(1)}(\theta,\beta_i)$ and evaluate the resulting expression at $\beta_{i0}$ to obtain

$$\partial^2_{\alpha_i\alpha_i}\mathcal{B}_i^{(1)}(\theta,\beta_i)|_{\beta_i=\beta_{i0}} = \frac{2\lambda_{i03}(\theta)\mathbb{E}_{\tau_0}[l_{i01}(\theta)l_{i02}(\theta)]}{T\lambda_{i02}^2(\theta)} - \frac{\mathbb{E}_{\tau_0}[l_{i02}^2(\theta)]}{T\lambda_{i02}(\theta)} - \frac{\mathbb{E}_{\tau_0}[l_{i01}(\theta)l_{i03}(\theta)]}{T\lambda_{i02}(\theta)} - \frac{\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)]\lambda_{i03}^2(\theta)}{T\lambda_{i02}^3(\theta)} + \frac{\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)]\lambda_{i04}(\theta)}{2T\lambda_{i02}^2(\theta)}.$$

In the next step, we need to take derivatives with respect to $\phi_i$. First, we compute

$$\frac{1}{T}\sum_{t=1}^T \partial_{\phi_i}\lambda_{it02}(\theta,\alpha_i;\tau_i)|_{\beta_i=\beta_{i0}} = \frac{1}{T}\sum_{t=1}^T \mathbb{E}_{\tau_0}[\ell_{it02}(\theta)\ell_{it01}]$$

and

$$\frac{1}{T}\sum_{t=1}^T \partial^2_{\phi_i\phi_i}\lambda_{it02}(\theta,\alpha_i;\tau_i)|_{\beta_i=\beta_{i0}} = \frac{1}{T}\sum_{t=1}^T \mathbb{E}_{\tau_0}[\ell_{it02}(\theta)\ell_{it02}] + \mathbb{E}_{\tau_0}[\ell_{it02}(\theta)\ell_{it01}^2],$$

so that by centering terms and using the fact that $\mathbb{E}_{\tau_0}[\ell_{it01}] = 0$ so that $\ell_{it01} = l_{it01}$, we obtain

$$\partial_{\phi_i}\lambda_{i02}(\theta,\alpha_i;\tau_i)|_{\beta_i=\beta_{i0}} = \mathbb{E}_{\tau_0}[l_{i02}(\theta)l_{i01}] \tag{G.1}$$

and

$$\partial^2_{\phi_i}\lambda_{i02}(\theta,\alpha_i;\tau_i)|_{\beta_i=\beta_{i0}} = \mathbb{E}_{\tau_0}[l_{i02}(\theta)l_{i02}] + \sqrt{T}\mathbb{E}_{\tau_0}[l_{i02}(\theta)l_{i01}^2]. \tag{G.2}$$

Moreover,

$$\frac{1}{T}\sum_{t=1}^T \partial^2_{\phi_i\alpha_i}\lambda_{it02}(\theta,\alpha_i;\tau_i)|_{\beta_i=\beta_{i0}} = \frac{1}{T}\sum_{t=1}^T \mathbb{E}_{\tau_0}[\ell_{it03}(\theta)\ell_{it01}],$$

which after centering yields

$$\partial^2_{\phi_i\alpha_i}\lambda_{i02}(\theta,\alpha_i;\tau_i)|_{\beta_i=\beta_{i0}} = \mathbb{E}_{\tau_0}[l_{i03}(\theta)l_{i01}].$$

Taking the derivative w.r.t. $\phi_i$ of $\mathbb{E}_{\tau_i}[l^2_{it01}(\theta,\alpha_i;\tau_i)]$ is more involved since $l_{it01}(\theta,\alpha_i) = \ell_{it01}(\theta,\alpha_i) - \mathbb{E}_{\tau_i}[\ell_{it01}(\theta,\alpha_i)]$. Therefore, $\phi_i$ appears twice in $\mathbb{E}_{\tau_i}[l^2_{it01}(\theta,\alpha_i;\tau_i)]$ whereas $\theta$ and $\alpha$ appear only once. First, using $\mathbb{E}_{\tau_0}[l_{it01}] = 0$,

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\partial_{\phi_i}\mathbb{E}_{\tau_i}[l^2_{it01}(\theta,\alpha_i)]|_{\beta_i=\beta_{i0}} &= \frac{1}{T}\sum_{t=1}^{T}\Big(\mathbb{E}_{\tau_0}[l^2_{i01}(\theta)l_{it01}] - 2\mathbb{E}_{\tau_0}[l_{it01}(\theta)l_{it01}]\mathbb{E}_{\tau_0}[l_{it01}] \\ &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l^2_{i01}(\theta)l_{it01}]\Big)\end{aligned}$$

and

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\partial^2_{\phi_i}\mathbb{E}_{\tau_i}[l^2_{it01}(\theta,\alpha_i)]|_{\beta_i=\beta_{i0}} = \frac{1}{T}\sum_{t=1}^{T}\Big(&\mathbb{E}_{\tau_0}[l^2_{i01}(\theta)l^2_{i01}] + \mathbb{E}_{\tau_0}[l^2_{i01}(\theta)l_{it02}] \\ &- 2\mathbb{E}_{\tau_0}[l^2_{it01}(\theta)l_{it01}]^2 - \mathbb{E}_{\tau_0}[l^2_{it01}(\theta)]\mathbb{E}_{\tau_0}[l^2_{it01}]\Big).\end{aligned} \tag{G.3}$$

By centering terms and using time independence together with $\mathbb{E}_{\tau_0}[l_{it01}(\theta)] = 0$ for every $\theta \in \Theta$, we therefore have

$$\partial_{\phi_i}\mathbb{E}_{\tau_i}[l^2_{i01}(\theta,\alpha_i)]|_{\beta_i=\beta_{i0}} = \sqrt{T}\mathbb{E}_{\tau_0}[l^2_{i01}(\theta)l_{i01}]. \tag{G.4}$$

Moreover, after some algebra,

$$\begin{aligned}(\partial^2_{\phi_i}\mathbb{E}_{\tau_i}[l^2_{i01}(\theta,\alpha_i)])|_{\beta_i=\beta_{i0}} = \sqrt{T}\mathbb{E}_{\tau_0}[l^2_{i01}(\theta)l_{i02}] - T\mathbb{E}_{\tau_0}[l^2_{i01}(\theta)]\mathbb{E}_{\tau_0}[l^2_{it01}] \\ + T\mathbb{E}_{\tau_0}[l^2_{i01}(\theta)l^2_{i01}] - 2T\mathbb{E}_{\tau_0}[l_{i01}(\theta)l_{i01}]^2.\end{aligned} \tag{G.5}$$

To see this, notice that

$$\begin{aligned}T\mathbb{E}_{\tau_0}[l^2_{i01}(\theta)l^2_{i01}] = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l^2_{it01}(\theta)l^2_{it01}] + 2T\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l_{it01}(\theta)l_{it01}]\right)^2 - \frac{2}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l_{it01}(\theta)l_{it01}]^2 \\ + T\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l^2_{it01}(\theta)]\right)\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l^2_{it01}]\right) - \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l^2_{it01}(\theta)]\mathbb{E}_{\tau_0}[l^2_{it01}].\end{aligned}$$

Now,

$$2T(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l_{it01}(\theta)l_{it01}])^2 = 2T\mathbb{E}_{\tau_0}[l_{i01}(\theta)l_{i01}]^2$$

and

$$T\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l^2_{it01}(\theta)]\right)\left(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l^2_{it01}]\right) = T\mathbb{E}_{\tau_0}[l^2_{i01}(\theta)]\mathbb{E}_{\tau_0}[l^2_{i01}].$$

Therefore,

$$T\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)l_{i01}^2] = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l_{it01}^2(\theta)l_{it01}^2] + 2T\mathbb{E}_{\tau_0}[l_{i01}(\theta)l_{i01}]^2 - \frac{2}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l_{it01}(\theta)l_{it01}]^2 + T\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)]\mathbb{E}_{\tau_0}[l_{i01}^2] - \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l_{it01}^2(\theta)]\mathbb{E}_{\tau_0}[l_{it01}^2]. \quad (G.6)$$

Finally plugging (G.6) into (G.5) again yields (G.3). Next, we consider

$$\partial^2_{\phi_i\alpha_i}\mathbb{E}_{\tau_i}[l_{i01}^2(\theta,\alpha_i)]|_{\beta_i=\beta_{i0}} = 2\mathbb{E}_{\tau_0}[l_{it01}(\theta)l_{it02}(\theta)l_{i01}],$$

so that

$$\partial^2_{\phi_i\alpha_i}\mathbb{E}_{\tau_i}[l_{it01}^2(\theta,\alpha_i)]|_{\beta_i=\beta_{i0}} = 2\sqrt{T}\mathbb{E}_{\tau_0}[l_{i01}(\theta)l_{i02}(\theta)l_{i01}].$$

Using (G.1) and (G.4),

$$\partial_{\phi_i}\mathcal{B}_i^{(1)}(\theta,\beta_i)|_{\beta_i=\beta_{i0}} = \frac{\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)]\partial_{\phi_i}\lambda_{i02}(\theta,\alpha_i;\tau_i)|_{\beta_i=\beta_{i0}}}{2T\lambda_{i02}^2(\theta)} - \frac{\partial_{\phi_i}\mathbb{E}_{\tau_i}[l_{i01}^2(\theta,\alpha_i)]|_{\beta_i=\beta_{i0}}}{2T\lambda_{i02}(\theta)}$$
$$= -\frac{\sqrt{T}\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)l_{i01}]}{2T\lambda_{i02}(\theta)} + \frac{\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)]\mathbb{E}_{\tau_0}[l_{i02}(\theta)l_{i01}]}{2T\lambda_{i02}^2(\theta)}.$$

Similarly, for $\partial^2_{\phi_i}\mathcal{B}_i^{(1)}(\theta,\alpha_i)\big|_{\beta_i=\beta_{i0}}$ we get

$$\partial^2_{\phi_i}\mathcal{B}_i^{(1)}(\theta,\beta_i)|_{\beta_i=\beta_{i0}} = -\frac{\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)](\partial_{\phi_i}\lambda_{i02}(\theta,\alpha_i;\tau_i)|_{\beta_i=\beta_{i0}})^2}{T\lambda_{i02}^3(\theta)} + \frac{\partial_{\phi_i}\lambda_{i02}(\theta,\alpha_i;\tau_i)|_{\beta_i=\beta_{i0}}\partial_{\phi_i}\mathbb{E}_{\tau_i}[l_{i01}^2(\theta,\alpha_i)]|_{\beta_i=\beta_{i0}}}{T\lambda_{i02}^2(\theta)} + \frac{\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)]\partial^2_{\phi_i}\lambda_{i02}(\theta,\alpha_i;\tau_i)|_{\beta_i=\beta_{i0}}}{2T\lambda_{i02}^2(\theta)} - \frac{\partial^2_{\phi_i}\mathbb{E}_{\tau_i}[l_{i01}^2(\theta,\alpha_i)]|_{\beta_i=\beta_{i0}}}{2T\lambda_{i02}(\theta)},$$

hence using (G.1), (G.2), (G.4), and (G.5),

$$\partial^2_{\phi_i}\mathcal{B}_i^{(1)}(\theta,\beta_i)|_{\beta_i=\beta_{i0}} = \frac{\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)]\mathbb{E}_{\tau_0}[l_{i02}(\theta)l_{i01}]^2}{T\lambda_{i02}^3(\theta)} + \frac{\sqrt{T}\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)l_{i01}]\mathbb{E}_{\tau_0}[l_{i02}(\theta)l_{i01}]}{T\lambda_{i02}^2(\theta)} + \frac{\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)](\mathbb{E}_{\tau_0}[l_{i02}(\theta)l_{i02}] + \sqrt{T}\mathbb{E}_{\tau_0}[l_{i02}(\theta)l_{i01}^2])}{2T\lambda_{i02}^2(\theta)} - \frac{\sqrt{T}\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)l_{i02}] + T\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)]\lambda_{i02} + T\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)l_{i01}^2] - 2T\mathbb{E}_{\tau_0}[l_{i01}(\theta)l_{i01}]^2}{2T\lambda_{i02}(\theta)}.$$

Finally, the mixed derivative $\partial^2_{\alpha_i\phi_i}\mathcal{B}_i^{(1)}(\theta,\beta_i)\big|_{\beta_i=\beta_{i0}}$ is

$$\partial^2_{\alpha_i\phi_i}\mathcal{B}_i^{(1)}(\theta,\beta_i)\big|_{\beta_i=\beta_{i0}} = \frac{\lambda_{i03}(\theta)\sqrt{T}\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)l_{i01}]}{2T\lambda_{i02}^2(\theta)} - \frac{\sqrt{T}\mathbb{E}_{\tau_0}[l_{i02}(\theta)l_{i01}(\theta)l_{i01}]}{T\lambda_{i02}(\theta)}$$

$$+\frac{\mathbb{E}_{\tau_0}[l_{i01}(\theta)l_{i02}(\theta)]\mathbb{E}_{\tau_0}[l_{i02}(\theta)l_{i01}]}{T\lambda_{i02}^2(\theta)}-\frac{\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)]\lambda_{i03}(\theta)\mathbb{E}_{\tau_0}[l_{i02}(\theta)l_{i01}]}{T\lambda_{i02}^3(\theta)}$$
$$+\frac{\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)]\mathbb{E}_{\tau_0}[l_{i03}(\theta)l_{i01}]}{2T\lambda_{i02}^2(\theta)}.$$

## Appendix H. Explicit terms for Section C.1

Here we collect complex algebraic terms resulting from the derivation of the first and second order profile likelihood bias.

