SECOND-ORDER BIAS REDUCTION FOR NONLINEAR PANEL DATA MODELS WITH FIXED EFFECTS BASED ON EXPECTED QUANTITIES

Martin Schumann
Maastricht University

In many nonlinear panel data models with fixed effects maximum likelihood estimators suffer from the incidental parameters problem, which often entails that point estimates are markedly biased. While the recent literature has mostly generated methods that yield a first-order bias reduction relative to maximum likelihood, we derive a first- and second-order bias correction of the profile likelihood based on "expected quantities" which differs from the corresponding correction based on "sample averages" derived in Dhaene and Sun (2021, *Journal of Econometrics* 220, 227–252). While consistency and asymptotic normality of our estimator are derived in a setting where both the number of individuals and the number of time periods grow to infinity, we illustrate in a simulation study that our second-order bias reduction indeed yields an estimator with substantially improved small sample properties relative to its first-order unbiased counterpart, especially when less than 10 time periods are available.

1. INTRODUCTION

Panel data models are popular among economists and other social scientists, since besides observed characteristics, they allow researchers to account for unobserved individual-specific attributes that do not vary over time. If the unobserved time-invariant individual heterogeneity is allowed to be arbitrarily correlated with observed explanatory variables, it is often referred to as "fixed effects." Failure to account for the presence of fixed effects may therefore lead to biased estimates and incorrect inference in econometric applications.

Many panel datasets for microeconomic analysis are "short," i.e., the number of individuals (n) is much larger than the number of time periods (T). Therefore,

693

the asymptotic theory has often been derived under the assumption that $n \to \infty$ while T is kept fixed. In this asymptotic framework, it is well known how to estimate panel data models in which the unobserved heterogeneity enters linearly and additively, without the need of making distributional assumptions (cf., e.g., Chamberlain, 1982, 1984; Arellano, 2003; Wooldridge, 2010). However, stronger parametric distributional assumptions are typically needed whenever the fixed effects enter the model nonlinearly. Yet, even then only few nonlinear models can be satisfactorily handled when the number of time periods is fixed, since the incidental parameters problem (IPP), first noted by Neyman and Scott (1948), often renders the maximum likelihood estimator (MLE) inconsistent.

Alternatively, one may focus on estimation approaches in which both n and T are allowed to grow to infinity. Yet, even in this asymptotic setting, the presence of fixed effects leads to an incorrectly centered asymptotic distribution of the MLE if $n/T \nrightarrow 0$. Since Hahn and Kuersteiner (2002), who proposed a bias correction for linear dynamic models, numerous corrections of the first-order bias (FOB), i.e., the leading bias term of order $O(T^{-1})$, have been proposed to mitigate the IPP (for reviews of the literature, see for instance, Arellano and Hahn, 2007 or Fernández-Val and Weidner, 2018). Simulation studies suggest that for moderately large T, FOB corrected estimators perform better in small samples than uncorrected estimators. Moreover, FOB corrected estimators can sometimes be shown to be asymptotically correctly centered if $n/T^3 \to 0$ (e.g., Hahn and Newey, 2004), so that n can be thought of as being "large" relative to T. Yet, when T is small (i.e., $T < 10$), even first-order corrected estimators exhibit some non-negligible bias (see, for instance, Hahn and Newey, 2004 or Section 5). It is therefore desirable to find methods that besides eliminating the leading FOB term also correct bias terms of higher order to allow for faster growth rates of n relative to T and to further improve the small-sample performance of the resulting estimators.

In this paper, we follow Arellano and Hahn (2016), henceforth AH, and focus on a bias correction of the "profile" or "concentrated" likelihood. Besides the FOB, our correction also removes the second-order bias (SOB) and thus leads to an objective function that is "second-order unbiased," i.e., its expectation coincides with the "target" likelihood up to an error of order $O(T^{-3})$. The maximizer of this objective function is shown to be consistent and asymptotically following a correctly centered normal distribution if $n, T \to \infty$ such that $n/T^5 \to 0$. A simulation study illustrates that our second-order corrected estimator substantially improves upon its first-order counterparts in terms of small sample performance and inference.

We further compare our approach to a closely related profile likelihood correction outlined in Dhaene and Sun (2021), henceforth DS, that also yields a second-order unbiased objective function. However, the approach in DS does not make full use of the known likelihood, since expectations are estimated based on "sample averages" (SA) rather than by "expected quantities" (EQ). We further illustrate that each approach relies on mutually exclusive properties: while DS use "score-factors," which is not possible with our approach based on EQ, the "centering"

based approach presented here is not applicable for second-order bias corrections based on SA. Finally, we also provide some evidence that estimators which make full use of the likelihood being known may be preferable in small samples as compared to estimators based on sample averages, albeit at the expense of higher computational complexity.

The rest of the paper is organized as follows: In Section 2, we introduce some notation and discuss the IPP before introducing our estimator together with an illustrative example. Section 3 discusses differences between corrections based on EQ and SA and highlights the main distinctions between our approach and DS. Assumptions and theory are provided in Section 4. Simulation results on the small sample performance of our estimator are presented in Section 5. Finally, Section 6 concludes. Additional tables as well as all proofs are collected in the Appendixes and the Online Supplementary Material for this paper.

2. SECOND-ORDER BIAS CORRECTION BASED ON EQ

For $n, T \geq 2$, let Y_{it} denote an outcome variable and X_{it} a (column) vector of explanatory variables for $i = 1, \ldots, n$ and $t = 1, \ldots, T$. Here, i refers to an individual whereas t indexes time. While both Y_{it} and X_{it} are assumed to be observed, we assume that the fixed effect α_{i0} is an unobserved random variable with unknown distribution. Moreover, for individual i, let $\mathcal{Y}_i := (Y_{i1}, \ldots, Y_{iT})$ denote the time-series of outcomes, and $\mathcal{X}_i := (X_{i1}, \ldots, X_{iT})$ the time-series of explanatory variables. The time-series of outcomes are drawn from the conditional distribution $f_{\mathcal{Y}_i | \mathcal{X}_i, \theta_0, \alpha_{i0}}$, which is known up to the p-dimensional parameter of interest $\theta_0 \in \Theta$. Note that this does not impose any restriction on the distribution of $(\mathcal{X}_i, \alpha_{i0})$, hence we allow for arbitrary dependence between the observed explanatory variables and the unobserved fixed effect. The average (or scaled) loglikelihood for individual i depends on $(\theta, \alpha_i) \in \Theta \times \mathcal{J}$ and is defined as $\ell_i(\theta, \alpha_i) := T^{-1} \log f_{\mathcal{Y}_i | \mathcal{X}_i, \theta, \alpha_i}(\mathcal{Y}_i)$. By "time-independence" specified in Assumption 4.1, we can further write $\ell_i(\theta, \alpha_i) = T^{-1} \sum_{t=1}^{T} \ell_{it}(\theta, \alpha_i)$, where $\ell_{it}(\theta, \alpha_i) = \log f_{Y_{it} | \mathcal{X}_i, \theta, \alpha_i}(Y_{it})$ is the likelihood contribution of individual i in time period t. For expansions of the loglikelihood, we indicate the order of partial derivatives with respect to the vector θ by the first number in the index of the loglikelihood whereas the order of partial derivatives with respect to the scalar α_i will be registered by the second number in the index. More precisely, for $m \in \{0, 1, \ldots, 9\}$, $\ell_{i0m}(\theta, \alpha_i) = \partial_{\alpha_i}^m \ell_i(\theta, \alpha_i)$ is a scalar, $\ell_{i1m}(\theta, \alpha_i) = \partial_\theta \circ \partial_{\alpha_i}^m \ell_i(\theta, \alpha_i)$ is a column vector and $\ell_{i2m}(\theta, \alpha_i) = \partial_{\theta\theta'} \circ \partial_{\alpha_i}^m \ell_i(\theta, \alpha_i)$ is a $p \times p$-dimensional matrix. Following usual convention, ∂^0 is interpreted as the identity operator. Since we derive a correction of the profile likelihood bias conditional on covariates that is valid globally, i.e., for any $\theta \in \Theta$, we further introduce notation for taking expectations conditional on the covariates.[1] For the SOB correction of the profile likelihood of individual i, let $\tau_i := (\gamma', \phi_i)' \in \Theta \times \mathcal{J}$. We then let the expectation operator $\mathbb{E}_{\tau_i}$ denote integration with respect to $f_{\mathcal{Y}_i | \mathcal{X}_i, \tau_i}$,