(1)

$$a_{-3/2}=-\frac{l_{i01}^3(\theta)\lambda_{i03}^2(\theta)}{2\lambda_{i02}^5(\theta)}+\frac{l_{i01}^3(\theta)\lambda_{i04}(\theta)}{6\lambda_{i02}^4(\theta)}+\frac{3l_{i01}^2(\theta)\lambda_{i03}(\theta)l_{i02}(\theta)}{2\lambda_{i02}^4(\theta)}-\frac{l_{i01}(\theta)l_{i02}^2(\theta)}{\lambda_{i02}^3(\theta)}-\frac{l_{i01}^2(\theta)l_{i03}(\theta)}{2\lambda_{i02}^3(\theta)}$$

(2)

$$\begin{aligned}a_{-2}=&\frac{l_{i01}(\theta)l_{i02}^3(\theta)}{\lambda_{i02}^4(\theta)}+\frac{3l_{i01}^2(\theta)l_{i02}(\theta)l_{i03}(\theta)}{2\lambda_{i02}^4(\theta)}+\frac{l_{i01}^3(\theta)l_{i04}(\theta)}{6\lambda_{i02}^4(\theta)}-\frac{3l_{i01}^2(\theta)l_{i02}^2(\theta)\lambda_{i03}(\theta)}{\lambda_{2}^5(\theta)}\\&-\frac{l_{i01}^3(\theta)l_{i03}(\theta)\lambda_{i03}(\theta)}{\lambda_{i02}^5(\theta)}+\frac{5l_{i01}^3(\theta)l_{i02}(\theta)\lambda_{i03}^2(\theta)}{2\lambda_{i02}^6(\theta)}-\frac{5l_{i01}^4(\theta)\lambda_{i03}^3(\theta)}{8\lambda_{i02}^7(\theta)}\\&-\frac{2l_{i01}^3(\theta)l_{i02}(\theta)\lambda_{i04}(\theta)}{3\lambda_{i02}^5(\theta)}+\frac{5l_{i01}^4(\theta)\lambda_{i03}(\theta)\lambda_{i04}(\theta)}{12\lambda_{i02}^6(\theta)}-\frac{l_{i01}^4(\theta)\lambda_{i05}(\theta)}{24\lambda_{i02}^5(\theta)},\end{aligned}$$

(3)

$$\begin{aligned}a_{-3}=&-\frac{l_{i01}(\theta)l_{i02}^4(\theta)}{\lambda_{i02}^5(\theta)}-\frac{3l_{i01}^2(\theta)l_{i02}^2(\theta)l_{i03}(\theta)}{\lambda_{i02}^5(\theta)}-\frac{l_{i01}^3(\theta)l_{i03}^2(\theta)}{2\lambda_{i02}^5(\theta)}-\frac{2l_{i01}^3(\theta)l_{i02}(\theta)l_{i04}(\theta)}{3\lambda_{i02}^5(\theta)}\\&-\frac{l_{i01}^4(\theta)l_{i05}(\theta)}{24\lambda_{i02}^5(\theta)}+\frac{5l_{i01}^2(\theta)l_{i02}^3(\theta)\lambda_{i03}(\theta)}{\lambda_{i02}^6(\theta)}+\frac{5l_{i01}^3(\theta)l_{i02}(\theta)l_{i03}(\theta)\lambda_{i03}(\theta)}{\lambda_{i02}^6(\theta)}\\&+\frac{5l_{i01}^4(\theta)l_{i04}(\theta)\lambda_{i03}(\theta)}{12\lambda_{i02}^6(\theta)}-\frac{15l_{i01}^3(\theta)l_{i02}^2(\theta)\lambda_{i03}^2(\theta)}{2\lambda_{i02}^7(\theta)}-\frac{15l_{i01}^4(\theta)l_{i03}(\theta)\lambda_{i03}^2(\theta)}{8\lambda_{i02}^7(\theta)}\\&+\frac{35l_{i01}^4(\theta)l_{i02}(\theta)\lambda_{i03}^3(\theta)}{8\lambda_{i02}^8(\theta)}-\frac{7l_{i01}^5(\theta)\lambda_{i03}^4(\theta)}{8\lambda_{i02}^9(\theta)}+\frac{5l_{i01}^3(\theta)l_{i02}^2(\theta)\lambda_{i04}(\theta)}{3\lambda_{i02}^6(\theta)}\\&+\frac{5l_{i01}^4(\theta)l_{i03}(\theta)\lambda_{i04}(\theta)}{12\lambda_{i02}^6(\theta)}-\frac{5l_{i01}^4(\theta)l_{i02}(\theta)\lambda_{i03}(\theta)\lambda_{i04}(\theta)}{2\lambda_{i02}^7(\theta)}+\frac{7l_{i01}^5(\theta)\lambda_{i03}^2(\theta)\lambda_{i04}(\theta)}{8\lambda_{i02}^8(\theta)}\\&-\frac{l_{i01}^5(\theta)\lambda_{i04}^2(\theta)}{12\lambda_{i02}^7(\theta)}+\frac{5l_{i01}^4(\theta)l_{i02}(\theta)\lambda_{i05}(\theta)}{24\lambda_{i02}^6(\theta)}-\frac{l_{i01}^5(\theta)\lambda_{i03}(\theta)\lambda_{i05}(\theta)}{8\lambda_{i02}^7(\theta)}+\frac{l_{i01}^5(\theta)\lambda_{i06}(\theta)}{120\lambda_{i02}^6(\theta)}\end{aligned}$$

## Appendix I. Proofs for Section 4.2

In this subsection we collect the proofs for the results in 4.2.

**Proof of (4.2)**. Let $\mathcal{M}$ denote the open ball centered at $\beta_{i0}$ specified in Assumption 4.3. For $k_1, k_2 \in \{1, \ldots, p+2\}$, let derivatives of $\mathcal{B}_i^{(2)}(\theta, \beta_i)$ with respect to components of $\beta_i$ be denoted by

$$\mathcal{B}_{i;k}^{(2)}(\theta, \beta_i) := \frac{\partial \mathcal{B}_i^{(2)}(\theta, \beta_i)}{\partial \beta_k}$$

$$\mathcal{B}_{i;k_1,k_2}^{(2)}(\theta, \beta_i) := \frac{\partial^2 \mathcal{B}_i^{(2)}(\theta, \beta_i)}{\partial \beta_{k_1} \partial \beta_{k_2}}.$$

Using an expansion (ignoring constants),

$$\mathcal{B}_i^{(2)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i} = \mathcal{B}_i^{(2)}(\theta, \beta_{i0}) + \sum_{k=1}^{p+2} \mathcal{B}_{i;k}^{(2)}(\theta, \beta_i)|_{\beta_i = \beta_{i0}} (\hat{\beta}_i - \beta_{i0})_k$$
$$+ \sum_{k_1=1}^{p+2} \sum_{k_2=1}^{p+2} \mathcal{B}_{i;k_1,k_2}^{(2)}(\theta, \beta_i)|_{\beta_i = \bar{\beta}_i} (\hat{\beta}_i - \beta_{i0})_{k_1} (\hat{\beta}_i - \beta_{i0})_{k_2},$$

where $\bar{\beta}_i$ lies between $\hat{\beta}_i$ and $\beta_{i0}$. We now show that partial derivatives of $\mathcal{B}_i^{(2)}(\theta, \beta_i)$ are bounded in probability. To do so, we note that the derivatives of $\mathcal{B}_i^{(2)}(\theta, \beta_i)$ consist of fractions with (powers of) $\mathbb{E}_{\tau_i}[\ell_{i02}(\theta, \alpha_i)]$ evaluated at $\beta_i = \beta_{i0}$ or $\beta_i = \bar{\beta}_i$ in the denominator. Since $\bar{\beta}_i$ lies in $\mathcal{M}$ with probability approaching one as $T \to \infty$, the denominators are bounded away from zero by Assumption 4.3(ii). To bound the numerator, we need to consider the first and second derivatives of

$$\sqrt{T}\mathbb{E}_{\tau_i}[l_{i01}^2(\theta, \alpha_i; \tau_i) l_{i02}(\theta, \alpha_i; \tau_i)], \tag{1}$$

$$\sqrt{T}\mathbb{E}_{\tau_i}[l_{i01}^3(\theta, \alpha_i; \tau_i)], \tag{2}$$

$$\mathbb{E}_{\tau_i}[\ell_{i03}(\theta, \alpha_i; \tau_i)], \tag{3}$$

$$\mathbb{E}_{\tau_i}[l_{i01}^2(\theta, \alpha_i; \tau_i) l_{i02}^2(\theta, \alpha_i; \tau_i)], \tag{4}$$

$$\mathbb{E}_{\tau_i}[l_{i01}^3(\theta(\alpha_i; \tau_i) l_{i03}(\theta, \alpha_i; \tau_i)], \tag{5}$$

$$\mathbb{E}_{\tau_i}[l_{i01}^3(\theta, \alpha_i; \tau_i) l_{i02}(\theta, \alpha_i; \tau_i)] \tag{6}$$

$$\mathbb{E}_{\tau_i}[l_{i01}^4(\theta, \alpha_i; \tau_i)]. \tag{7}$$

Here, (1) and (2) are products of three centered likelihood terms. In order to illustrate how these terms can be bounded, we note that using time-independence together with the mean-zero property of centered likelihood terms of individual $i$ in time period $t$, we can write

$$(1) = \sqrt{T}\mathbb{E}_{\tau_i}[l_{i01}^2(\theta, \alpha_i; \tau_i) l_{i02}(\theta, \alpha_i; \tau_i)] = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{\tau_i}[l_{it01}^2(\theta, \alpha_i; \tau_i) l_{it02}(\theta, \alpha_i; \tau_i)].$$

Similarly, (3) can be expressed as

$$\mathbb{E}_{\tau_i}[\ell_{i03}(\theta, \alpha_i; \tau_i)] = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{\tau_i}[\ell_{it03}(\theta, \alpha_i; \tau_i)].$$

As can be seen from (L.6), we can also express expectations of products of four centered likelihood terms as scaled sums. For example, we can express (7) as

$$\mathbb{E}_{\tau_i}[l_{i01}^4(\theta, \alpha_i; \tau_i)] = \frac{1}{T^2}\mathbb{E}_{\tau_i}[l_{it01}^4(\theta, \alpha_i; \tau_i)] \\ + 3(\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_{\tau_i}[l_{it01}^2(\theta, \alpha_i; \tau_i)])^2 - \frac{3}{T^2}\sum_{t=1}^{T} \mathbb{E}_{\tau_i}[l_{it01}^2(\theta, \alpha_i; \tau_i)]^2.$$

As in the proof of Lemma E.2, we can therefore bound each derivative (uniformly across $\beta_i$) by repetitive use of the Jensen inequality and the Cauchy-Schwarz inequality together with Assumption 4.3(i). Next, since derivatives of first order of $\mathcal{B}_i^{(2)}(\theta, \beta_i)$ do not depend on the outcome data when evaluated at $\beta_i = \beta_{i0}$, (C.14) and (C.13) together with (C.16) imply

$$\mathbb{E}_{\tau_0}[\mathcal{B}_{i;k}^{(2)}(\theta, \beta_i)|_{\beta_i=\beta_{i0}}(\hat{\beta}_i - \beta_{i0})_k] = \mathcal{B}_{i;k}^{(2)}(\theta, \beta_i)|_{\beta_i=\beta_{i0}}\mathbb{E}_{\tau_0}[(\hat{\beta}_i - \beta_{i0})_k] = O_{\mathrm{p}}(T^{-1})$$

for every $k \in \{1, \ldots, p+2\}$. Since the second derivatives of $\mathcal{B}_i^{(2)}(\theta, \beta_i)$ are bounded in probability uniformly across $\beta_i$ and since $(\hat{\beta}_i - \beta_{i0})_{k_1}(\hat{\beta}_i - \beta_{i0})_{k_2}$ is of order $O_{\mathrm{p}}(T^{-1})$ or lower by (C.1), (I.4) and Assumption 4.4(ii,iii), we further see that

$$\mathbb{E}_{\tau_0}[\mathcal{B}_{i;k_1,k_2}^{(2)}(\theta, \beta_i)|_{\beta_i=\bar{\beta}_i}(\hat{\beta}_i - \beta_{i0})_{k_1}(\hat{\beta}_i - \beta_{i0})_{k_2}] = O_{\mathrm{p}}(T^{-1}).$$

In total we thus have $\mathbb{E}_{\tau_0}[\mathcal{B}_i^{(2)}(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}] = \mathcal{B}_i^{(2)}(\theta) + O_{\mathrm{p}}(T^{-1})$ for every $\theta \in \Theta$.

□

**Proof of (C.12)**. Consider the $\psi$-th component of $\partial_{\beta_i}\mathcal{B}_i^{(1)}(\theta, \beta_i)$ where $\psi \in \{1, ..., p+2\}$ denoted as $\mathcal{B}_{i;\psi}^{(1)}(\theta, \beta_i) := \partial_{\beta_\psi}\mathcal{B}_i^{(1)}(\theta, \beta_i)$. As in the proof of (C.9), we denote further derivatives with respect to components of $\beta_i$ by additional indices. We can write $\mathrm{Rem}_i^{(\rho)}(\theta) := \mathrm{Rem}_i^{(\rho,a)}(\theta) + \mathrm{Rem}_i^{(\rho,b)}(\theta)$, where, ignoring constants,

$$\mathrm{Rem}_i^{(\psi,a)}(\theta) := \sum_{k=1}^{p+2} \mathcal{B}_{i;\psi,k}^{(1)}(\theta, \beta_i)|_{\beta_i=\beta_{i0}}(\hat{\beta}_i - \beta_{i0})_k$$

and

$$\text{Rem}_i^{(\psi,b)}(\theta) := \sum_{k_1=1}^{p+2}\sum_{k_2=1}^{p+2} \mathcal{B}^{(1)}_{i;\psi,k_1,k_2}(\theta,\beta_i)|_{\beta_i=\bar{\beta}_i}(\hat{\beta}_i-\beta_{i0})_{k_1}(\hat{\beta}_i-\beta_{i0})_{k_2}+\sum_{k=1}^{p} \mathcal{B}^{(1)}_{i,\psi,k}(\theta,\beta_i)|_{\beta_i=\bar{\beta}_i}(\hat{\beta}_i-\beta_{i0})_k,$$

where $\bar{\beta}_i$ lies between $\hat{\beta}_i$ and $\beta_{i0}$. The rest of the proof now closely follows the arguments used in the proof of (C.9). Since $\bar{\beta}_i \xrightarrow{p} \beta_{i0}$, the first and second derivatives of $\mathcal{B}^{(1)}_{i\rho}(\theta,\beta_i)$ with respect to the respective component of $\beta_i$ are of order $O_{\text{p}}(1)$ when evaluated at $\bar{\beta}_i$ by Lemma E.2. Thus, as $(\hat{\beta}_i - \beta_{i0})_k = O_{\text{p}}(T^{-1/2})$ for $k \in \{1,\dots,p+2\}$ by Assumption 4.4(ii,iii), (C.1) and (I.4), $\text{Rem}_i^{(\rho)}(\theta) = O_{\text{p}}(T^{-1/2})$. Next, we notice again that derivatives of $\mathcal{B}^{(1)}_{i\rho}(\theta,\beta_i)$ do not depend on outcome data when evaluated at $\beta_i = \beta_{i0}$, so that

$$\mathbb{E}_{\tau_0}[\text{Rem}_i^{(\rho,a)}(\theta)] = \sum_{k=1}^{p+2} \mathcal{B}^{(1)}_{i;\psi,k}(\theta,\beta_i)|_{\beta_i=\beta_{i0}}\mathbb{E}_{\tau_0}[(\hat{\beta}_i-\beta_{i0})_k] = O_{\text{p}}(T^{-1})$$

by (C.8), (I.4) and (C.16). Again by Assumption 4.4(ii,iii), we further see that

$$\sum_{k=2}^{p+1}\mathbb{E}_{\tau_0}[\mathcal{B}^{(1)}_{i;\psi,k}(\theta,\beta_i)|_{\beta_i=\bar{\beta}_i}(\hat{\beta}_i-\beta_{i0})_k] = O_{\text{p}}(T^{-1}).$$

Thus, it is left to show that $(\hat{\beta}_i-\beta_{i0})_{k_1}(\hat{\beta}_i-\beta_{i0})_{k_2} = O_{\text{p}}(T^{-1})$, which follows from (C.1), (I.4) and Assumption 4.4(ii,iii). Therefore, $\mathbb{E}_{\tau_0}[\text{Rem}_i^{(\rho,b)}(\theta)] = O_{\text{p}}(T^{-1})$, which finishes the proof. □

**Proof of (C.13)**. Score unbiasedness implies $\mathbb{E}_{\tau_0}[\ell_{it01}(\theta_0,\alpha_{i0})] = 0$, which combined with first order condition of $\alpha_i^*(\theta)$ evaluated at $\theta=\theta_0$ and Assumption 4.2(ii) shows that $\alpha_i^*(\theta_0)=\alpha_{i0}$. Next,

$$\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0} = \delta_i(\tilde{\theta}) + \alpha_i^*(\tilde{\theta}) - \alpha_{i0} \tag{I.1}$$

so that $\mathbb{E}_{\tau_0}[\hat{\alpha}_i(\tilde{\theta})-\alpha_{i0}] = \mathbb{E}[\delta_i(\tilde{\theta})] + \mathbb{E}_{\tau_0}[\alpha_i^*(\tilde{\theta})-\alpha_{i0}]$. By a Taylor expansion,

$$\alpha_i^*(\tilde{\theta}) = \alpha_{i0} + \partial_{\theta'}\alpha_i^*(\theta)|_{\theta=\theta_0}(\tilde{\theta}-\theta_0) + (\tilde{\theta}-\theta_0)'\partial_{\theta\theta'}\alpha_i^*(\theta)|_{\theta=\bar{\theta}_1}(\tilde{\theta}-\theta_0), \tag{I.2}$$

where $\bar{\theta}_1$ lies between $\tilde{\theta}$ and $\theta_0$. By (D.5) $\partial_{\theta'}\alpha_i^*(\theta) = O_{\text{p}}(1)$, which does not depend upon the outcome data when evaluated at $\theta=\theta_0$. Moreover,

$$\partial_{\theta\theta'}\alpha_i^*(\theta) = \frac{2\lambda_{i11}(\theta)\lambda_{i12}(\theta)}{\lambda_{i02}^2(\theta)} - \frac{\lambda_{i11}(\theta)\lambda_{i11}'(\theta)\lambda_{i03}(\theta)}{\lambda_{i02}^3(\theta)} - \frac{\lambda_{i21}(\theta)}{\lambda_{i02}(\theta)}, \tag{I.3}$$

which is component-wise of order $O_{\text{p}}(1)$ uniformly across $\theta$ by Assumption 4.3(i,ii) for $T$ large enough. Hence, $\alpha_i^*(\tilde{\theta}) - \alpha_{i0} = O_{\text{p}}(T^{-1})$ by Assumption 4.4(ii,iii). Therefore, from (I.1) and (C.1), we get

$$\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0} = \delta_i(\tilde{\theta}) + O_{\text{p}}(T^{-1}) = O_{\text{p}}(T^{-1/2}). \tag{I.4}$$

Next,

$$\mathbb{E}_{\tau_0}[\partial_{\theta'}\alpha_i^*(\theta)|_{\theta=\theta_0}(\tilde{\theta}-\theta_0)] = \partial_{\theta'}\alpha_i^*(\theta)|_{\theta=\theta_0}\mathbb{E}_{\tau_0}[\tilde{\theta}-\theta_0] = O_{\mathrm{p}}(T^{-1})$$

by (C.16). By Assumption 4.4(ii,iii), we further have

$$\mathbb{E}_{\tau_0}[(\tilde{\theta}-\theta_0)'\partial_{\theta\theta'}\alpha_i^*(\theta)|_{\theta=\bar{\theta}}(\tilde{\theta}-\theta_0)] = O_{\mathrm{p}}(T^{-2}).$$

In total, $\mathbb{E}_{\tau_0}[\alpha_i^*(\tilde{\theta})-\alpha_{i0}] = O_{\mathrm{p}}(T^{-1})$. Next, we expand

$$\delta_i(\tilde{\theta}) = \delta_i(\theta_0) + \partial_{\theta'}\delta_i(\theta)|_{\theta=\theta_0}(\tilde{\theta}-\theta_0) + (\tilde{\theta}-\theta_0)'\partial_{\theta\theta'}\delta_i(\theta)|_{\theta=\bar{\theta}_2}(\tilde{\theta}-\theta_0), \tag{I.5}$$

where $\bar{\theta}_2$ lies between $\tilde{\theta}$ and $\theta_0$. Using (C.4),

$$\mathbb{E}_{\tau_0}[\partial_{\theta'}\delta_i(\theta)] = T^{-1/2}\partial_{\theta'}a_i(\theta,\beta_{i0}) + O_{\mathrm{p}}(T^{-1}), \tag{I.6}$$

where

$$\partial_{\theta'}a_i(\theta,\beta_{i0}) = \frac{l_{i01}(\theta)(\lambda_{i12}'(\theta)+\lambda_{i03}(\theta)\partial_{\theta'}\alpha_i^*(\theta))}{\lambda_{i02}(\theta)} - \frac{l_{i11}'(\theta)+l_{i02}(\theta)\partial_{\theta'}\alpha_i^*(\theta)}{\lambda_{i02}(\theta)}.$$

Using $\mathbb{E}_{\tau_0}[\partial_{\theta'}a_i(\theta,\beta_{i0})] = 0$, (I.11), independence across individuals and the fact that

$$\Big(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\tau_0}[\partial_{\theta\theta'}\ell_i(\theta)]\big|_{\theta=\theta_0}\Big)^{-1} = O_{\mathrm{p}}(1),$$

we see that the stochastic order of $\mathbb{E}_{\tau_0}[\partial_{\theta'}\delta_i(\theta)|_{\theta=\theta_0}(\tilde{\theta}-\theta_0)]$ is, up to a term of order $O_{\mathrm{p}}(T^{-2})$, determined by

$$\frac{1}{n\sqrt{T}}\mathbb{E}_{\tau_0}[\partial_{\theta'}a_i(\theta,\beta_{i0})\partial_{\theta}\ell_i(\theta,\hat{\alpha}_i(\theta))]\big|_{\theta=\theta_0}.$$

Now, $\partial_{\theta}\ell_i(\theta,\hat{\alpha}_i(\theta)) = \partial_{\theta}\ell_i(\theta) + T^{-1}Rem(\theta)$, where $Rem(\theta) = \partial_{\theta}\mathbb{E}_{\tau_0}[l_{i01}(\theta,\bar{\alpha})]/(2\lambda_{i02}(\theta,\bar{\alpha}))$ with $\bar{\alpha}$ lies between $\hat{\alpha}_i(\theta)$ and $\alpha_i^*(\theta)$. Thus, since $\sup_{\theta} Rem(\theta) = O_{\mathrm{p}}(1)$ by Assumption 4.3, $\partial_{\theta}\ell_i(\theta,\hat{\alpha}_i(\theta)) = \partial_{\theta}\ell_i(\theta)+O_{\mathrm{p}}(T^{-1})$. Since further $\partial_{\theta}\ell_i(\theta) = \lambda_{i10}(\theta)+(l_{i10}(\theta)+l_{i01}(\theta)\partial_{\theta}\alpha_i^*(\theta))/\sqrt{T}$, noting that $\mathbb{E}_{\tau_0}[\partial_{\theta'}a_i(\theta,\beta_{i0})\lambda_{i10}(\theta)] = \mathbb{E}_{\tau_0}[\partial_{\theta'}a_i(\theta,\beta_{i0})]\lambda_{i10}(\theta) = 0$ yields

$$\begin{aligned}&\frac{1}{n\sqrt{T}}\mathbb{E}_{\tau_0}[\partial_{\theta'}a_i(\theta,\beta_{i0})\partial_{\theta}\ell_i(\theta,\hat{\alpha}_i(\theta))]\big|_{\theta=\theta_0}\\ &= \frac{1}{nT}\mathbb{E}_{\tau_0}[\partial_{\theta'}a_i(\theta,\beta_{i0})(l_{i10}(\theta)+l_{i02}(\theta)\partial_{\theta}\alpha_i^*(\theta))]\big|_{\theta=\theta_0} + O_{\mathrm{p}}\Big(\frac{1}{nT^{3/2}}\Big) = O_{\mathrm{p}}((nT)^{-1}) = O_{\mathrm{p}}(T^{-2}),\end{aligned}$$

where the second equation follows from the form of $\partial_{\theta'}a_i(\theta,\beta_{i0})$ together with (E.3) and the third equation follows from Assumption 4.4(iii). Since $\mathbb{E}_{\tau_0}[\partial_{\theta'}a_i(\theta,\beta_{i0})] = 0$ for every $\theta\in\Theta$, the same arguments yield

$$\mathbb{E}_{\tau_0}[\partial_{\theta'}\delta_i(\theta)(\tilde{\theta}-\theta_0)] = O_{\mathrm{p}}(T^{-2}). \tag{I.7}$$

Using again (C.4),

$$\partial_{\theta\theta'}\delta_i(\theta) = -\frac{l_{i01}(\theta)\partial_\theta\lambda_{i02}(\theta)\partial_{\theta'}\lambda_{i02}(\theta)}{\sqrt{T}\lambda_{i02}^2(\theta)} + \frac{\partial_\theta l_{i01}(\theta)\partial_{\theta'}\lambda_{i02}(\theta)}{\sqrt{T}\lambda_{i02}(\theta)} + \frac{l_{i01}(\theta)\partial_{\theta\theta'}\lambda_{i02}(\theta)}{\sqrt{T}\lambda_{i02}(\theta)} + \frac{\partial_\theta l_{i01}(\theta)\partial_{\theta'}\lambda_{i02}(\theta)}{\sqrt{T}\lambda_{i02}^2(\theta)} - \frac{\partial_{\theta\theta'} l_{i01}(\theta)}{\sqrt{T}\lambda_{i02}(\theta)} + O_{\mathrm{p}}(T^{-1}). \tag{I.8}$$

When evaluated at $\theta = \bar{\theta}_2$, where $\bar{\theta}_2 \xrightarrow{p} \theta_0$, the denominators of the terms on the right hand side can be bounded away from zero by Assumption 4.3(ii). Moreover, the numerators can be uniformly bounded over $\Theta$ using Assumption 4.3(i). Therefore, $\partial_{\theta\theta'}\delta_i(\theta)|_{\theta=\bar{\theta}_2} = O_{\mathrm{p}}(T^{-1/2})$. Hence, (C.1) together with Assumption 4.4(ii,iii) implies that

$$\mathbb{E}_{\tau_0}[(\tilde{\theta} - \theta_0)'\partial_{\theta\theta'}\delta_i(\theta)|_{\theta=\bar{\theta}_2}(\tilde{\theta} - \theta_0)] = O_{\mathrm{p}}(T^{-5/2}).$$

Summing up, we therefore obtain

$$\mathbb{E}_{\tau_0}[\delta_i(\tilde{\theta})] = \mathbb{E}_{\tau_0}[\delta_i(\theta_0)] + O_{\mathrm{p}}(T^{-2}).$$