---

[1]In the following, "bias" refers to the bias conditional on covariates unless stated otherwise.

which denotes the conditional density evaluated at τ_i for individual i. For Section 4, where we derive the conditional bias of our objective function (which depends on preliminary estimators and thus on the data of the whole sample), let $\tau_0 := (\theta_0', \alpha_{10}, \ldots, \alpha_{n0})' \in \Theta \times \mathcal{J} \times \cdots \times \mathcal{J}$ and let $\mathbb{E}_{\tau_0}$ denote integration with respect to $f_{\mathcal{Y}_1, \ldots, \mathcal{Y}_n | \mathcal{X}_1, \ldots, \mathcal{X}_n, \tau_0}$, which denotes the joint density of outcomes, conditional on all explanatory variables, for the entire sample of individuals. Notice that if the integrand only depends on outcome data of individual i, $\mathbb{E}_{\tau_0}$ reduces to an expectation with respect to the individual density $f_{\mathcal{Y}_i | \mathcal{X}_i, \theta_0, \alpha_{i0}}$ so that in this case, $\mathbb{E}_{\tau_0}$ coincides with $\mathbb{E}_{\tau_i}$ with $\tau_i = (\theta_0', \alpha_{i0})'$.[2] The symbol $\mathbb{E}$ without subscript is reserved for unconditional expectations evaluated at the true parameters. Finally, we introduce a separate notation for centered and scaled likelihood terms, i.e., we write

$$l_{ikm}(\theta, \alpha_i; \tau_i) := \sqrt{T}[\ell_{ikm}(\theta, \alpha_i) - \lambda_{ikm}(\theta, \alpha_i; \tau_i)] = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} l_{itkm}(\theta, \alpha_i; \tau_i),$$

where $\lambda_{ikm}(\theta, \alpha_i; \tau_i) := \mathbb{E}_{\tau_i}[\ell_{ikm}(\theta, \alpha_i)]$ and $l_{itkm}(\theta, \alpha_i; \tau_i) = \ell_{itkm}(\theta, \alpha_i) - \lambda_{itkm}(\theta, \alpha_i; \tau_i)$, with $\lambda_{itkm}(\theta, \alpha_i; \tau_i) = \mathbb{E}_{\tau_i}[\ell_{itkm}(\theta, \alpha_i)]$. By definition, $\mathbb{E}_{\tau_i}[l_{ikm}(\theta, \alpha_i; \tau_i)] = 0$ and, under mild assumptions specified in Section 4.1, $\mathbb{E}_{\tau_i}[l_{ikm}^2(\theta, \alpha_i; \tau_i)] = O_p(1)$, implying that $l_{ikm}(\theta, \alpha_i; \tau_i) = O_p(1)$.[3] In the following, we omit the argument τ_i when $\tau_i = (\theta_0', \alpha_{i0})'$.

2.1. Maximum Likelihood Estimation in the Presence of Fixed Effects

Since we assume that the distribution of $\mathcal{Y}_i | \mathcal{X}_i, \alpha_i$ is known up to θ_0, we can compute the MLE of θ_0 while treating α_i as a "nuisance parameter" to be estimated. Writing $\hat{\alpha}_i(\theta) := \operatorname{argmax}_{\alpha_i \in \mathcal{J}} \ell_i(\theta, \alpha_i)$, and obtaining $\ell_i(\theta, \hat{\alpha}_i(\theta))$ known as the "profile likelihood" for individual i, we can compute the MLE of θ_0 in the presence of $\alpha_1, \ldots, \alpha_n$ as $\tilde{\theta} := \operatorname{argmax}_{\theta \in \Theta} n^{-1} \sum_{i=1}^{n} \ell_i(\theta, \hat{\alpha}_i(\theta))$. Unfortunately, the IPP often renders the MLE inconsistent if T is fixed while $n \to \infty$. As shown in Hahn and Newey (2004), the MLE suffers from FOB, which leads to an incorrectly centered asymptotic distribution if $n, T \to \infty$ such that $n/T \to c$, where c is some positive constant. In order to remedy this problem, one may attempt to correct the profile likelihood in order to approximate the infeasible "target likelihood" $\ell_i(\theta, \alpha_i^*(\theta))$ (which is score-unbiased) based on the "target value" or "population level MLE" $\alpha_i^*(\theta) := \operatorname{argmax}_{\alpha_i \in \mathcal{J}} \mathbb{E}_{\tau_0}[\ell_i(\theta, \alpha_i)]$. As discussed in AH, the profile likelihood correction can be based on EQ (see AH, p. 259), where the FOB is calculated explicitly before replacing the unknown true parameters with plug-in-estimates,

---

[2] Since $\mathbb{E}_{\tau_i}$ and $\mathbb{E}_{\tau_0}$ denote integration with respect to a conditional density which thus results in a random variable, the same $O_p(\cdot)$-notation as for the likelihood is used, even though the respective probability measures may differ. If expectations are taken unconditionally, we use standard $O(\cdot)$-notation.

[3] When expectations and variances are defined conditional on explanatory variables and individual fixed effects, equalities and inequalities involving them hold w.p.1. In order to avoid the proliferation of "w.p.1" qualifiers, we do not state them explicitly for the remainder of the paper.

and based on SA, where the FOB is estimated by replacing population means with sample means.[4] We now introduce our second-order corrected profile likelihood based on EQ before discussing the distinction between this correction and the one of DS, who provide a second-order correction of the profile likelihood based on SA. Notice that we omit the argument for α_i whenever the likelihood is evaluated at $\alpha_i^*(\theta)$ to shorten notation. For example, $\ell_{ikm}(\theta) := \ell_{ikm}(\theta, \alpha_i^*(\theta))$. Moreover, we frequently omit both arguments when the likelihood is evaluated at (θ_0, α_{i0}). For instance, $\ell_{ikm} := \ell_{ikm}(\theta_0, \alpha_{i0})$.

2.2. The Estimator

Recall $\tau_i := (\gamma', \phi_i)' \in \Theta \times \mathcal{J}$. Let $\beta_i := (\tau_i', \alpha_i)'$, $\beta_{i0} := (\tau_{i0}', \alpha_i^*(\theta))'$ with $\tau_{i0} := (\theta_0', \alpha_{i0})'$ and $\hat{\beta}_i := (\tilde{\theta}', \hat{\alpha}_i(\tilde{\theta}), \hat{\alpha}_i(\theta))'$. By expanding the profile likelihood, we find the first- and second-order profile likelihood bias terms $T^{-1}\mathcal{B}_i^{(1)}(\theta, \beta_{i0})$ and $T^{-2}\mathcal{B}_i^{(2)}(\theta, \beta_{i0})$, where

$$\mathcal{B}_i^{(1)}(\theta, \beta_i) := -\frac{\mathbb{E}_{\tau_i}[l_{i01}^2(\theta, \alpha_i; \tau_i)]}{2\lambda_{i02}(\theta, \alpha_i; \tau_i)} \tag{2.1}$$

and

$$\mathcal{B}_i^{(2)}(\theta, \beta_i)$$

$$:= \frac{\sqrt{T}\mathbb{E}_{\tau_i}[l_{i01}^2(\theta, \alpha_i; \tau_i)l_{i02}(\theta, \alpha_i; \tau_i)]}{2\lambda_{i02}^2(\theta, \alpha_i; \tau_i)} - \frac{\sqrt{T}\mathbb{E}_{\tau_i}[l_{i01}^3(\theta, \alpha_i; \tau_i)]\lambda_{i03}(\theta, \alpha_i; \tau_i)}{6\lambda_{i02}^3(\theta, \alpha_i; \tau_i)}$$

$$- \frac{\mathbb{E}_{\tau_i}[l_{i01}^2(\theta, \alpha_i; \tau_i)l_{i02}^2(\theta, \alpha_i; \tau_i)]}{2\lambda_{i02}^3(\theta, \alpha_i; \tau_i)} - \frac{\mathbb{E}_{\tau_i}[l_{i01}^3(\theta, \alpha_i; \tau_i)l_{i03}(\theta, \alpha_i; \tau_i)]}{6\lambda_{i02}^3(\theta, \alpha_i; \tau_i)}$$

$$+ \frac{\mathbb{E}_{\tau_i}[l_{i01}^3(\theta, \alpha_i; \tau_i)l_{i02}(\theta, \alpha_i; \tau_i)]\lambda_{i03}(\theta, \alpha_i; \tau_i)}{2\lambda_{i02}^4(\theta, \alpha_i; \tau_i)} - \frac{\mathbb{E}_{\tau_i}[l_{i01}^4(\theta, \alpha_i; \tau_i)]\lambda_{i03}^2(\theta, \alpha_i; \tau_i)}{8\lambda_{i02}^5(\theta, \alpha_i; \tau_i)}$$

$$+ \frac{\mathbb{E}_{\tau_i}[l_{i01}^4(\theta, \alpha_i; \tau_i)]\lambda_{i04}(\theta, \alpha_i; \tau_i)}{24\lambda_{i02}^4(\theta, \alpha_i; \tau_i)}. \tag{2.2}$$