To show the second part of (C.13), notice that by (I.2),

$$\delta_i(\theta)(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0}) = \delta_i(\theta)\delta_i(\tilde{\theta}) - \partial_{\theta'}\alpha_i^*(\theta)|_{\theta=\theta_0}(\tilde{\theta} - \theta_0)\delta_i(\theta) - (\tilde{\theta} - \theta_0)'\partial_{\theta\theta'}\alpha_i^*(\theta)|_{\theta=\bar{\theta}_1}(\tilde{\theta} - \theta_0)\delta_i(\theta),$$

where $\bar{\theta}_1$ lies between $\tilde{\theta}$ and $\theta_0$. Now, we note that by (I.7)

$$\mathbb{E}_{\tau_0}[\partial_{\theta'}\alpha_i^*(\theta)|_{\theta=\theta_0}(\tilde{\theta} - \theta_0)\delta_i(\theta)] = \partial_{\theta'}\alpha_i^*(\theta)|_{\theta=\theta_0}\mathbb{E}_{\tau_0}[(\tilde{\theta} - \theta_0)\delta_i(\theta)] = O_{\mathrm{p}}(T^{-2}),$$

since $\partial_{\theta'}\alpha_i^*(\theta)|_{\theta=\theta_0}$ does not depend upon outcome data. Moreover, since the second derivative of $\alpha_i^*(\theta)$ is uniformly bounded by (I.3) and the discussion following it,

$$\mathbb{E}_{\tau_0}[(\tilde{\theta} - \theta_0)'\partial_{\theta\theta'}\alpha_i^*(\theta)|_{\theta=\bar{\theta}_1}(\tilde{\theta} - \theta_0)\delta_i(\theta)] = O_{\mathrm{p}}(T^{-5/2})$$

by (C.1) and Assumption 4.4(ii,iii). Thus,

$$\mathbb{E}_{\tau_0}[\delta_i(\theta)(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0})] = \mathbb{E}_{\tau_0}[\delta_i(\theta)\delta_i(\tilde{\theta})] + O_{\mathrm{p}}(T^{-2}).$$

Next, using (I.5),

$$\mathbb{E}_{\tau_0}[\delta_i(\theta)\delta_i(\tilde{\theta})] - \mathbb{E}_{\tau_0}[\delta_i(\theta)\delta_i(\theta_0)] = \mathbb{E}_{\tau_0}[\delta_i(\theta)\partial_{\theta'}\delta_i(\theta)|_{\theta=\theta_0}(\tilde{\theta} - \theta_0)] + \mathbb{E}_{\tau_0}[\delta_i(\theta)(\tilde{\theta} - \theta_0)'\partial_{\theta\theta'}\delta_i(\theta)|_{\theta=\bar{\theta}_2}(\tilde{\theta} - \theta_0)],$$

where $\bar{\theta}_2$ lies between $\tilde{\theta}$ and $\theta_0$. By (I.6), $\partial_\theta\delta_i(\theta) = O_{\mathrm{p}}(T^{-1/2})$, so that by Assumption 4.4(ii,iii), the first term on the right hand side is of order $O_{\mathrm{p}}(T^{-2})$. Following the argument after (I.8), $\partial_{\theta\theta'}\delta_i(\theta)|_{\theta=\bar{\theta}_2} = O_{\mathrm{p}}(T^{-1/2})$. Thus, using (C.1) together with Assumption 4.4(ii,iii) implies that the second term on the right hand side is of order $O_{\mathrm{p}}(T^{-3})$. Hence, in total,

$$\mathbb{E}_{\tau_0}[\delta_i(\theta)\delta_i(\tilde{\theta})] - \mathbb{E}_{\tau_0}[\delta_i(\theta)\delta_i(\theta_0)] = O_{\mathrm{p}}(T^{-2}).$$

To show the third part of (C.13), notice that by (I.2) and the discussion following it,

$$(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0})^2 = (\delta_i^2(\tilde{\theta}) + \partial_{\theta'}\alpha_i^*(\theta)|_{\theta=\theta_0}(\tilde{\theta} - \theta_0) + O_{\mathrm{p}}(T^{-2}))^2.$$

Since $\partial_{\theta'}\alpha_i^*(\theta) = O_{\mathrm{p}}(1)$ in each component by (D.5) and Assumption 4.3(i,ii), we see that $(\partial_{\theta'}\alpha_i^*(\theta)|_{\theta=\theta_0}(\tilde{\theta} - \theta_0))^2 = O_{\mathrm{p}}(T^{-2})$ by Assumption 4.4(ii,iii). Hence,

$$\mathbb{E}_{\tau_0}[(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0})^2] = \mathbb{E}_{\tau_0}[\delta_i^2(\tilde{\theta})] + 2\partial_{\theta'}\alpha_i^*(\theta)|_{\theta=\theta_0}\mathbb{E}_{\tau_0}[\delta_i(\tilde{\theta})(\tilde{\theta} - \theta_0)] + O_{\mathrm{p}}(T^{-2}).$$

The second term on the right hand side can be handled with arguments similar to those leading to (I.7), as $a_i(\theta, \beta_{i0})$ and its derivative have mean zero and are of order $O_{\mathrm{p}}(T^{-1/2})$ uniformly across $\theta$. Thus,

$$2\partial_{\theta'}\alpha_i^*(\theta)|_{\theta=\theta_0}\mathbb{E}_{\tau_0}[\delta_i(\tilde{\theta})(\tilde{\theta} - \theta_0)] = O_{\mathrm{p}}(T^{-2}).$$

Further notice that this together with (C.8) implies

$$\mathbb{E}_{\tau_0}[(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0})^2] = O_{\mathrm{p}}(T^{-1}).$$

Next, using an expansion,

$$\delta_i^2(\tilde{\theta}) = \delta_i^2(\theta_0) + 2\delta_i(\bar{\theta}_3)\partial_{\theta'}\delta_i(\theta)|_{\theta=\bar{\theta}_3}(\tilde{\theta} - \theta_0),$$

where again $\bar{\theta}_3$ lies between $\tilde{\theta}$ and $\theta_0$. Since both $\delta_i(\theta)$ and $\partial_{\theta'}\delta_i(\theta)|_{\theta=\bar{\theta}_3}$ are of order $O_{\mathrm{p}}(T^{-1/2})$, Assumption 4.4(ii,iii) implies

$$\mathbb{E}_{\tau_0}[2\delta_i(\bar{\theta}_3)\partial_{\theta'}\delta_i(\theta)|_{\theta=\bar{\theta}_3}(\tilde{\theta} - \theta_0)] = O_{\mathrm{p}}(T^{-2}),$$

which finishes the proof of the third part. □

**Proof of the second equation in (4.4)**. As in (C.15), we use an expansion (ignoring constants) to obtain

$$\begin{aligned}
A(\theta, \beta_i)|_{\theta=\tilde{\theta},\beta_i=\hat{\beta}_i} &= A(\theta_0, \beta_{i0}) + A_\theta(\theta_0, \beta_{i0})'(\tilde{\theta} - \theta_0) + A_{\alpha_i}(\theta_0, \beta_{i0})\delta_i(\tilde{\theta}) + A_\gamma(\theta_0, \beta_{i0})'(\tilde{\theta} - \theta_0) \\
&+ A_{\phi_i}(\theta_0, \beta_{i0})(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0}) + A_{\alpha_i\alpha_i}(\theta, \beta_i)|_{\theta=\bar{\theta},\beta_i=\bar{\beta}_i}\delta_i^2(\tilde{\theta}) + A_{\phi_i\phi_i}(\theta, \beta_i)|_{\theta=\bar{\theta},\beta_i=\bar{\beta}_i}(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0})^2 \\
&+ A_{\alpha_i\gamma}(\theta, \beta_i)'|_{\theta=\bar{\theta},\beta_i=\bar{\beta}_i}(\tilde{\theta} - \theta_0)\delta_i(\tilde{\theta}) + A_{\phi_i\gamma}(\theta, \beta_i)'|_{\theta=\bar{\theta},\beta_i=\bar{\beta}_i}(\tilde{\theta} - \theta_0)(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0}) \\
&+ A_{\alpha_i\theta}(\theta, \beta_i)'|_{\theta=\bar{\theta},\beta_i=\bar{\beta}_i}(\tilde{\theta} - \theta_0)\delta_i(\tilde{\theta}) + A_{\phi_i\theta}(\theta, \beta_i)'|_{\theta=\bar{\theta},\beta_i=\bar{\beta}_i}(\tilde{\theta} - \theta_0)(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0}).
\end{aligned}$$

The rest of the argument can be carried out as in the proof of (4.4), using that derivatives of $A(\theta, \beta_i)$ are of order $O_{\mathrm{p}}(T^{-1})$ together with (C.13), (C.14), (C.8), (C.1), (I.4), (C.16) and Assumption 4.4. □

**Proof of (4.5)**. The arguments used here closely resemble those in the proof of (4.4). For brevity, we only show the first part of (4.5), as the arguments for the second part are analogous. Using an expansion (ignoring constants),

$$\begin{aligned} V(\theta,\beta_i)|_{\theta=\tilde{\theta},\beta_i=\hat{\beta}_i} &= V(\theta_0,\beta_{i0}) + V_\theta(\theta_0,\beta_{i0})'(\tilde{\theta}-\theta_0) + V_{\alpha_i}(\theta_0,\beta_{i0})\delta_i(\tilde{\theta}) + V_\gamma(\theta_0,\beta_{i0})'(\tilde{\theta}-\theta_0) \\ &+ V_{\phi_i}(\theta_0,\beta_{i0})(\hat{\alpha}_i(\tilde{\theta})-\alpha_{i0}) + V_{\alpha_i\alpha_i}(\theta,\beta_i)|_{\theta=\bar{\theta},\beta_i=\bar{\beta}_i}\delta_i^2(\tilde{\theta}) + V_{\phi_i\phi_i}(\theta,\beta_i)|_{\theta=\bar{\theta},\beta_i=\bar{\beta}_i}(\hat{\alpha}_i(\tilde{\theta})-\alpha_{i0})^2 \\ &+ V_{\alpha_i\gamma}(\theta,\beta_i)'|_{\theta=\bar{\theta},\beta_i=\bar{\beta}_i}(\tilde{\theta}-\theta_0)\delta_i(\tilde{\theta}) + V_{\phi_i\gamma}(\theta,\beta_i)'|_{\theta=\bar{\theta},\beta_i=\bar{\beta}_i}(\tilde{\theta}-\theta_0)(\hat{\alpha}_i(\tilde{\theta})-\alpha_{i0}) \\ &+ V_{\alpha_i\theta}(\theta,\beta_i)'|_{\theta=\bar{\theta},\beta_i=\bar{\beta}_i}(\tilde{\theta}-\theta_0)\delta_i(\tilde{\theta}) + V_{\phi_i\theta}(\theta,\beta_i)'|_{\theta=\bar{\theta},\beta_i=\bar{\beta}_i}(\tilde{\theta}-\theta_0)(\hat{\alpha}_i(\tilde{\theta})-\alpha_{i0}). \end{aligned} \tag{I.9}$$

First, we determine the order of derivatives of $V(\theta,\beta_i)$. Recall that by definition

$$V(\theta,\beta_i) = \frac{\mathbb{E}_{\tau_i}[l_{i01}^2(\theta,\alpha_i)]}{T\mathbb{E}_{\tau_i}[\ell_{i02}(\theta,\alpha_i)]^2},$$

so besides the factor $T^{-1}$, derivatives of $V(\theta,\beta_i)$ consist of fractions with (powers of) $\mathbb{E}[\ell_{i02}(\theta,\alpha]$ in the denominator. When evaluated at $\beta_{i0}$ or $\bar{\beta}_i$ with $\bar{\beta}_i \xrightarrow{p} \beta_{i0}$, the latter terms are bounded away from zero by Assumption 4.3 for $T$ large enough. The numerator of the derivatives of $V(\theta,\beta_i)$ consist of derivatives of $\mathbb{E}[\ell_{i02}(\theta,\alpha_i)]$ and $\mathbb{E}[l_{i01}^2(\theta,\alpha_i)]$. As in the proof of Lemma E.2, these derivatives can be uniformly bounded across $\theta$ and $\beta_i$ using Assumption 4.3(i) together with the Cauchy-Schwarz inequality and the Jensen inequality. Taking into account (C.1), (I.4) and Assumption 4.4(ii,iii), we see that $V(\theta,\beta_i)|_{\theta=\tilde{\theta},\beta_i=\hat{\beta}_i} = V(\theta_0,\beta_{i0}) + O_{\mathrm{p}}(T^{-3/2})$. Moreover, we again use that derivatives of $V(\theta,\beta_i)$ do not depend on outcome data when evaluated at $(\theta_0,\beta_{i0})$ so that they act as constants with respect to $\mathbb{E}_{\tau_0}$. Hence, expectations of terms in (I.9) that involve derivatives of $V(\theta,\beta_i)$ of first order can be shown to be of order $O_{\mathrm{p}}(T^{-2})$ using (C.8), (I.4) and (C.16). Terms in (I.9) that involve derivatives of $V(\theta,\beta_i)$ of second order are of order $O_{\mathrm{p}}(T^{-2})$ or lower by (C.1) (I.4) and Assumption 4.4(ii,iii), so that $\mathbb{E}_{\tau_0}[V(\theta,\beta_i)|_{\theta=\tilde{\theta},\beta_i=\hat{\beta}_i}] = V(\theta_0,\beta_{i0}) + O_{\mathrm{p}}(T^{-2})$. □

**Deriving the first order term in** $\mathbb{E}_{\tau_0}[\tilde{\theta}] - \theta_0$. To simplify notation, let $\dim(\theta) = 1$ and $\ell_{nT}^p(\theta) := \frac{1}{n}\sum_{i=1}^n \ell_i(\theta,\hat{\alpha}_i(\theta))$ denote the average profile likelihood. Using the first order condition of the MLE $\tilde{\theta}$ together with an expansion, we obtain

$$0 = \partial_\theta \ell_{nT}^p(\theta)|_{\theta=\tilde{\theta}} = \partial_\theta \ell_{nT}^p(\theta)|_{\theta=\theta_0} + \partial_\theta^2 \ell_{nT}^p(\theta)|_{\theta=\theta_0}(\tilde{\theta}-\theta_0) + (\tilde{\theta}-\theta_0)^2 \partial_\theta^3 \ell_{nT}^p(\theta)|_{\theta=\bar{\theta}}, \tag{I.10}$$

where $\bar{\theta}$ lies between $\tilde{\theta}$ and $\theta_0$. Under Assumptions 4.1(iii) and 4.3(i, iii) the third derivative of each individual profile likelihood $\ell_i(\theta,\hat{\alpha}_i(\theta))$ is bounded in probability uniformly across $\theta$. Since also $\partial_\theta^3 \ell_{nT}^p(\theta)|_{\theta=\bar{\theta}} = O_{\mathrm{p}}(1)$, this implies that the stochastic properties of the last term on the right hand side are determined by $(\tilde{\theta}-\theta_0)^2 = O_{\mathrm{p}}(T^{-2})$, where we have used Assumption

4.4(ii, iii). Next, we consider

$$\partial_\theta^2 \ell_{nT}^p(\theta)|_{\theta=\theta_0}(\tilde{\theta}-\theta_0) = \frac{1}{n}\sum_{i=1}^n \partial_\theta^2 \ell_i(\theta,\hat{\alpha}_i(\theta))(\tilde{\theta}-\theta_0).$$

Using an expansion,

$$\partial_\theta^2 \ell_i(\theta,\hat{\alpha}_i(\theta)) = \partial_\theta^2 \ell_i(\theta) + \partial_\theta^2 \ell_{i01}(\theta)\delta_i(\theta) + \frac{1}{2}\partial_\theta^2 \ell_{i02}(\theta,\bar{\alpha}(\theta))\delta_i^2(\theta),$$

where $\bar{\alpha}(\theta)$ lies between $\hat{\alpha}_i(\theta)$ and $\alpha_i^*(\theta)$. Since $\delta_i(\theta) = O_{\mathrm{p}}(T^{-1/2})$ and $\partial_\theta^2 \ell_{i02}(\theta,\bar{\alpha}(\theta))$ is bounded by Assumption 4.3(i), the last term is of order $O_{\mathrm{p}}(T^{-1})$. Further writing

$$\partial_\theta^2 \ell_{i01}(\theta)\delta_i(\theta) = \partial_\theta^2 \mathbb{E}_{\tau_0}[\ell_{i01}(\theta)]\delta_i(\theta) + \frac{1}{\sqrt{T}}\partial_\theta^2 l_{i01}(\theta)\delta_i(\theta) = O_{\mathrm{p}}(T^{-1}),$$

where we have used that by the definition of the target value $\mathbb{E}_{\tau_0}[\ell_{i01}(\theta)] = 0$ for every $\theta$, so that $\partial_\theta^2 \mathbb{E}_{\tau_0}[\ell_{i01}(\theta)] = 0$. In total, since stochastic orders that hold for each individual also hold for the average over all individuals by Assumption 4.1(iii),

$$\partial_\theta^2 \ell_{nT}^p(\theta)|_{\theta=\theta_0} = \frac{1}{n}\sum_{i=1}^n \partial_\theta^2 \ell_i(\theta) + O_{\mathrm{p}}(T^{-1}).$$

Since $(\tilde{\theta}-\theta_0) = O_{\mathrm{p}}(\frac{1}{\sqrt{nT}}) + O_{\mathrm{p}}(T^{-1}) = O_{\mathrm{p}}(T^{-1})$ by Assumption 4.4(ii,iii),

$$\partial_\theta^2 \ell_{nT}^p(\theta)|_{\theta=\theta_0}(\tilde{\theta}-\theta_0) = \frac{1}{n}\partial_\theta^2 \ell_i(\theta)(\tilde{\theta}-\theta_0) + O_{\mathrm{p}}(T^{-2}).$$

Next, we centralize the second derivative of the target likelihood to obtain

$$\partial_\theta^2 \ell_i(\theta) = \mathbb{E}_{\tau_0}[\partial_\theta^2 \ell_i(\theta)] + \frac{1}{\sqrt{T}}\partial_\theta^2 l_i(\theta)$$

and notice that since $\mathbb{E}_{\tau_0}[\partial_\theta^2 l_i(\theta)] = 0$ by definition, independence over individuals and time together with Assumption 4.3(i,ii) implies

$$\mathbb{E}_{\tau_0}[(\frac{1}{n}\sum_{i=1}^n \frac{1}{\sqrt{T}}\partial_\theta^2 l_i(\theta))^2] = \frac{1}{n^2 T}\sum_{i=1}^n \mathbb{E}_{\tau_0}[(\partial_\theta^2 l_i(\theta))^2] = O_{\mathrm{p}}(\frac{1}{nT}),$$

which shows that

$$\frac{1}{n}\sum_{i=1}^n \frac{1}{\sqrt{T}}\partial_\theta^2 l_i(\theta) = O_{\mathrm{p}}(\frac{1}{\sqrt{nT}}).$$

Therefore,

$$\frac{1}{n}\sum_{i=1}^n \partial_\theta^2 \ell_i(\theta)|_{\theta=\theta_0}(\tilde{\theta}-\theta_0) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\tau_0}[\partial_\theta^2 \ell_i(\theta)]|_{\theta=\theta_0}(\tilde{\theta}-\theta_0) + O_{\mathrm{p}}(T^{-2}),$$

so that in (I.10) we get

$$\tilde{\theta} - \theta_0 = (-\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{\tau_0}[\partial_\theta^2 \ell_i(\theta)|_{\theta=\theta_0})^{-1}\frac{1}{n}\sum_{i=1}^{n} \partial_\theta \ell_i(\theta, \hat{\alpha}_i(\theta)))|_{\theta=\theta_0} + O_{\mathrm{p}}(T^{-2}). \tag{I.11}$$

Taking expectations while noting that $\mathbb{E}_{\tau_0}[\ell_i(\theta, \hat{\alpha}_i(\theta))] = T^{-1}B_i^{(1)}(\theta) + O_{\mathrm{p}}(T^{-2})$, we get

$$\mathbb{E}_{\tau_0}[\tilde{\theta}] - \theta_0 = (-\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{\tau_0}[\partial_\theta^2 \ell_i(\theta)]|_{\theta=\theta_0})^{-1}\frac{1}{nT}\sum_{i=1}^{n} \partial_\theta B_i^{(1)}(\theta) + O_{\mathrm{p}}(T^{-2}).$$

Since $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_{\tau_0}[\partial_\theta^2 \ell_i(\theta)]|_{\theta=\theta_0}$ is bounded away from zero by Assumption 4.3(v) and (D.2) while $\partial_\theta B_i^{(1)}(\theta) = O_{\mathrm{p}}(1)$ by Lemma E.2, this further shows that $\mathbb{E}_{\tau_0}[\tilde{\theta}] - \theta_0 = O_{\mathrm{p}}(T^{-1})$. □

## Appendix J. Bartlett identities

For this section, let $\tau_0 := (\theta', \alpha_i)'$ so that expectations are taken with respect to the conditional density that is evaluated at the same parameter values as the likelihood. Since integration and differentiation is interchangeable,

$$\mathbb{E}_{\tau_0}[\ell_{it01}(\theta, \alpha_i)] = \int_{\mathrm{supp}(Y_{it})} \frac{\partial_{\alpha_i} f_{Y_{it}|\mathbb{X}_i,\theta,\alpha_i}}{f_{Y_{it}|\mathbb{X}_i,\theta,\alpha_i}} f_{Y_{it}|\mathbb{X}_i,\theta,\alpha_i} = \partial_{\alpha_i} \int_{\mathrm{supp}(Y_{it})} f_{Y_{it}|\mathbb{X}_i,\theta,\alpha_i} = 0$$

for every $i$ and $t$. Thus,

$$\mathbb{E}_{\tau_0}[\ell_{i01}(\theta, \alpha_i)] = 0 \quad \forall\ \theta, \alpha.$$

Taking derivatives with respect to $\alpha_i$ while noting that $\frac{\partial}{\partial \alpha_i} f_{\mathbb{y}_i|\mathbb{X}_i,\theta,\alpha_i} = T\ell_{i01}(\theta, \alpha_i) f_{\mathbb{y}_i|\mathbb{X}_i,\alpha_i;\theta}$ and rearranging terms, the second Bartlett identity is

$$T\mathbb{E}_{\tau_0}[\ell_{i01}^2(\theta, \alpha_i)] = -\mathbb{E}_{\tau_0}[\ell_{i02}(\theta, \alpha_i)], \tag{J.1}$$

which can be equivalently written as

$$\mathbb{E}_{\tau_0}[l_{i01}^2(\theta, \alpha_i)] = -\mathbb{E}_{\tau_0}[\ell_{i02}(\theta, \alpha_i)]. \tag{J.2}$$

Repeating this procedure yields the third Bartlett identity

$$T^2\mathbb{E}_{\tau_0}[\ell_{i01}^3(\theta, \alpha_i)] = -\mathbb{E}_{\tau_0}[\ell_{i03}(\theta, \alpha_i)] - 3T\mathbb{E}_{\tau_0}[\ell_{i02}(\theta, \alpha_i)\ell_{i01}(\theta, \alpha_i)],$$

which, noting that $\mathbb{E}_{\tau_0}[l_{ij}(\theta, \alpha_i)] = 0$ for all $i, j \in \mathbb{N}$ by definition, can be written as

$$\mathbb{E}_{\tau_0}[l_{i01}^3(\theta, \alpha_i))] = -T^{-1/2}\mathbb{E}_{\tau_0}[\ell_{i03}(\theta, \alpha_i)] - 3T^{-1/2}\mathbb{E}_{\tau_0}[l_{i02}(\theta, \alpha_i)l_{i01}(\theta, \alpha_i)].$$

One further repetition yields the fourth Bartlett identity

$$\begin{aligned} T^3\mathbb{E}_{\tau_0}[\ell_{i01}^4(\theta, \alpha_i)] = -\mathbb{E}_{\tau_0}[\ell_{i04}(\theta, \alpha_i)] - 4T\mathbb{E}_{\tau_0}[\ell_{i03}(\theta, \alpha_i)\ell_{i01}(\theta, \alpha_i)] \\ - 3T\mathbb{E}_{\tau_0}[\ell_{i02}^2(\theta, \alpha_i)] - 6T^2\mathbb{E}_{\tau_0}[\ell_{i02}(\theta, \alpha_i)\ell_{i01}^2(\theta, \alpha_i)], \end{aligned}$$

which can be written as

$$\begin{aligned}\mathbb{E}_{\tau_0}[l_{i01}^4(\theta,\alpha_i)] &= -T^{-1}\mathbb{E}_{\tau_0}[\ell_{i04}(\theta,\alpha_i)] - 4T^{-1}\mathbb{E}_{\tau_0}[l_{i03}(\theta,\alpha_i)l_{i01}(\theta,\alpha_i)]\\ &\quad - 3T^{-1}\mathbb{E}_{\tau_0}[l_{i02}^2(\theta,\alpha_i)] - 6T^{-1/2}\mathbb{E}_{\tau_0}[l_{i02}(\theta,\alpha_i)l_{i01}^2(\theta,\alpha_i)] + 3(\mathbb{E}_{\tau_0}[\ell_{i02}(\theta,\alpha_i)])^2, \quad (J.3)\end{aligned}$$

where we have used (J.1) in order to show that

$$6T^2\mathbb{E}_{\tau_0}[\ell_{i02}(\theta,\alpha_i)\ell_{i01}^2(\theta,\alpha_i)] = 6\sqrt{T}\mathbb{E}_{\tau_0}[l_{i02}(\theta,\alpha_i)l_{i01}^2(\theta,\alpha_i)] - 6T(\mathbb{E}_{\tau_0}[\ell_{i02}(\theta,\alpha_i)])^2.$$

Taking the derivative with respect to $\theta$ in (J.1) further yields

$$\begin{aligned}T^2\mathbb{E}_{\tau_0}[\ell_{i01}^2(\theta,\alpha_i)\ell_{i10}(\theta,\alpha_i)] &= -\mathbb{E}_{\tau_0}[[\ell_{i12}(\theta,\alpha_i)] - T\mathbb{E}_{\tau_0}[\ell_{i02}(\theta,\alpha_i)\ell_{i10}(\theta,\alpha_i)]\\ &\quad - 2T\mathbb{E}_{\tau_0}[\ell_{i11}(\theta,\alpha_i)\ell_{i10}(\theta,\alpha_i)],\end{aligned}$$

which can be equivalently written as

$$\begin{aligned}\sqrt{T}\mathbb{E}_{\tau_0}[l_{i01}^2(\theta,\alpha_i)l_{i10}(\theta,\alpha_i)] &= -\mathbb{E}_{\tau_0}[\ell_{i12}(\theta,\alpha_i)] - \mathbb{E}_{\tau_0}[l_{i02}(\theta,\alpha_i)l_{i10}(\theta,\alpha_i)]\\ &\quad - 2\mathbb{E}_{\tau_0}[l_{i11}(\theta,\alpha_i)l_{i10}(\theta,\alpha_i)].\end{aligned}$$

## Appendix K. Consistency of $\hat{\theta}$

This section provides an explicit proof of consistency of $\hat{\theta}$, i.e. we show the following theorem.