A naive bias correction might then involve $T^{-1}\mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i}$. However, this would induce a second-order bias due to the slow convergence of $\hat{\beta}_i$. Thus, we use the corrected version $T^{-1}\tilde{\mathcal{B}}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i}$, where

$$\tilde{\mathcal{B}}_i^{(1)}(\theta, \beta_i)$$

$$:= \mathcal{B}_i^{(1)}(\theta, \beta_i) - \mathcal{B}_{i\alpha_i}^{(1)}(\theta, \beta_i)A(\theta, \beta_i) - \frac{1}{2}\mathcal{B}_{i\alpha_i\alpha_i}^{(1)}(\theta, \beta_i)V(\theta, \beta_i) - \mathcal{B}_{i\gamma}^{(1)}(\theta, \beta_i)'\hat{\mathcal{T}}$$

---

[4]AH refer to SA as the "trace-based approach" since they consider models with a vector of fixed effects. Here, we assume that α_{i0} is a scalar.

$$- \mathcal{B}_{i\phi_i}^{(1)}(\theta,\beta_i)A(\theta,\beta_i)|_{\theta=\tilde{\theta}} - \frac{1}{2}\mathcal{B}_{i\phi_i\phi_i}^{(1)}(\theta,\beta_i)V(\theta,\beta_i)|_{\theta=\tilde{\theta}} - \mathcal{B}_{i\alpha_i\phi_i}^{(1)}(\theta,\beta_i)\tilde{V}(\theta,\beta_i).$$

$$\text{(2.3)}$$

In this definition, partial derivatives of $\mathcal{B}_i^{(1)}(\theta,\beta_i)$ with respect to the components of β_i are denoted by the corresponding parameters in the index. In the following, the β_i-argument is further omitted when evaluating at β_{i0}. For instance,

$$\mathcal{B}_{i\alpha_i}^{(1)}(\theta) := \partial_{\alpha_i}\mathcal{B}_i^{(1)}(\theta,\beta_i)|_{\beta_i=\beta_{i0}} = \partial_{\alpha_i}\mathcal{B}_i^{(1)}(\theta,\gamma,\phi_i,\alpha_i)|_{\gamma=\theta_0,\phi_i=\alpha_{i0},\alpha_i=\alpha_i^*(\theta)}.$$

The explicit expressions for all derivatives appearing in (2.3) are collected in Appendix A. Moreover, for the right-hand side of (2.3) we require

$$A(\theta,\beta_i) := -(2T\lambda_{i02}^2(\theta,\alpha_i;\tau_i))^{-1}$$

$$\left(\mathbb{E}_{\tau_i}[l_{i01}^2(\theta,\alpha_i;\tau_i)]\lambda_{i03}(\theta,\alpha_i;\tau_i) - 2\mathbb{E}_{\tau_i}[l_{i01}(\theta,\alpha_i;\tau_i)l_{i02}(\theta,\alpha_i;\tau_i)]\right)$$

and

$$V(\theta,\beta_i) := \frac{\mathbb{E}_{\tau_i}[l_{i01}^2(\theta,\alpha_i;\tau_i)]]}{T\lambda_{i02}^2(\theta,\alpha_i;\tau_i)}, \quad \tilde{V}(\theta,\beta_i) := \frac{\mathbb{E}_{\tau_i}[l_{i01}(\theta,\alpha_i;\tau_i)l_{i01}(\gamma,\phi_i;\tau_i)]}{T\lambda_{i02}(\theta,\alpha_i;\tau_i)\lambda_{i02}(\gamma,\phi_i;\tau_i)}.$$

Finally,

$$\hat{\mathcal{T}} := \left[\frac{1}{n}\sum_{i=1}^n \hat{F}_i(\theta,\beta_i)|_{\theta=\tilde{\theta},\beta_i=\hat{\beta}_i}\right]^{-1} \frac{1}{nT}\sum_{i=1}^n \partial_\theta \mathcal{B}_i^{(1)}(\theta,\beta_i)|_{\theta=\tilde{\theta},\beta_i=\hat{\beta}_i}$$

with $\hat{F}_i(\theta,\beta_i) := -\lambda_{i20}(\theta,\alpha_i;\tau_i) + \lambda_{i02}^{-1}(\theta,\alpha_i;\tau_i)\lambda_{i11}(\theta,\alpha_i;\tau_i)\lambda_{i11}'(\theta,\alpha_i;\tau_i)$. The estimator $\hat{\theta}$ derived here is now obtained as the maximizer of a second-order bias corrected profile likelihood, i.e.,

$$\hat{\theta} := \underset{\theta\in\Theta}{\text{argmax}} \frac{1}{n}\sum_{i=1}^n \ell_i^*(\theta,\beta_i)|_{\beta_i=\hat{\beta}_i}, \qquad \text{(2.4)}$$

where

$$\ell_i^*(\theta,\beta_i) := \ell_i(\theta,\hat{\alpha}_i(\theta)) - T^{-1}\tilde{\mathcal{B}}_i^{(1)}(\theta,\beta_i) - T^{-2}\mathcal{B}_i^{(2)}(\theta,\beta_i) \qquad \text{(2.5)}$$

and where $T^{-1}\tilde{\mathcal{B}}_i^{(1)}(\theta,\beta_i)|_{\beta_i=\hat{\beta}_i}$ and $T^{-2}\mathcal{B}_i^{(2)}(\theta,\beta_i)|_{\beta_i=\hat{\beta}_i}$ are second-order unbiased approximations of the first- and second-order profile likelihood bias terms. The intuition behind the form of the SOB corrected profile likelihood based on EQ is provided in Section 4.2.

2.3. Example

We now show how the second-order bias correction based on EQ can be applied in the heterogeneous means model of Neyman and Scott (1948). In this simple setup, an explicit form of the bias and its correction can be