**Theorem K.1.** *Let Assumptions* 4.1–4.4 *hold. Then,*

$$\hat{\theta} \xrightarrow{p} \theta_0 \qquad \text{as } n, T \to \infty.$$

**Proof**. Let $\hat{Q}_n^*(\theta) := n^{-1}\sum_{i=1}^n \ell_i(\theta, \alpha_{iT}^*(\theta))$ denote the target loglikelihood of $\theta$ for the entire sample, and $\bar{Q}_n^*(\theta) := \mathbb{E}[\hat{Q}_n^*(\theta)]$. By (2.5), we further have

$$\ell_i^*(\theta,\beta_i)|_{\beta_i=\hat{\beta}_i} = \ell_i(\theta,\hat{\alpha}_i(\theta)) + R_i(\theta),$$

where $R_i(\theta) := -T^{-1}\mathcal{B}_i^{(1)}(\theta,\beta_i)|_{\beta_i=\hat{\beta}_i} - T^{-2}\mathcal{B}_i^{(2)}(\theta,\beta_i)|_{\beta_i=\hat{\beta}_i}$.[22] Notice that by Lemma E.2 in Appendix C, there exists an open ball $B$ centered at the true value $\beta_{i0}$ such that $\mathcal{B}_i^{(1)}(\theta,\beta_i)$ is uniformly bounded on $\Theta \times B$. Since consistency of $\hat{\beta}_i$ implies that $\hat{\beta}_i \in B$ with probability approaching one as $T \to \infty$, $\sup_{\theta\in\Theta} T^{-1}|\mathcal{B}_i^{(1)}(\theta,\beta_i)|_{\beta_i=\hat{\beta}_i}| = o_{\mathrm{p}}(1)$. A similar argument shows that $\sup_{\theta\in\Theta} T^{-2}|\mathcal{B}_i^{(2)}(\theta,\beta_i)|_{\beta_i=\hat{\beta}_i}| = o_{\mathrm{p}}(1)$. Therefore, using Assumption 4.1(iii) shows that $n^{-1}\sum_{i=1}^n \sup_{\theta\in\Theta}|R_i(\theta)| = o_{\mathrm{p}}(1)$ as $n, T \to \infty$. Moreover, the profile likelihood approximates the target likelihood, i.e.

$$\ell_i(\theta,\hat{\alpha}_i(\theta)) = \ell_i(\theta,\alpha_i^*(\theta)) + \tilde{R}_i(\theta),$$

[22]Notice the abuse of notation as the remainder term here is different from the one in (4.7).

where

$$\tilde{R}_i(\theta) := \ell_i(\theta, \bar{\alpha}(\theta))\delta_i(\theta)$$

and $\bar{\alpha}(\theta)$ lies between $\hat{\alpha}_i(\theta)$ and $\alpha_i^*(\theta)$. Assumption 4.3(i,ii) and similar arguments as those leading to (C.1) imply that $\mathbb{E}[\sup_\theta |\tilde{R}_i(\theta)|] = o(1)$ as $n, T \to \infty$. Thus, by Assumption 4.1(iii), $n^{-1}\sum_{i=1}^n \sup_{\theta\in\Theta} |\tilde{R}_i(\theta)| = o_{\mathrm{p}}(1)$. In the next step, we show uniform convergence, i.e.

$$\sup_{\theta\in\Theta} |\hat{Q}_n^*(\theta) - \mathbb{E}[\hat{Q}_n^*(\theta)]| = o_{\mathrm{p}}(1). \tag{K.1}$$

Following Newey (1991, Corollary 2.2), a sufficient condition for (K.1) to hold is that: (a) for each $\theta \in \Theta$, $\hat{Q}_n^*(\theta) - \mathbb{E}[\hat{Q}_n^*(\theta)] = o_{\mathrm{p}}(1)$, and (b) for $\theta_1, \theta_2 \in \Theta$, $|\hat{Q}_n^*(\theta_2) - \hat{Q}_n^*(\theta_1)| \leq B_n q(\|\theta_2 - \theta_1\|)$, where $(B_n)_{n\in\mathbb{N}}$ is a sequence of nonnegative random variables such that $\mathbb{E}[B_n] = O(1)$, and $q : [0, \infty) \to [0, \infty)$ is a function such that $q$ is continuous at 0 and $q(0) = 0$. A sufficient condition for (b) is that $\hat{Q}_n^*$ be continuously differentiable with the derivative dominated by a random sequence that is bounded in probability. Thus, (a) follows immediately from Assumptions 4.1(iii) and 4.3(i) and the law of large numbers. Moreover, as $\mathbb{E}[\partial_\theta \ell_i(\theta)] = \mathbb{E}[\ell_{i\theta}(\theta)] + \mathbb{E}[\ell_{i\alpha}(\theta)\partial_\theta\alpha_i^*(\theta)]$ where $\partial_\theta\alpha_i^*(\theta) = -\lambda_{i11}(\theta)/\lambda_{i02}(\theta)$ is bounded by Assumption 4.3, the sufficient condition for (b) is satisfied as well. First, we note that the preceding discussion implies

$$\begin{aligned}\sup_{\theta\in\Theta} |\ell_n^*(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i} - \hat{Q}_n^*(\theta)| &\leq \sup_{\theta\in\Theta} |\ell_n^*(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i} - \frac{1}{n}\sum_{i=1}^n \ell_i(\theta, \hat{\alpha}_i(\theta))| \\ &\quad + \sup_{\theta\in\Theta} |\frac{1}{n}\sum_{i=1}^n \ell_i(\theta, \hat{\alpha}_i(\theta)) - \hat{Q}_n^*(\theta)| = o_{\mathrm{p}}(1).\end{aligned}$$

In addition, by (K.1),

$$\sup_{\theta\in\Theta} |\ell_n^*(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i} - \bar{Q}_n^*(\theta)| = o_{\mathrm{p}}(1). \tag{K.2}$$

Now, let $\mathcal{B} \subset \Theta$ be an open ball around $\theta_0$. By Assumption 4.3(iv), there exists some $\delta > 0$ such that

$$\delta < \bar{Q}_n^*(\theta_0) - \sup_{\theta\in\Theta\setminus\mathcal{B}} \bar{Q}_n^*(\theta).$$

Therefore,

$$\hat{\theta} \in \Theta \setminus \mathcal{B} \Rightarrow \bar{Q}_n^*(\hat{\theta}) \leq \sup_{\theta\in\Theta\setminus\mathcal{B}} \bar{Q}_n^*(\theta) < \bar{Q}_n^*(\theta_0) - \delta$$

for all $n$ large enough. Thus, $\hat{\theta} \in \Theta \setminus \mathcal{B}$ implies

$$\begin{aligned}\delta < \bar{Q}_n^*(\theta_0) - \bar{Q}_n^*(\hat{\theta}) &= \bar{Q}_n^*(\theta_0) - \ell_n^*(\theta, \beta_i)|_{\theta=\tilde{\theta},\beta_i=\hat{\beta}_i} + \ell_n^*(\theta, \beta_i)|_{\theta=\tilde{\theta},\beta_i=\hat{\beta}_i} - \bar{Q}_n^*(\hat{\theta}) \\ &= \bar{Q}_n^*(\theta_0) - \ell_n^*(\theta, \beta_i)|_{\theta=\tilde{\theta},\beta_i=\hat{\beta}_i} + o_{\mathrm{p}}(1)\end{aligned}$$

by (K.2). Since by definition $\ell_n^*(\theta, \beta_i)|_{\theta=\tilde{\theta}, \beta_i=\hat{\beta}_i} \geq \ell_n^*(\theta_0, \beta_i)|_{\beta_i=\hat{\beta}_i}$,

$$\delta < \bar{Q}_n^*(\theta_0) - \ell_n^*(\theta_0, \beta_i)|_{\beta_i=\hat{\beta}_i} + o_{\mathrm{p}}(1).$$

Again using (K.2), we finally see that $\hat{\theta} \notin \mathcal{B} \implies \delta < o_{\mathrm{p}}(1)$, i.e. $\Pr(\hat{\theta} \notin \mathcal{B}) \to 0$. Since $\mathcal{B}$ can be chosen arbitrarily small, $\operatorname{plim}_{n,T\to\infty} \hat{\theta} = \theta_0$.

□

## Appendix L. Feasible approximation of FOB and SOB

In this section, we provide the details on the derivation of the feasible approximations of the FOB and SOB in static panel logit and probit which have been used to generate the simulation results. Henceforth, $\Lambda(u) := e^u/(1+e^u)$, $u \in \mathbb{R}$, is the logistic cdf and $\Phi(u)$ denotes the standard normal cdf.

L.1. **Estimation in the static logit model.** Let $Y_{it} = \mathbb{1}(X_{it}'\theta_0 + \alpha_{i0} + U_{it} > 0)$, where $U_{i1}, \dots, U_{iT} \big| \mathcal{X}_i, \alpha_{i0} \overset{\mathrm{d}}{=} \text{LogisticIID}$. Since the observations are independent across $t$, the scaled loglikelihood for the $i$-th individual is $\ell_i(\theta, \alpha_i) = T^{-1} \sum_{t=1}^T Y_{it} \log \Lambda(X_{it}'\theta + \alpha_i) + (1 - Y_{it}) \log(1 - \Lambda(X_{it}'\theta + \alpha_i))$. While implementing $\hat{\theta}$ may appear to be complicated due to the large number of terms that need to be computed to approximate the FOB and SOB with a bias of order $O_{\mathrm{p}}(T^{-3})$, substantial simplification arise due to the form of the loglikelihood. Let $\Lambda^{(k)}(\cdot)$ denote the $k$-th derivative of $\Lambda(\cdot)$. Implementing $\hat{\theta}$ requires calculating $\Lambda^{(1)}(\cdot) = \Lambda^{(0)}(\cdot)(1 - \Lambda^{(0)}(\cdot))$, $\Lambda^{(2)}(\cdot) = \Lambda^{(1)}(\cdot)(1 - 2\Lambda^{(0)}(\cdot))$ and $\Lambda^{(3)}(\cdot) = \Lambda^{(2)}(\cdot)(1 - 2\Lambda^{(0)}(\cdot)) - 2(\Lambda^{(1)}(\cdot))^2$. Now, taking derivatives of the loglikelihood with respect to $\alpha_i$ yields $\ell_{i01}(\theta, \alpha_i) = T^{-1} \sum_{t=1}^T (Y_{it} - \Lambda(X_{it}'\theta + \alpha_i))$ and $\ell_{i02}(\theta, \alpha_i) = -T^{-1} \sum_{t=1}^T \Lambda^{(1)}(X_{it}'\theta + \alpha_i)$. The latter expression does not depend on the outcome variables, which implies that derivatives of at least second order are not affected by taking expectations conditional on explanatory variables. Therefore, $\lambda_{i0k}(\theta, \alpha_i) = -T^{-1} \sum_{t=1}^T \Lambda^{(k-1)}(X_{it}'\theta + \alpha_i)$. This in turn also implies $l_{i0k} = 0$ for $k \geq 2$. Moreover, by definition of $\alpha_i^*(\theta)$,

$$0 = \mathbb{E}_{\tau_0}[\ell_{i01}(\theta)] \Leftrightarrow \frac{1}{T} \sum_{t=1}^T \Lambda(X_{it}'\theta_0 + \alpha_{i0}) = \frac{1}{T} \sum_{t=1}^T \Lambda(X_{it}'\theta + \alpha_i^*(\theta)),$$

which implies that $\ell_{i01}(\theta) = \ell_{i01}(\theta_0, \alpha_{i0})$. Hence, moments of $\ell_{i01}(\theta)$ coincide with moments of $\ell_{i01}$. Since the likelihood evaluated at the true values satisfies the Bartlett identities, it is possible to find simplified expressions for moments of $l_{i01}(\theta)$. Using the fact that $l_{i01}(\theta) = l_{i01}$ together with (J.2) for example yields

$$\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)] = \mathbb{E}_{\tau_0}[l_{i01}^2] = -\lambda_{i02}, \tag{L.1}$$

while with (J.3) together with $l_{i02} = \mathbb{E}_{\tau_0}[l_{i01}(\theta)] = 0$ in the second equation, we get

$$\mathbb{E}_{\tau_0}[l_{i01}^3(\theta)] = \mathbb{E}_{\tau_0}[l_{i01}^3] = -T^{-1/2}\lambda_{i03}. \tag{L.2}$$

Further using (J.3) and $l_{i02} = l_{i03} = 0$

$$\mathbb{E}_{\tau_0}[l_{i01}^4(\theta)] = \mathbb{E}_{\tau_0}[l_{i01}^4] = -T^{-1}\lambda_{i04} + 3\lambda_{i02}^2. \tag{L.3}$$

It is therefore not necessary to compute higher moments of $l_{i01}(\theta)$ in the static logit example.[23] Consequently, (4.1) simplifies to

$$\begin{aligned}\mathbb{E}_{\tau_0}[\ell_i(\theta,\hat{\alpha}_i(\theta)) - \ell_i(\theta,\alpha_i^*(\theta))] = &-\frac{1}{T}\frac{\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)]}{2\lambda_{i02}(\theta)} - \frac{1}{T^{3/2}}\frac{\mathbb{E}_{\tau_0}[l_{i01}^3(\theta)]\lambda_{i03}(\theta)}{6\lambda_{i02}^3(\theta)} \\ &+ \frac{1}{T^2}\Big(\frac{\mathbb{E}_{\tau_0}[l_{i01}^4(\theta)]\lambda_{i04}(\theta)}{24\lambda_{i02}^4(\theta)} - \frac{\mathbb{E}_{\tau_0}[l_{i01}^4(\theta)]\lambda_{i03}^2(\theta)}{8\lambda_{i02}^5(\theta)}\Big) + O_{\mathrm{p}}(T^{-3}).\end{aligned}$$