derived.[5] Let $Y_{it} = \alpha_{i0} + U_{it}$ and assume $U_{i1}, \ldots, U_{iT} | \alpha_{i0} \overset{d}{=} \text{NIID}(0, \sigma_0^2)$, where $\theta := \sigma^2 \in (0, \infty)$. Ignoring constants, the loglikelihood can then be written as $\ell_i(\theta, \alpha_i) = -\frac{1}{2} \log \theta - \sum_{t=1}^{T} (Y_{it} - \alpha_i)^2/(2T\theta)$. Taking derivatives with respect to α_i yields $\ell_{i01}(\theta, \alpha_i) = \sum_{t=1}^{T} (Y_{it} - \alpha_i)/(T\theta)$ and $\ell_{i02}(\theta, \alpha_i) = -1/\theta = \lambda_{i02}(\theta)$, where the second equality holds because $\ell_{i02}(\theta, \alpha_i)$ does not depend on Y_{it} or α_i. Since $\mathbb{E}_{\tau_0}[Y_{it}] - \alpha_i = 0 \Leftrightarrow \alpha_i = \alpha_{i0}$, the target value $\alpha_i^*(\theta)$ coincides with the true value α_{i0} for any θ. Thus, the expression for the FOB of the profile likelihood of individual i in the Neyman–Scott model is $T^{-1} \mathcal{B}_i^{(1)}(\theta, \beta_{i0})$ with $\mathcal{B}_i^{(1)}(\theta, \beta_i) := \gamma/(2\theta)$. Notice that $\mathcal{B}_i^{(1)}(\theta, \beta_i)$ depends on $\beta_i = (\gamma', \phi_i, \alpha_i)'$ only through γ. The first-order corrected estimator based on expected quantities $\check{\theta}_{exp}$ can now be computed as $\check{\theta}_{exp} = \text{argmax}_{\theta \in (0, \infty)} n^{-1} \sum_{i=1}^{n} [\ell_i(\theta, \hat{\alpha}_i(\theta)) - T^{-1} \mathcal{B}_i^{(1)}(\theta, \beta_i)_{\beta_i = \hat{\beta}_i}]$, where $\hat{\alpha}_i(\theta) = \bar{Y}_{i\cdot} := T^{-1} \sum_{t=1}^{T} Y_{it}$ for every $\theta \in (0, \infty)$ and $\tilde{\theta} = (nT)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T} (Y_{it} - \bar{Y}_{i\cdot})^2$. Since $\partial_\theta \ell_i(\theta, \hat{\alpha}_i(\theta)) = -(2\theta)^{-1} + \sum_{t=1}^{T} (Y_{it} - \bar{Y}_{i\cdot})^2/(2T\theta^2)$ while $T^{-1} \partial_\theta \mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i} = -\tilde{\theta}/(2T\theta^2)$, the first-order condition yields $\check{\theta}_{exp} = \tilde{\theta}(1 + T^{-1})$. Clearly, $\mathbb{E}_{\tau_0}[\tilde{\theta}] = \theta_0(1 - T^{-1})$, showing that the bias of the MLE is exactly of order $O(T^{-1})$. In comparison, the bias of the estimator derived from the first-order corrected profile likelihood based on expected quantities is of order $O(T^{-2})$, as $\mathbb{E}_{\tau_0}[\check{\theta}_{exp}] = \theta_0(1 - T^{-1})(1 + T^{-1}) = \theta_0 - \theta_0/T^2$. In order to derive the second-order corrected profile likelihood based on expected quantities, we first notice that $l_{i0m}(\theta) = 0$ for every $m > 1$ and $\lambda_{i0q} = 0$ for every $q > 2$, which by (2.2) implies that the SOB of the profile likelihood is zero. Let $\mathring{\beta}_i := (\mathring{\theta}', \hat{\alpha}_i(\mathring{\theta}), \hat{\alpha}_i(\mathring{\theta}))'$, where $\mathring{\theta}$ is a consistent preliminary estimator of θ_0. Expanding the approximation of the FOB and taking expectations yields

$$\mathbb{E}_{\tau_0}[\mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \mathring{\beta}_i}] = \mathcal{B}_i^{(1)}(\theta, \beta_{i0}) + \partial_\gamma \mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \beta_{i0}} \mathbb{E}_{\tau_0}[\mathring{\theta} - \theta_0],$$

where we have used that derivatives of $\mathcal{B}_i^{(1)}(\theta, \beta_i)$ with respect to ϕ_i and α_i as well as derivatives with respect to $\mathring{\theta}$ of at least second-order are zero. If the preliminary estimator is the MLE, i.e., $\mathring{\theta} = \tilde{\theta}$, then the expectation of the MLE suggests estimating $\mathbb{E}_{\tau_0}[\tilde{\theta} - \theta_0]$ by $-\tilde{\theta}/T$. Further, since $T^{-1} \partial_\gamma \mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i} = (2T\theta)^{-1}$, we can estimate the FOB up to a bias of order $O(T^{-3})$ by $T^{-1} \tilde{\mathcal{B}}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i} = \tilde{\theta}(1 + T^{-1})/(2T\theta)$. Thus, letting $\hat{\theta} := \text{argmax}_{\theta \in (0, \infty)} n^{-1} \sum_{i=1}^{n} [\ell_i(\theta, \hat{\alpha}_i(\theta) - T^{-1} \tilde{\mathcal{B}}_i^{(1)}(\theta, \tilde{\theta})]$, we obtain $\hat{\theta} = \tilde{\theta}(1 + T^{-1} + T^{-2})$ with $\mathbb{E}_{\tau_0}[\hat{\theta}] = \theta_0 - \theta_0/T^3$. Therefore, the bias of the maximizer of the second-order corrected profile likelihood based on expected quantities is indeed of order $O(T^{-3})$. Notice that the estimator using $\tilde{\theta}$ as the preliminary estimator coincides with the SOB corrected estimator based on SA derived in Example 2 of DS, denoted as $\hat{\theta}^{(2)}$. However, if the preliminary estimator of θ_0 is chosen to be the first-order corrected estimator based on expected quantities, i.e., $\mathring{\theta} = \check{\theta}_{exp}$, the bias of the corrected profile likelihood can be further reduced. Since

---

[5] In Appendix L, we further provide a detailed analysis of the static panel logit and probit models.

$\mathbb{E}_{\tau_0}[\check{\theta}_{exp}] = \theta_0(1 - T^{-2})$, the bias of $\check{\theta}_{exp}$ can be estimated by $-\check{\theta}_{exp}/T^2$, so that $T^{-1}\tilde{\mathcal{B}}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \check{\beta}_i^{exp}} = \check{\theta}_{exp}(1 + T^{-2})/(2T\theta)$ with $\check{\beta}_i^{exp} = (\check{\theta}'_{exp}, \hat{\alpha}_i(\check{\theta}_{exp}), \hat{\alpha}_i(\theta))'$. Letting $\hat{\theta}^* := \text{argmax}_{\theta \in (0, \infty)} n^{-1}\sum_{i=1}^{n}[\ell_i(\theta, \hat{\alpha}_i(\theta)) - T^{-1}\tilde{\mathcal{B}}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \check{\beta}_i^{exp}}]$, we see that $\hat{\theta}^*$ is unbiased up to order $O(T^{-5})$, as $\mathbb{E}_{\tau_0}[\hat{\theta}^*] - \theta_0 = -\theta_0/T^5$. Notice that this iteration step is not possible with the SA-based approach of DS. Further iterating this procedure infinitely many times finally yields $\hat{\theta}^{(\infty)} := \tilde{\theta}\sum_{k=0}^{\infty} T^{-k}$, with expected value $\mathbb{E}_{\tau_0}[\hat{\theta}^{(\infty)}] = \theta_0(1 - T^{-1})\sum_{k=0}^{\infty} T^{-k} = \theta_0$, i.e., the bias has now been removed completely.[6]

3. COMPARING BIAS CORRECTIONS BASED ON EQ AND SA

We now discuss differences between profile likelihood bias corrections based on EQ and SA with particular focus on the implications for differences between our approach and DS. As can be seen by comparing b_1 and b_2 in Proposition 1 of DS with $\tilde{\mathcal{B}}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i}$ and $\mathcal{B}_i^{(2)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i}$ given in Section 2.2, the bias correction based on SA of DS is in general different from the bias correction based on EQ as derived here. This can be attributed to the differences in the definition of the bias and the sources of randomness, as discussed in Section 3.1. Potential consequences in terms for the small sample performance are discussed in Section 3.2.

3.1. Conditional Profile Likelihood Bias

A first crucial distinction between our approach and DS is the benchmark likelihood relative to which the profile likelihood bias is defined. Recall that we are defining the target value as $\alpha_i^*(\theta) = \text{argmax}_{\alpha_i \in \mathcal{J}} \mathbb{E}_{\tau_0}[\ell_i(\theta, \alpha_i)]$, which is used, for instance, in Pace and Salvan (2006), Severini (2000, Chap. 4.6), Schumann, Severini, and Tripathi (2021a), or Schumann, Severini, and Tripathi (2021b). Since the likelihood in many models of interest is obtained by making distributional assumptions on the model error conditional on observed covariates and the unobserved fixed effect, the bias of the profile likelihood is defined as $\mathbb{E}_{\tau_0}[\ell_i(\theta, \hat{\alpha}_i(\theta)) - \ell_i(\theta, \alpha_i^*(\theta))]$. Clearly, cancellation of the first- and second-order conditional profile likelihood bias terms implies that also the corresponding unconditional bias terms are canceled. This definition of the bias has important consequences for the bias correction. For instance, it is not possible to use "score-factors," which are fundamental to the bias correction in DS, as $\mathbb{E}_{\tau_0}[\ell_{it01}(\theta, \alpha_i^*(\theta))] \neq 0$ unless $\theta = \theta_0$. To illustrate this, let $Y_{it} = X_{it}'\theta_0 + \alpha_{i0} + U_{it}$ where $U_{i1}, \dots, U_{iT}|X_{i1}, \dots, X_i, \alpha_{i0} \overset{d}{=} \text{NIID}(0, 1)$. Then $\ell_{i01}(\theta, \alpha_i) = T^{-1}\sum_{t=1}^{T}(Y_{it} - X_{it}'\theta - \alpha_i)$ so that $\alpha_i^*(\theta) = T^{-1}\sum_{t=1}^{T} X_{it}'(\theta_0 - \theta) + \alpha_{i0}$. Since $\ell_{it01}(\theta, \alpha_i) = Y_{it} -$

---

[6]Of course, this result may not hold in more complicated models in which the approximation of the FOB depends on plug-in estimators of $\alpha_i^*(\theta)$ and α_{i0}.

$X'_{it}\theta - \alpha_i$, this implies that $\mathbb{E}_{\tau_0}[\ell_{it01}(\theta, \alpha_i^*(\theta))] = X'_{it}(\theta_0 - \theta) - T^{-1}\sum_{t=1}^{T} X'_{it}(\theta_0 - \theta)$ does in general not equal zero unless for instance $\theta = \theta_0$ or if there is no time variation in the observed covariates.[7] Moreover, the "centering" approach used here by which the likelihood terms are centered around their expectations conditional on covariates is not possible with SA, as the centered terms depend on the unknown true parameters. It is thus not surprising that the bias terms in (2.1) and (2.2) differ from those presented in Lemma 4 of DS. The choice between EQ and SA also has consequences for the derivation of a feasible approximation of the FOB that is unbiased up to an error of order $O_p(T^{-3})$. Initial approximations of the FOB based on EQ and SA are $T^{-1}\mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i}$ and $T^{-1}\check{\mathcal{B}}_i^{(1)}(\theta) := -T^{-1}\sum_{t=1}^{T} \ell_{it01}^2(\theta, \hat{\alpha}_i(\theta))/[2T\ell_{i02}(\theta, \hat{\alpha}_i(\theta))]$, respectively. Conditional on covariates, randomness in $\check{\mathcal{B}}_i^{(1)}(\theta)$ stems from the outcome data in the likelihood derivatives and $\hat{\alpha}_i(\theta)$. In $\mathcal{B}_i^{(1)}(\theta, \beta_{i0})$, all outcome data have been integrated out, so that randomness stems from replacing β_{i0} with its plug-in estimator $\hat{\beta}_i$. Not surprisingly, our approximation of the FOB differs from the one of DS. Moreover, as the use of score-factors is not permitted when the profile likelihood bias is considered conditional on covariates, the derivations of DS cannot be used for a correction based on EQ. For instance, the terms A_2 and B_1 in Lemma 4 of DS have "score-factor" and thus, according to DS, have "a zero-mean dominant stochastic term, which can be ignored." However, as $\mathbb{E}_{\tau_0}[\ell_{it01}(\theta)] \neq 0$ in general, the definition of the target value and simple algebra show that for $\theta \neq \theta_0$, $T^{-1}\sum_{t=1}^{T}\sum_{s=1, s\neq t}^{T} \mathbb{E}_{\tau_0}[\ell_{it01}(\theta)]\mathbb{E}_{\tau_0}[\ell_{is01}(\theta)] = -T^{-1}\sum_{t=1}^{T}\mathbb{E}_{\tau_0}[\ell_{it01}(\theta)]^2 < 0$ so that the statement in DS does not hold when expectations are derived conditionally on covariates. Similarly, if $\theta \neq \theta_0$, $\mathbb{E}_{\tau_0}[Z] \neq 0$, where "[Z]" refers to the term of the same notation that appears for instance in the proof of Lemma 5 in DS.

3.2. Small Sample Performance

While both the estimator of DS and the estimator derived here remove the bias of the profile likelihood up to an $O(T^{-3})$ term, the small sample performance may differ due to the theoretical differences between the two approaches. Hahn, Kuersteiner, and Newey (2004) use higher-order expansions of the MLE in a cross-sectional model to show that using sample averages instead of integrals in first-order bias corrections does not affect the higher-order variance. However, while their result suggests that also in panel data the limiting distribution (as $n, T \to \infty$) of FOB and SOB corrections may be equivalent, a theoretical analysis of potential advantages of SA and EQ in small samples is very challenging. For instance, comparisons in specific models are either noninformative (e.g., in the Neyman–Scott model in Section 2.3, the SOB corrections based on SA and EQ coincide), or very difficult, as closed forms for parameter estimates such as $\tilde{\theta}$ are

---

[7]While this simple model does not suffer from the IPP, a similar argument provided in Appendix L.1.1 shows that score-factors also cannot be used in the static logit.

not available. Nevertheless, we can gain some intuition on possible advantages of each approach. First of all, as can be seen by comparing $\mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i}$ and $\check{\mathcal{B}}_i^{(1)}(\theta)$, the estimated first-order profile likelihood bias for individual i depends on T observations of individual i only for SA and on data of the whole sample (through $\hat{\beta}_i$) for EQ, which may explain why the bias correction based on EQ can perform better in small samples, particularly when T is small. Moreover, unlike the estimator based on SA, our estimator can in principle be iterated, which can be used to remove dependence on initial estimators and starting values for numerical optimization.[8] Further notice that if $\hat{\beta}_i = \beta_{i0}$, the EQ FOB approximation coincides exactly with the FOB, whereas the SA approximation $\check{\mathcal{B}}_i^{(1)}(\theta)$ does so only up to an error of order $O_p(T^{-3/2})$ and with a conditional bias of order $O_p(T^{-2})$. As the same arguments holds for the SOB corrections based on EQ and SA, one may expect the EQ based approximations to be closer to the actual bias terms if the parameter estimates are sufficiently close to the true values. Finally, the fact that our bias correction corrects the conditional profile likelihood bias for every $\theta \in \Theta$ whereas DS in general reduce the conditional bias for $\theta = \theta_0$ (score-factors hold at the true parameters only) might also help explaining why the SOB corrections based on EQ and SA perform similarly well in static logit while the EQ correction appears to perform better in static probit. As shown in detail in Appendix L.1, various simplifications arise in the static logit. For instance, $\ell_{i02}(\theta)$ coincides with $\lambda_{i02}(\theta)$, so that "[Z]" in the proof of Lemma 5 in DS, which stems from expanding $\ell_{i02}(\theta)$ in the denominator around $\lambda_{i02}(\theta)$, vanishes. In contrast, no such simplifications arise in the static probit model, so that our approach corrects the conditional bias due to "[Z]" for any $\theta \in \Theta$ whereas the approach presented in DS does so only for $\theta = \theta_0$. Unfortunately, the potential improvement in terms of small sample performance comes at the cost of an increase in algebraic and computational complexity. A further potential disadvantage of EQ bias corrections may be that it relies on the correct specification of the likelihood not only through the likelihood derivatives needed for the bias correction (which is the case also for SA) but also through the integration step when computing expected quantities. While we do not find evidence for larger misspecification bias of EQ estimators relative to SA estimators in a small simulation study (see the discussion on Table B.3 in Section 5), misspecification bias remains a legitimate concern.

4. ASSUMPTIONS AND THEORY

We now present the assumptions before discussing the underlying intuition for the second-order correction presented here. Finally, we provide some asymptotic results for $\hat{\theta}$.

---

[8]In additional simulation results available upon request we found no clear evidence that iterating improves the small sample properties of our estimator.

4.1. Assumptions

We begin the discussion with the fundamental conditions on our model which we will maintain in the rest of the paper.

Assumption 4.1. (i) The true parameter θ_0 lies in the interior of Θ, which is a compact and convex subset of $\mathbb{R}^p$. (ii) For every i, α_{i0} is a random variable that, with probability one, lies in the interior of $\mathcal{J}$, which is a finite and closed interval in $\mathbb{R}$. (iii) The random variables $(\mathcal{Y}_1, \mathcal{X}_1, \alpha_{10}), \ldots, (\mathcal{Y}_n, \mathcal{X}_n, \alpha_{n0})$ are i.i.d. (iv) For each i, conditional on $\mathcal{X}_i$ and α_{i0}, the outcomes $Y_{i1}, \ldots, Y_{iT}$ are independent. (v) For each i, the process $(Y_{it}, X_{it})_{t \in \mathbb{N}}$ is strictly stationary conditional on α_{i0}.