By using (L.1), (L.2) and (L.3), this expression can be written as

$$\mathbb{E}_{\tau_0}[\ell_i(\theta,\hat{\alpha}_i(\theta)) - \ell_i(\theta)] = \frac{\lambda_{i02}}{2T\lambda_{i02}(\theta)} + \frac{\lambda_{i03}\lambda_{i03}(\theta)}{6T^2\lambda_{i02}^3(\theta)} + \frac{\lambda_{i02}^2\lambda_{i04}(\theta)}{8T^2\lambda_{i02}^4(\theta)} - \frac{3\lambda_{i03}^2(\theta)\lambda_{i02}^2}{8T^2\lambda_{i02}^5(\theta)} + O_{\mathrm{p}}(T^{-3}),$$

which does not involve moments of $l_{i01}(\theta)$. It can easily be seen that the FOB in logit is

$$B_i^{(1)}(\theta) = \frac{\lambda_{i02}}{2T\lambda_{i02}(\theta)},$$

which, for $\beta_i = (\gamma', \phi_i, \alpha_i)'$, can be estimated by $T^{-1}\mathcal{B}_i^{(1)}(\theta,\beta_i)|_{\beta_i=\hat{\beta}_i}$ with

$$T^{-1}\mathcal{B}_i^{(1)}(\theta,\beta_i) = \frac{\ell_{i02}(\gamma,\phi_i)}{2T\ell_{i02}(\theta,\alpha_i)},$$

where we have used that $\ell_{i0k}(\theta,\alpha_i)$ does not depend on the outcome variables and therefore coincides with $\lambda_{i0k}(\theta,\alpha_i)$ for $k \geq 2$. Taking derivatives for the derivation of a feasible approximation of the FOB that is unbiased up to order $O_{\mathrm{p}}(T^{-3})$ is hence facilitated. We show this first with $\partial_{\alpha_i}\mathcal{B}_i^{(1)}(\theta,\beta_i)\big|_{\beta_i=\beta_{i0}}$, which can be written as

$$\mathcal{B}_{i\alpha_i}^{(1)}(\theta,\beta_{i0}) = -\frac{\lambda_{i02}\lambda_{i03}(\theta)}{2\lambda_{i02}^2(\theta)}.$$

Similarly,

$$\mathcal{B}_{i\gamma}^{(1)}(\theta,\beta_{i0}) = \frac{\lambda_{i12}}{2\lambda_{i02}(\theta)},$$

and

$$\mathcal{B}_{i\phi_i}^{(1)}(\theta,\beta_{i0}) = -\frac{\sqrt{T}\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)l_{i01}]}{2\lambda_{i02}(\theta)} = \frac{\lambda_{i03}}{2T\lambda_{i02}(\theta)}.$$

[23]Since the approach of DS does not involve preliminary estimators of $\theta_0$ and $\alpha_{i0}$, using the Bartlett identities is not possible in the derivation of their estimator.

For the second derivative of $\mathcal{B}_i^{(1)}(\theta, \beta_i)$ with respect to $\alpha_i$ we obtain

$$\mathcal{B}^{(1)}_{i\alpha_i\alpha_i}(\theta, \beta_{i0}) = \frac{\lambda_{i02}\lambda^2_{i03}(\theta)}{\lambda^3_{i02}(\theta)} - \frac{\lambda_{i02}\lambda_{i04}(\theta)}{2\lambda^2_{i02}(\theta)},$$

whereas for the second derivative with respect to $\phi_i$ we get

$$\mathcal{B}^{(1)}_{i\phi_i\phi_i}(\theta, \beta_{i0}) = -\frac{T\mathbb{E}_{\tau_0}[l^2_{i01}(\theta)l^2_{i01}]}{2\lambda_{i02}(\theta)} = \frac{\lambda_{i04}}{2\lambda_{i02}(\theta)}.$$

Finally

$$\mathcal{B}^{(1)}_{i\alpha_i\phi_i}(\theta, \beta_{i0}) = -\frac{\lambda_{i03}(\theta)\lambda_{i03}}{2\lambda^2_{i02}(\theta)}.$$

L.1.1. *Remark on the use of "score-factors" in static logit.* We show here that in general $\mathbb{E}_{\tau_0}[\ell_{it01}(\theta)] \neq 0$ for $\theta \neq \theta_0$ in the static logit example. This implies that the use of "score-factors" which are crucial to the bias correction approach of DS are not permitted for the correction of the conditional profile likelihood bias $\mathbb{E}_{\tau_0}[\ell_i(\theta, \hat{\alpha}_i(\theta)) - \ell_i(\theta, \alpha^*_i(\theta))]$. Recall that

$$\ell_{i01}(\theta, \alpha_i) = \frac{1}{T}\sum_{t=1}^{T} Y_{it} - \Lambda(X'_{it}\theta + \alpha_i).$$

Notice that by definition of the target likelihood

$$\mathbb{E}_{\tau_0}[\ell_{i01}(\theta)] = \frac{1}{T}\sum_{t=1}^{T} \Lambda(X'_{it}\theta_0 + \alpha_{i0}) - \Lambda(X'_{it}\theta + \alpha^*_i(\theta)) = 0.$$

If the score factor property holds with respect to $\mathbb{E}_{\tau_0}$, we further have $\mathbb{E}_{\tau_0}[\ell_{it01}(\theta)] = 0$. For $T = 2$ this implies that

$$\Lambda(X'_{i1}\theta_0 + \alpha_{i0}) = \Lambda(X'_{i1}\theta + \alpha^*_i(\theta)) \Leftrightarrow X'_{i1}\theta_0 + \alpha_{i0} = X'_{i1}\theta + \alpha^*_i(\theta)$$
$$\Lambda(X'_{i2}\theta_0 + \alpha_{i0}) = \Lambda(X'_{i2}\theta + \alpha^*_i(\theta)) \Leftrightarrow X'_{i2}\theta_0 + \alpha_{i0} = X'_{i2}\theta + \alpha^*_i(\theta),$$

since the CDF $\Lambda()$ is strictly increasing. Rewriting this statement yields

(A) $X'_{i1}(\theta_0 - \theta) = \alpha^*_i(\theta) - \alpha_{i0}$
(B) $X'_{i2}(\theta_0 - \theta) = \alpha^*_i(\theta) - \alpha_{i0}$.

Subtracting equation (B) from (A) finally yields

$$(X_{i1} - X_{i2})(\theta_0 - \theta) = 0.$$

This equation can again only be satisfied if $\theta = \theta_0$ or when there is no time variation in $X_{it}$, which would imply that $X_{i1} = X_{i2}$. However, the latter is not possible when the true parameter $\theta_0$ is assumed to be identified, as this rules out any time-invariant regressors.

L.2. **Estimation in the static probit model.** Let $Y_{it} = \mathbb{1}(X'_{it}\theta_0 + \alpha_{i0} + U_{it} > 0)$, where $U_{i1}, \ldots, U_{iT} \big| \mathcal{X}_i, \alpha_{i0} \overset{\text{d}}{=} \text{NIID}(0, 1)$. By independence across $t$, the loglikelihood for the $i$-th individual is

$$\ell_i(\theta, \alpha_i) = \frac{1}{T}\sum_{t=1}^{T}[Y_{it}\log\Phi(X_{it}'\theta + \alpha_i) + (1 - Y_{it})\log(1 - \Phi(X_{it}'\theta + \alpha_i))],$$

where $\Phi(\cdot)$ denotes the standard normal cdf. Taking the first derivative with respect to $\alpha_i$ yields

$$\ell_{i01}(\theta, \alpha_i) = \frac{1}{T}\sum_{t=1}^{T}[(Y_{it} - \Phi(X_{it}'\theta + \alpha_i))G(X_{it}'\theta + \alpha_i)],$$

where $G(u) := \frac{\varphi(u)}{\Phi(u)\Phi(-u)}$ and $\varphi(\cdot)$ denotes the standard normal pdf. Thus,

$$\mathbb{E}[\ell_{i01}(\theta, \alpha_i); \gamma, \phi_i] = \frac{1}{T}\sum_{t=1}^{T}[(\Phi(X_{it}'\gamma + \phi_i) - \Phi(X_{it}'\theta + \alpha_i))G(X_{it}'\theta + \alpha_i)].$$

Unlike in static logit, derivatives of the likelihood with respect to $\alpha_i$ do depend on the outcome variables because of the presence of $G$.[24] Hence, in general $\ell_{i0k}(\theta, \alpha_i) \neq \lambda_{i0k}(\theta, \alpha_i)$ and $l_{i0k}(\theta, \alpha_i) \neq 0$ for $k \in \mathbb{N}$. However, the likelihood provides some structure leading to patterns that can be useful in finding $\hat{\theta}$: Notice first that $G(u) = h(u) + h(-u)$, where $h(u)/\Phi(-u)$ denotes the Gaussian hazard function. For $k \in \mathbb{N}$, let further $h^{(k)}(\cdot)$ denote the $k$-th derivative of $h$. It is then easy to confirm that the necessary derivatives are

(1) $h^{(1)}(x) = h(x)(h(x) - x)$,
(2) $h^{(2)}(x) = h^{(1)}(x)(h(x) - x) + h(x)(h^{(1)}(x) - 1)$ and
(3) $h^{(3)}(x) = h^{(2)}(x)(h(x) - x) + 2h^{(1)}(x)(h^{(1)}(x) - 1) + h(x)h^{(2)}(x)$.

In order to avoid repeating the argument, let $k \in \mathbb{N}_0$ and

$$G_{it}^{(k)} := \frac{\partial^k}{\partial \alpha_i^k} G(X_{it}'\theta + \alpha_i).$$

Then, $G_{it}^{(k)} = (-1)^k h^{(k-1)}(-X_{it}'\theta - \alpha_i) + h^{(k-1)}(X_{it}'\theta + \alpha_i)$. Next, let

$$\Phi_{it}^{(k)} := \frac{\partial^k}{\partial \alpha_i^k} \Phi(X_{it}'\theta + \alpha_i).$$

Then,

(1) $\Phi_{it}^{(1)} = \varphi(X_{it}'\theta + \alpha_i)$
(2) $\Phi_{it}^{(2)} = -(X_{it}'\theta + \alpha_i)\varphi(X_{it}'\theta + \alpha_i)$
(3) $\Phi_{it}^{(3)} = \varphi(X_{it}'\theta + \alpha_i)[(X_{it}'\theta + \alpha_i)^2 - 1]$
(4) $\Phi_{it}^{(4)} = \varphi(X_{it}'\theta + \alpha_i)[3(X_{it}'\theta + \alpha_i) - (X_{it}'\theta + \alpha_i)^3]$.

[24]The function $G(\cdot)$ is dubbed the "probit weight function" in Schumann and Tripathi (2018). It is further shown there that $G$ is convex and $U$-shaped on the real line.

Moreover, we indicate evaluation at the true values with "0", e.g. $\Phi_{it0}^{(2)} := \Phi_{it}^{(2)}(X_{it}'\theta+\alpha_i)|_{\theta=\theta_0,\alpha_i=\alpha_{i0}}$. We then find the following:

(1)

$$\lambda_{i02}(\theta,\alpha_i) = \frac{1}{T}\sum_{t=1}^{T}(\Phi_{it0}-\Phi_{it})G_{it}^{(1)} - \Phi_{it}^{(1)}G_{it}^{(0)}$$

(2)

$$\lambda_{i03}(\theta,\alpha_i) = \frac{1}{T}\sum_{t=1}^{T}(\Phi_{it0}-\Phi_{it})G_{it}^{(2)} - 2\Phi_{it}^{(1)}G_{it}^{(1)} - \Phi_{it}^{(2)}G_{it}^{(0)}$$

(3)

$$\lambda_{i04}(\theta,\alpha_i) = \frac{1}{T}\sum_{t=1}^{T}(\Phi_{it0}-\Phi_{it})G_{it}^{(3)} - 3\varphi_{it}G_{it}^{(2)} - 3\Phi_{it}^{(2)}G_{it}^{(1)} - \Phi_{it}^{(3)}G_{it}^{(0)}$$

(4)

$$\lambda_{11}(\theta,\alpha_i) = \frac{1}{T}\sum_{t=1}^{T}X_{it}[(\Phi_{it0}-\Phi_{it})G_{it}^{(1)} - \Phi_{it}^{(1)}G_{it}^{(0)}]$$

(5)

$$\lambda_{12}(\theta,\alpha_i) = \frac{1}{T}\sum_{t=1}^{T}X_{it}[(\Phi_{it0}-\Phi_{it})G_{it}^{(2)} - 2\Phi_{it}^{(1)}G_{it}^{(1)} - \Phi_{it}^{(2)}G_{it}^{(0)}]$$

(6)

$$\lambda_{20}(\theta,\alpha_i) = \frac{1}{T}\sum_{t=1}^{T}X_{it}X_{it}'[(\Phi_{it0}-\Phi_{it})G_{it}^{(1)} - \Phi_{it}^{(1)}G_{it}^{(0)}]$$

Moreover,

$$l_{i10}(\theta,\alpha_i) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_{it}(Y_{it}-\Phi_{it0})G_{it}^{(0)},$$

and, for $k \in \mathbb{N}_0$,

$$l_{i0k}(\theta,\alpha_i) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}(Y_{it}-\Phi_{it0})G_{it}^{(k-1)}.$$

Given these simple forms of the centralized likelihoods, it is easy to derive the expectations of products of centralized likelihood derivatives that are needed here. To illustrate the computation, let $a, b, c, \cdots \in \mathbb{N}$ and notice that time-independence together with $\mathbb{E}_{\tau_0}[l_{it}(\theta,\alpha_i)] = 0$ implies

$$\mathbb{E}_{\tau_0}[l_{iab}(\theta_1,\alpha_1)l_{icd}(\theta_2,\alpha_2)] = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l_{itab}(\theta_1,\alpha_1)l_{itcd}(\theta_2,\alpha_2)]. \tag{L.4}$$

The same argument yields

$$\mathbb{E}_{\tau_0}[l_{iab}(\theta_1,\alpha_1)l_{icd}(\theta_2,\alpha_2)l_{ief}(\theta_3,\alpha_3)] = \frac{1}{T^{3/2}}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l_{itab}(\theta_1,\alpha_1)l_{itcd}(\theta_2,\alpha_2)l_{itef}(\theta_3,\alpha_3)]. \quad \text{(L.5)}$$

Further, the expectation of four centralized likelihood terms can be derived as

$$\begin{aligned}
&\mathbb{E}_{\tau_0}[l_{iab}(\theta_1,\alpha_1)l_{icd}(\theta_2,\alpha_2)l_{ief}(\theta_3,\alpha_3)l_{igh}(\theta_4,\alpha_4)] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l_{itab}(\theta_1,\alpha_1)l_{itcd}(\theta_2,\alpha_2)l_{itef}(\theta_3,\alpha_3)l_{itgh}(\theta_4,\alpha_4)] \\
&+ \Big(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l_{itab}(\theta_1,\alpha_1)l_{itcd}(\theta_2,\alpha_2)]\Big)\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l_{itef}(\theta_3,\alpha_3)l_{itgh}(\theta_4,\alpha_4)]\Big) \\
&- \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l_{itab}(\theta_1,\alpha_1)l_{itcd}(\theta_2,\alpha_2)]\mathbb{E}_{\tau_0}[l_{itef}(\theta_3,\alpha_3)l_{itgh}(\theta_4,\alpha_4)] \\
&+ \Big(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l_{itab}(\theta_1,\alpha_1)l_{itef}(\theta_3,\alpha_3)]\Big)\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l_{itcd}(\theta_2,\alpha_2)l_{itgh}(\theta_4,\alpha_4)]\Big) \\
&- \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l_{itab}(\theta_1,\alpha_1)l_{itef}(\theta_3,\alpha_3)]\mathbb{E}_{\tau_0}[l_{itcd}(\theta_2,\alpha_2)l_{itgh}(\theta_4,\alpha_4)] \\
&+ \Big(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l_{itab}(\theta_1,\alpha_1)l_{itgh}(\theta_4,\alpha_4)]\Big)\Big(\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l_{itcd}(\theta_2,\alpha_2)l_{itef}(\theta_3,\alpha_3)]\Big) \\
&- \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[l_{itab}(\theta_1,\alpha_1)l_{itgh}(\theta_4,\alpha_4)]\mathbb{E}_{\tau_0}[l_{itcd}(\theta_2,\alpha_2)l_{itef}(\theta_3,\alpha_3)]. \quad \text{(L.6)}
\end{aligned}$$

For example, using the formula for central moments of the Bernoulli distribution together with (L.4), (L.5) and (L.6), we now get

(i)

$$\mathbb{E}_{\tau_0}[l_{i01}^2(\theta,\alpha_i)] = \frac{1}{T}\sum_{t=1}^{T}\Phi_{it0}(1-\Phi_{it0})(G_{it}^{(0)})^2,$$

(ii)

$$\begin{aligned}
\mathbb{E}_{\tau_0}[l_{i01}^3(\theta,\alpha_i)] =& \frac{1}{T^{3/2}}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[(Y_{it}-\Phi_{it0})^3](G_{it}^{(0)})^3 \\
=& \frac{1}{T^{3/2}}\sum_{t=1}^{T}\Phi_{it0}(1-\Phi_{it0})(1-2\Phi_{it0})(G_{it}^{(0)})^3
\end{aligned}$$

(iii) and

$$\mathbb{E}_{\tau_0}[l_{i01}^4(\theta,\alpha_i)] = \frac{1}{T^2}\sum_{t=1}^T \Phi_{it0}(1-\Phi_{it0})(3\Phi_{it0}^2 - 3\Phi_{it0}+1)(G_{it}^{(0)})^4 + (\frac{1}{T}\sum_{t=1}^T (G_{it}^{(0)})^2\Phi_{it0}(1-\Phi_{it0}))^2$$
$$-\frac{1}{T^2}\sum_{t=1}^T(\Phi_{it0}(1-\Phi_{it0})G_{it}^{(0)})^2 = (\frac{1}{T}\sum_{t=1}^T (G_{it}^{(0)})^2\Phi_{it0}(1-\Phi_{it0}))^2 + O_{\mathrm{p}}(T^{-1}).$$

As another example,

$$\mathbb{E}_{\tau_0}[l_{i01}^2(\theta,\alpha_i)l_{i10}] = \frac{1}{T^{3/2}}\sum_{t=1}^T X_{it}\mathbb{E}_{\tau_0}[l_{it01}^2(\theta,\alpha_i)l_{it01}] = \frac{1}{T^{3/2}}\sum_{t=1}^T X_{it}(G_{it}^{(0)})^2 G_{it0}^{(0)}\Phi_{it0}(1-\Phi_{it0})(1-2\Phi_{it0}).$$

Other expressions can be derived in a similar manner. Notice that $\sup_{(\theta,\alpha_i)\in\Theta\times\mathcal{J}} \ell_i(\theta,\alpha_i)\in(0,1)$ for every $i$ and $T$ by the compactness of $\Theta$ and $\mathcal{J}$. In particular, the target likelihood which we are approximating takes values in $(0,1)$. Since unlike in the static logit model higher order derivatives of the likelihood depend on outcome data, the second order corrected likelihood may take values outside $(0,1)$ for certain values of $\theta$ and $\hat{\beta}_i$. We found in our simulation exercise that also restricting the second order corrected likelihood to assume only values in $(0,1)$ improves the numerical properties of the optimization problem for small values of $T$. Therefore, our maximization problem in static probit is

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \ell_i^*(\theta,\beta_i)|_{\beta_i=\hat{\beta}_i} \quad \text{subject to } \ell_i^*(\theta,\beta_i)|_{\beta_i=\hat{\beta}_i}\in(0,1).$$

Alternatively, instead of using the general formulas, one can make use of the known probit likelihood and derive the necessary derivatives of the FOB directly, i.e. we first explicitly compute

$$\mathbb{E}_{\tau_i}[l_{i01}^2(\theta,\alpha_i)] = \frac{1}{T}\sum_{t=1}^T \Phi(X_{it}'\gamma+\phi_i)(1-\Phi(X_{it}'\gamma+\phi_i))G_{it}(X_{it}'\theta+\alpha_i)$$

and

$$\lambda_{i02}(\theta,\alpha_i;\tau_i) = \frac{1}{T}\sum_{t=1}^T(\Phi(X_{it}'\gamma+\phi_i)-\Phi(X_{it}'\theta+\alpha_i))G_{it}^{(1)}(X_{it}'\theta+\alpha_i) - \phi(X_{it}'\theta+\alpha_i)G_{it}(X_{it}'\theta+\alpha_i)$$

before taking the derivatives with respect to $\alpha$, $\gamma$ and $\phi_i$. Writing $G_{it}^{(k)}(\theta) := G_{it}^{(k)}(X_{it}'\theta+\alpha_i^*(\theta))$ for $k\in\mathbb{N}$, we then obtain

(I)

$$\partial_\alpha\mathbb{E}_{\tau_i}[l_{i01}^2(\theta,\alpha_i)]|_{\beta_i=\beta_{i0}} = \frac{1}{T}\sum_{t=1}^T \Phi_{it0}(1-\Phi_{it0})G_{it}^{(1)}(\theta)$$

(II)

$$\partial^2_{\alpha_i}\mathbb{E}_{\tau_i}[l^2_{i01}(\theta,\alpha_i)]|_{\beta_i=\beta_{i0}} = \frac{1}{T}\sum_{t=1}^{T}\Phi_{it0}(1-\Phi_{it0})G^{(2)}_{it}(\theta)$$

(III)

$$\partial_{\phi_i}\lambda_{i02}(\theta,\alpha_i;\tau_i)|_{\beta_i=\beta_{i0}} = \frac{1}{T}\sum_{t=1}^{T}\varphi_{it0}G^{(1)}_{it}(\theta)$$

(IV)

$$\partial^2_{\phi_i}\lambda_{i02}(\theta,\alpha_i;\tau_i)|_{\beta_i=\beta_{i0}} = -\frac{1}{T}\sum_{t=1}^{T}(X'_{it}\theta_0+\alpha_{i0})\varphi_{it0}G^{(1)}_{it}(\theta)$$

(V)

$$\partial_{\phi_i}\mathbb{E}_{\tau_i}[l^2_{i01}(\theta,\alpha_i)]|_{\beta_i=\beta_{i0}} = \frac{1}{T}\sum_{t=1}^{T}(\varphi_{it0}-2\varphi_{it0}\Phi_{it0})G^2_{it}(\theta)$$

(VI)

$$\partial^2_{\phi_i}\mathbb{E}_{\tau_i}[l^2_{i01}(\theta,\alpha_i)]|_{\beta_i=\beta_{i0}} = \frac{1}{T}\sum_{t=1}^{T}[2(X'_{it}\theta_0+\alpha_{i0})\varphi_{it0}\Phi_{it0}-(X'_{it}\theta_0+\alpha_{i0})\varphi_{it0}-2\varphi^2_{it0}]G^2_{it}(\theta)$$

(VII)

$$\partial_{\gamma}\lambda_{i02}(\theta,\alpha_i;\tau_i)|_{\beta_i=\beta_{i0}} = \frac{1}{T}\sum_{t=1}^{T}X_{it}\varphi_{it0}G^{(1)}_{it}(\theta)$$

(VIII)

$$\partial_{\gamma}\mathbb{E}_{\tau_i}[l^2_{i01}(\theta,\alpha_i)]|_{\beta_i=\beta_{i0}} = \frac{1}{T}\sum_{t=1}^{T}X_{it}(\varphi_{it0}-2\varphi_{it0}\Phi_{it0})G^2_{it}(\theta).$$

Using (I)-(VIII) directly yields

$$\mathcal{B}^{(1)}_{i\phi_i}(\theta,\beta_{i0}) \;=\; \frac{\sum_{t=1}^{T}G^2_{it}\Phi_{it0}(1-\Phi_{it0})\sum_{t=1}^{T}\varphi_{it0}G^{(1)}_{it}}{2T(\sum_{t=1}^{T}(\Phi_{it0}-\Phi_{it})G^{(1)}_{it}-\varphi_{it}G_{it})^2} - \frac{\sum_{t=1}^{T}(\varphi_{it0}-2\varphi_{it0}\Phi_{it0})G^2_{it}}{2T\sum_{t=1}^{T}(\Phi_{it0}-\Phi_{it})G^{(1)}_{it}-\varphi_{it}G_{it}}.$$

Similarly,

$$\begin{aligned}\mathcal{B}^{(1)}_{i\phi_i\phi_i}(\theta,\beta_{i0}) = &-\frac{\sum_{t=1}^{T}G^2_{it}\Phi_{it0}(1-\Phi_{it0})[\sum_{t=1}^{T}\varphi_{it0}G^{(1)}_{it}]^2}{(\sum_{t=1}^{T}(\Phi_{it0}-\Phi_{it})G^{(1)}_{it}-\varphi_{it}G_{it})^3} + \frac{\sum_{t=1}^{T}(\varphi_{it0}-2\varphi_{it0}\Phi_{it0})G^2_{it}\sum_{t=1}^{T}\varphi_{it0}G^{(1)}_{it}}{2\sum_{t=1}^{T}(\Phi_{it0}-\Phi_{it})G^{(1)}_{it}-\varphi_{it}G_{it}}\\ &+\frac{\sum_{t=1}^{T}(\varphi_{it0}-2\varphi_{it0}\Phi_{it0})G^2_{it}\sum_{t=1}^{T}(X'_{it}\theta_0+\alpha_{i0})\varphi_{it0}G^{(1)}_{it}}{2(\sum_{t=1}^{T}(\Phi_{it0}-\Phi_{it})G^{(1)}_{it}-\varphi_{it}G_{it})^2}\\ &-\frac{\sum_{t=1}^{T}[2(X'_{it}\theta_0+\alpha_{i0})\varphi_{it0}\Phi_{it0}-(X'_{it}\theta_0+\alpha_{i0})\varphi_{it0}-2\varphi^2_{it0}]G^2_{it}}{2\sum_{t=1}^{T}(\Phi_{it0}-\Phi_{it})G^{(1)}_{it}-\varphi_{it}G_{it}}.\end{aligned}$$

Moreover,

$$\mathcal{B}^{(1)}_{i\gamma}(\beta_{i0}) \quad = \quad \frac{\sum_{t=1}^{T} G_{it}^2 \Phi_{it0}(1-\Phi_{it0}) \sum_{t=1}^{T} X_{it}\varphi_{it0} G_{it}^{(1)}}{2(\sum_{t=1}^{T}(\Phi_{it0}-\Phi_{it})G_{it}^{(1)} - \varphi_{it}G_{it})^2} - \frac{\sum_{t=1}^{T} X_{it}(\varphi_{it0} - 2\varphi_{it0}\Phi_{it0})G_{it}^2}{2\sum_{t=1}^{T}(\Phi_{it0}-\Phi_{it})G_{it}^{(1)} - \varphi_{it}G_{it}}.$$

Finally,

$$\mathcal{B}^{(1)}_{i\alpha_i\phi_i}(\beta_{i0}) = \frac{\sum_{t=1}^{T} G_{it}^2 \Phi_{it0}(1-\Phi_{it0}) \sum_{t=1}^{T}[(\Phi_{it0}-\Phi_{it})G_{it}^{(2)} - 2\varphi_{it}G_{it}^{(1)} + (X_{it}'\theta + \alpha_i)\varphi_{it}G_{it}] \sum_{t=1}^{T} \varphi_{it0} G_{it}^{(1)}}{(\sum_{t=1}^{T}(\Phi_{it0}-\Phi_{it})G_{it}^{(1)} - \varphi_{it}G_{it})^3}.$$