Assumptions (i) and (ii) are standard assumptions in maximum likelihood theory which ensure that the MLEs for θ_0 and α_{i0} are bounded in probability. Moreover, (iii) is standard in microeconometric applications and is, for instance, used in Chamberlain (2010), Alvarez and Arellano (2003) or, more recently, Schumann et al. (2021a) and Schumann et al. (2021b). Notice that it treats the fixed effect as a random variable that is unobserved (which distinguishes it from the observed covariates) and that we impose no restrictions on the correlation between $\mathcal{X}_i$ and α_{i0}. While the conditional independence assumption in (iv), subsequently referred to as "time-independence" or "independence over time," can in principle be relaxed, it is used here to make the proofs which require higher-order expansions more tractable. Finally, (v) rules out time effects and time trends.[9] The next assumption is used to ensure uniqueness of the MLE and the population level MLE for α_{i0}.

Assumption 4.2. For every i, (i) $\hat{\alpha}_i(\theta)$ is the unique maximizer of $\ell_i(\theta, u)$ in u; (ii) $\alpha_i^*(\theta)$ is the unique maximizer of $\mathbb{E}_{\tau_0}[\ell_i(\theta, u)]$ in u; and (iii) $\hat{\alpha}_i(\theta)$ and $\alpha_i^*(\theta)$ lie in the interior of $\mathcal{J}$ with probability one.

In the derivation of our estimator, we will frequently use the following assumption:

Assumption 4.3. (i) Let $\ell_{it}(\theta, \alpha_i)$ denote the likelihood contribution of individual i in time period t when evaluated at (θ, α_i), $v = (v_1, \ldots, v_p)$ be a vector of nonnegative integers with $|v| := \sum_{s=1}^p v_s$ and μ be a nonnegative integer. Define

$$D^{v+\mu} \ell_{it}(\theta, \alpha_i) = \frac{\partial^{|v|+\mu} \ell_{it}(\theta, \alpha_i)}{\partial \theta_1^{v_1} \cdots \partial \theta_p^{v_p} \partial \alpha_i^{\mu}}.$$

There exists a function $M_{it} := M(Y_{it}, X_{it})$ such that $\sup_{(\theta, \alpha_i) \in \Theta \times \mathcal{J}} |D^{v+\mu} \ell_{it}(\theta, \alpha_i)| \leq M_{it}$ and $\mathbb{E}[\sup_{\tau_i \in \Theta \times \mathcal{J}} \mathbb{E}_{\tau_i}[M_{it}^{32}]] = O(1)$ and $\mathbb{E}[\sup_{\tau_i \in \Theta \times \mathcal{J}} \mathbb{E}_{\tau_i}[M_{it}]]^{32} = O(1)$ for all v with $|v| \leq 4$ and $\mu \leq 9$. (ii) Recall $\beta_i = (\tau_i', \alpha_i)' \in \Upsilon = \Theta \times \mathcal{J} \times \mathcal{J}$ and $\beta_{i0} =$

---

[9]Similar assumptions are used in the majority of papers on nonlinear panels with fixed effects. A noteworthy exception is Fernández-Val and Weidner (2016).

$(\theta'_0, \alpha_{i0}, \alpha^*_i(\theta))'$. There exists an open ball $\mathcal{M} \subset \Upsilon$ around β_{i0} and a constant $D_1 > 0$ such that $\inf_i \inf_{(\theta, \beta_i) \in \Theta \times \mathcal{M}} |\mathbb{E}_{\tau_i}[\ell_{i02}(\theta, \alpha_i)]| > D_1$ for all T large enough. (iii) There exists a constant $D_2 > 0$ such that $\inf_i \inf_{(\theta, \alpha_i) \in \Theta \times \mathcal{J}} (-\ell_{i02}(\theta, \alpha_i)) > D_2$ w.p.1. (iv) For every $\eta > 0$, $\inf_i [\mathbb{E}_{\tau_0}[\ell_i(\theta_0, \alpha_{i0})] - \sup_{\theta \in \Theta : \|\theta - \theta_0\| > \eta} \mathbb{E}_{\tau_0}[\ell_i(\theta, \alpha^*_i(\theta))]] > 0$, where $\|\cdot\|$ denotes the euclidean norm. (v) Let $F_i := T\mathbb{E}_{\tau_0}[\partial_\theta \ell_i(\theta) \partial_{\theta'} \ell_i(\theta)]|_{\theta=\theta_0}$. The partial information $F := \lim_{n\to\infty} n^{-1} \sum_{i=1}^n F_i$ exists and is nonsingular.

Assumption 4.3(i) is used to derive stochastic orders of likelihood terms and remainder terms in Taylor expansions when computing the profile likelihood bias of a given individual. Similar assumptions are frequently used in the literature.[10] Notice that here $\mathbb{E}$ on the outside denotes the unconditional expectation. By the Markov inequality, (i) also implies that conditional expectations of likelihood derivatives are bounded in probability, i.e., $\mathbb{E}_{\tau_0}[\sup_{(\theta, \alpha_i) \in \Theta \times \mathcal{J}} |D^{\nu+\mu} \ell_{it}(\theta, \alpha_i)|] = O_p(1)$. Part (ii) is used to control the behavior of individual i's profile likelihood bias expression and its feasible approximation by bounding the denominator away from zero for T large enough. Part (iii) states that the observed information is uniformly bounded away from zero across the parameters. This holds in all our examples given the compactness of Θ and $\mathcal{J}$, which is guaranteed by Assumption 4.1. Assumption (iv) is an identification condition similar to Condition 1 of Hahn and Kuersteiner (2011). Finally, (v) is required to show asymptotic normality of our estimator.

Since our feasible approximation of the bias terms requires a preliminary estimator of θ_0 which will typically be the MLE, we make the following assumption.

Assumption 4.4. As $n, T \to \infty$: (i) $\tilde{\theta} \xrightarrow{p} \theta_0$; (ii) $\tilde{\theta} - \theta_0 = O_p((nT)^{-1/2}) + O_p(T^{-1})$, $\mathbb{E}[\tilde{\theta}] - \theta_0 = O(T^{-1})$; and (iii) $n/T \nrightarrow 0$.

Unlike the previous conditions, this assumption is maintained mostly for convenience, as (i) and (ii) can be shown under Assumptions 4.1–4.3. Since proofs are available for instance in Hahn and Newey (2004) they are omitted here. Part (iii) restricts our analysis to a setup in which the asymptotic distribution of the MLE is not correctly centered (which makes bias corrections necessary). It is used to simplify the stochastic order of remainder terms.[11]

4.2. Derivation of the SOB Correction

We now provide some intuition behind the estimator presented in Section 2.2. First, we use Assumption 4.3 and follow the approach in Severini (2000, Sect. 5.2) to derive the bias of the profile likelihood for each $\theta \in \Theta$. As shown in Appendix C.1,

---

[10] Assumptions similar to (i) and (ii) can be found in Hahn and Kuersteiner (2011) or Hahn and Newey (2004).

[11] For example, using (iii), we can reduce $O_p(\frac{1}{nT^2}) + O_p(T^{-3})$ to $O_p(T^{-3})$.

the bias of the profile likelihood can be expressed as

$$\mathbb{E}_{\tau_0}[\ell_i(\theta, \hat{\alpha}_i(\theta)) - \ell_i(\theta)] = \frac{\mathcal{B}_i^{(1)}(\theta, \beta_{i0})}{T} + \frac{\mathcal{B}_i^{(2)}(\theta, \beta_{i0})}{T^2} + \mathrm{Rem}^{(0)}(\theta), \tag{4.1}$$

where $\sup_{\theta \in \Theta} \mathrm{Rem}^{(0)}(\theta) = O_p(T^{-3})$. Note that $\mathcal{B}_i^{(1)}(\theta, \beta_{i0})$ and $\mathcal{B}_i^{(2)}(\theta, \beta_{i0})$ are of order $O_p(1)$. While $T^{-2}\mathcal{B}_i^{(2)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i}$ is a simple second-order unbiased approximation of the SOB, as

$$T^{-2}\mathbb{E}_{\tau_0}[\mathcal{B}_i^{(2)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i}] = T^{-2}\mathcal{B}_i^{(2)}(\theta, \beta_{i0}) + O_p(T^{-3}), \tag{4.2}$$

using $T^{-1}\mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i}$ as an approximation of the FOB will induce a bias of second order. To refine this approximation, we use a Taylor expansion (adapting the idea outlined in Ferrari et al., 1996) and take expectations (conditional on covariates), which yields

$$\mathcal{B}_i^{(1)}(\theta, \beta_{i0}) = \mathbb{E}_{\tau_0}[\mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i}] - \mathcal{B}_{i\alpha_i}^{(1)}(\theta)\mathbb{E}_{\tau_0}[\hat{\alpha}_i(\theta) - \alpha_i^*(\theta)]$$

$$- \frac{1}{2}\mathcal{B}_{i\alpha_i\alpha_i}^{(1)}(\theta)\mathbb{E}_{\tau_0}[(\hat{\alpha}_i(\theta) - \alpha_i^*(\theta))^2]$$

$$- \mathcal{B}_{i\gamma}^{(1)}(\theta)'\mathbb{E}_{\tau_0}[\tilde{\theta} - \theta_0] - \mathcal{B}_{i\phi_i}^{(1)}(\theta)\mathbb{E}_{\tau_0}[\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0}] - \frac{1}{2}\mathcal{B}_{i\phi_i\phi_i}^{(1)}(\theta)\mathbb{E}_{\tau_0}[(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0})^2]$$

$$- \mathcal{B}_{i\alpha_i\phi_i}^{(1)}(\theta)\mathbb{E}_{\tau_0}[(\hat{\alpha}_i(\theta) - \alpha_i^*(\theta))(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0})] + \mathbb{E}_{\tau_0}[\mathrm{Rem}^{(1)}(\theta)], \tag{4.3}$$

where $\mathrm{Rem}^{(1)}(\theta)$ is specified in (C.10).[12] Notice that we have used that derivatives of $\mathcal{B}_i^{(1)}(\theta, \beta_i)$ do not depend on outcome data if none of the arguments do. As a consequence, they act as constants with respect to the conditional expectation operator $\mathbb{E}_{\tau_0}$.[13] As shown in Appendix I, $T^{-1}\mathbb{E}_{\tau_0}[\mathrm{Rem}^{(1)}(\theta)] = O_p(T^{-3})$ so that the remainder can be ignored. For a feasible approximation, we require approximations for the terms on the right-hand side of (4.3). Derivatives of the FOB can easily be approximated up to an error of sufficiently low order by replacing β_{i0} with $\hat{\beta}_i$, as $\mathbb{E}_{\tau_0}[\partial_{\beta_i}\mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i}] = \partial_{\beta_i}\mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \beta_{i0}} + O_p(T^{-1})$. Next, as shown explicitly in Appendix C.2, an expansion shows that $\mathbb{E}_{\tau_0}[\hat{\mathcal{T}}] = \mathbb{E}_{\tau_0}[\tilde{\theta} - \theta_0] + O_p(T^{-2})$, i.e., $\hat{\mathcal{T}}$ is a first-order unbiased estimator of $\mathbb{E}_{\tau_0}[\tilde{\theta} - \theta_0]$. Similarly,

$$\mathbb{E}_{\tau_0}[A(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i}] = \mathbb{E}_{\tau_0}[\hat{\alpha}_i(\theta) - \alpha_i^*(\theta)] + O_p(T^{-2}) \text{ and}$$

$$\mathbb{E}_{\tau_0}[A(\theta, \beta_i)|_{\theta = \tilde{\theta}, \beta_i = \hat{\beta}_i}] = \mathbb{E}_{\tau_0}[\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0}] + O_p(T^{-2}). \tag{4.4}$$

---

[12]Notice that the expectation computes $\int \mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i} f_{y_1, \dots, y_n | x_1, \dots, x_n, \tau_0}$, where the randomness with respect to the conditional density is coming from $\hat{\beta}_i$ only.

[13]If the FOB is approximated based on SA, derivatives of the FOB approximation depend on outcome data even when evaluated at the true parameters. Thus, they are not constant with respect to $\mathbb{E}_{\tau_0}$. Consequently, this step highlights one of the main conceptual differences between bias corrections based on EQ and SA.

Furthermore,

$$\mathbb{E}_{\tau_0}[V(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}] = \mathbb{E}_{\tau_0}[(\hat{\alpha}_i(\theta) - \alpha_i^*(\theta))^2] + O_p(T^{-2}),$$

$$\mathbb{E}_{\tau_0}[V(\theta, \beta_i)|_{\theta=\tilde{\theta}, \beta_i=\hat{\beta}_i}] = \mathbb{E}_{\tau_0}[(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0})^2] + O_p(T^{-2}), \text{ and}$$

$$\mathbb{E}_{\tau_0}[\tilde{V}(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}] = \mathbb{E}_{\tau_0}[(\hat{\alpha}_i(\theta) - \alpha_i^*(\theta))(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0})] + O_p(T^{-2}). \tag{4.5}$$

Consequently, as is shown in Appendix C.2, $\tilde{\mathcal{B}}_i^{(1)}(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}$ defined in (2.3) satisfies

$$T^{-1}\mathbb{E}_{\tau_0}[\tilde{\mathcal{B}}_i^{(1)}(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}] = T^{-1}\mathcal{B}_i^{(1)}(\theta, \beta_{i0}) + O_p(T^{-3}). \tag{4.6}$$

Therefore, $T^{-1}\tilde{\mathcal{B}}_i^{(1)}(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}$ constitutes a feasible approximation of the true FOB $T^{-1}\mathcal{B}_i^{(1)}(\theta, \beta_{i0})$ that is unbiased up to third order. Finally, combining (4.1), (4.2), and (4.6),

$$\mathbb{E}_{\tau_0}[\ell_i^*(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}]$$

$$= \mathbb{E}_{\tau_0}[\ell_i(\theta, \hat{\alpha}_i(\theta))] - T^{-1}\mathbb{E}_{\tau_0}[\tilde{\mathcal{B}}_i^{(1)}(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}] - T^{-2}\mathbb{E}_{\tau_0}[\mathcal{B}_i^{(2)}(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}]$$

$$= \mathbb{E}_{\tau_0}[\ell_i(\theta, \alpha_i^*(\theta))] + R_i(\theta), \text{ with } R_i(\theta) = O_p(T^{-3}), \tag{4.7}$$

so that the difference between the second-order corrected profile likelihood and the target likelihood exhibits a bias of order $O_p(T^{-3})$.[14] Notice that the remainder term $R_i(\theta)$ is a combination of the remainder terms in (4.1), (4.2), and (4.6). Next, we discuss the asymptotic properties of $\hat{\theta}$ defined in (2.4).

4.3. Asymptotic Theory

We now show asymptotic normality of $\hat{\theta}$. While we provide some intuition on the main steps, we defer the detailed proofs to Appendix D. If not stated otherwise, all limits are taken as $n, T \to \infty$. Consistency for $\hat{\theta}$ can be shown in a similar way as the consistency of the MLE $\tilde{\theta}$, since the correction term to the profile likelihood in our bias corrected objective function becomes negligible in the limit as $n, T \to \infty$.[15] While showing consistency does not require us to impose a rate at which T grows relative to n, we will need rate conditions to show asymptotic normality of $\hat{\theta}$. In the first step, we expand the first-order condition

---

[14] In principle, the ideas used to derive a SOB corrected profile likelihood can also be used to derive SOB corrected estimators of average partial effects (APEs), which in practical applications are often the ultimate object of interest. Since several authors (e.g., Hahn and Newey, 2004; Fernández-Val, 2009) have found a rather small bias of standard maximum likelihood estimators of APEs in static models which are the focus of this paper, we leave the explicit derivation of SOB corrections of APEs for future research.

[15] An explicit proof of the consistency of $\hat{\theta}$ is provided in Appendix K.

$n^{-1} \sum_{i=1}^{n} \ell_i^*(\theta, \beta_i)|_{\theta=\bar{\theta}, \beta_i=\hat{\beta}_i} = 0$ of $\hat{\theta}$ as

$$\sqrt{nT}(\hat{\theta} - \theta_0) = \left[-\frac{1}{n} \sum_{i=1}^{n} \partial_{\theta\theta'} \ell_i^*(\theta, \beta_i)|_{\theta=\bar{\theta}, \beta_i=\hat{\beta}_i} \right]^{-1} \sqrt{\frac{T}{n}} \sum_{i=1}^{n} \partial_\theta \ell_i^*(\theta, \beta_i)|_{\theta=\theta_0, \beta_i=\hat{\beta}_i},$$

$$\textbf{(4.8)}$$

where $\bar{\theta}$ lies between $\hat{\theta}$ and θ_0. Since the target likelihood is a "genuine" likelihood that satisfies the Bartlett identities and the information equality in particular, the first term on the right hand side of (4.8) can be shown to converge to F as defined in Assumption 4.3(v), i.e.,

$$-\frac{1}{n} \sum_{i=1}^{n} \partial_{\theta\theta'} \ell_i^*(\theta, \beta_i)|_{\theta=\bar{\theta}, \beta_i=\hat{\beta}_i} \xrightarrow{p} F \qquad \text{as } n, T \to \infty. \qquad \textbf{(4.9)}$$

Next, we consider the second term on the right-hand side of (4.8). Recall that by (4.7), $\mathbb{E}_{\tau_0}[\ell_i^*(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}] = \mathbb{E}_{\tau_0}[\ell_i(\theta, \alpha_i^*(\theta))] + R_i(\theta)$ with $R_i(\theta) = O_p(T^{-3})$. In order to guarantee that the score with respect to θ of $\ell_i^*(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}$ is unbiased up to an error of order $O(T^{-3})$, we impose the next "high-level" assumption.

Assumption 4.5. $\mathbb{E}[\partial_\theta R_i(\theta)|_{\theta=\theta_0}] = O(T^{-3})$ for every $\theta \in \Theta$.

Assumptions of this kind are frequently used in the literature (e.g., in Hahn and Newey, 2004, and implicitly in Arellano and Bonhomme, 2009). Here, it is used to avoid the explicit derivation of the derivative of $R_i(\theta)$, which is algebraically tedious without offering further insight. However, careful inspections of the proofs of Section 4.2 reveals that $R_i(\theta)$ is of the form $T^{-3} r_i(\theta)$, where $r_i(\theta)$ consist of a sum of fractions. The numerators of these fractions consist of products of likelihood derivatives which have bounded unconditional expectation by Lemma E.1, while the denominators are bounded away from zero by Assumption 4.3. The same pattern arises when taking the derivative with respect to θ, albeit with more complicated algebra. Further notice that the argument in $r_i(\theta)$ is fixed and thus does not drift with n, T. Further justification for Assumption 4.5 is given in Appendix D.

Since differentiation and integration is interchangeable by Assumption 4.3(i), using an expansion of the profile likelihood together with Assumption 4.5 then implies

$$\mathbb{E}_{\tau_0}[\partial_\theta \ell_i^*(\theta, \beta_i)|_{\theta=\theta_0, \beta_i=\hat{\beta}_i}] = \mathbb{E}_{\tau_0}[\partial_\theta \ell_i(\theta, \alpha_i^*(\theta))|_{\theta=\theta_0}] + \partial_\theta R_i(\theta)|_{\theta=\theta_0} = O_p(T^{-3}).$$

It is then shown in Appendix D that

$$\sqrt{\frac{T}{n}} \sum_{i=1}^{n} \partial_\theta \ell_i^*(\theta, \beta_i)|_{\theta=\theta_0, \beta_i=\hat{\beta}_i} = \sqrt{\frac{T}{n}} \sum_{i=1}^{n} \partial_\theta \ell_i(\theta)|_{\theta=\theta_0} + O_p\left(\sqrt{\frac{n}{T^5}}\right) + O_p\left(\frac{1}{\sqrt{T}}\right),$$

$$\textbf{(4.10)}$$

which illustrates that the rate $n/T^5 \to 0$ needs to be imposed in order to ensure that the asymptotic distribution is centered around the true parameter θ_0. We further show in Appendix D that

$$\sqrt{\frac{T}{n}} \sum_{i=1}^{n} \partial_\theta \ell_i(\theta)|_{\theta=\theta_0} \xrightarrow{d} \text{N}(0, F) \qquad \text{as } n, T \to \infty. \tag{4.11}$$

Combining these results then leads to the following theorem.

THEOREM 4.1. *Let Assumptions 4.1–4.5 hold and let $n/T^5 \to 0$ as $n, T \to \infty$. Then,*

$$\sqrt{nT}(\hat{\theta} - \theta_0) \xrightarrow{p} \text{N}(0, F^{-1}) \qquad \text{as } n, T \to \infty.$$

5. SIMULATIONS

In order to assess the potential gains of our SOB correction while allowing for a comparison with the simulation results of DS, we conduct a simulation study using Design 3 of DS.[16] Here, $Y_{it} = \mathbb{1}(X_{it}'\theta_0 + \alpha_{i0} + U_{it} > 0)$, for $i = 1, \ldots, n$ and $t = 1, \ldots, T$ with $\alpha_{i0} \sim \text{N}(0, 1/16)$ and $X_{it} \sim \text{N}(\alpha_{i0}, 1)$. Further, we set $\theta_0 = 1$. Finally, independent of $(\mathcal{X}_i, \alpha_{i0})$, we let $U_{i1}, \ldots, U_{iT} \stackrel{d}{=} \text{LogisticIID}$ in panel logit and $U_{i1}, \ldots, U_{iT} \stackrel{d}{=} \text{NIID}(0, 1)$ in panel probit. In a first set of simulations, we choose $n = 100$ and consider $T \in \{3, 4, 5, 6, 10\}$, since we are particularly interested in the bias reducing properties when T is small relative to n. We further report numbers for $n = 1,000$ in static logit and static probit with $T \in \{5, 10, 20\}$. The code for the experiments is written in R. The results in the first set of simulations are based on $M = 2,000$ Monte Carlo iterations whereas, due to an increase in computational burden, the results in the second set of simulations are based on 250 iterations.[17]

The estimators considered here include the MLE $(\tilde{\theta})$, the first-order corrected estimators based on sample averages $(\check{\theta}_{avg})$ and based on expected quantities $(\check{\theta}_{exp})$ where, as discussed in Section 3.1, the FOB is estimated using $T^{-1}\check{\mathcal{B}}_i^{(1)}(\theta)$ and $T^{-1}\mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}$ respectively, and the second-order corrected estimator based on expected quantities $(\hat{\theta})$. Moreover, we report results for both the first- and second-order split-panel jackknife bias corrected estimators of Dhaene and Jochmans (2015) denoted as $\hat{\theta}_{SPJ}^{(FO)}$ and $\hat{\theta}_{SPJ}^{(SO)}$. For static logit, we also report the conditional maximum likelihood estimator $(\tilde{\theta}_{CMLE})$ as a benchmark.

---

[16] In Appendix F, we further provide simulation results based on a data generating process introduced in Arellano and Bonhomme (2009) and subsequently used in Schumann et al. (2021a). The results are qualitatively similar (see Table 4).

[17] The actual number of Monte Carlo iterations may be lower, as iterations where at least one of the estimators cannot be computed due to the failure of the standard maximization routine `optim()` are discarded. However, these cases are very rare when $T \geq 4$.

We further present coverage rates based on the Wald test, i.e., for a given estimator $\mathring{\theta}$ of θ_0 we compute $\hat{W} := nT(\mathring{\theta} - \theta_0)^2/\hat{F}^{-1}(\mathring{\theta})$, where $\hat{F}(\mathring{\theta}) := n^{-1}\sum_{i=1}^{n} \partial_{\theta\theta'}\ell_i(\theta, \hat{\alpha}_i(\theta))|_{\theta=\mathring{\theta}}$ is a consistent estimator of F as $n, T \to \infty$. For a given nominal coverage level c, we then report the coverage rate as $M^{-1}\sum_{m=1}^{M} \mathbb{1}\{\hat{W}_m < \mathcal{Q}\chi_1^2(c)\}$, where $\hat{W}_m$ denotes the value of the Wald statistic in Monte Carlo iteration m and $\mathcal{Q}\chi_1^2(c)$ denotes the c-quantile of the chi-square distribution with one degree of freedom. In all our simulations, we set $c = 95\%$.

5.1. Results

The results in both models shown in Tables B.1 and B.2 are qualitatively similar. First, they confirm the well-known fact that the MLE exhibits a large bias, especially when T is small. In comparison, the FOB corrected estimators perform much better, in particular when T is moderately large (i.e., $T \geq 5$), both in terms of bias and variance.[18] Moreover, our results suggest that the EQ FOB correction of the profile likelihood works better in small samples than its SA counterpart, since $\check{\theta}_{exp}$ greatly outperforms $\check{\theta}_{avg}$ in terms of bias, standard deviation and MSE for any value of T. The difference in performance is particularly striking in static probit, where the MSE of $\check{\theta}_{avg}$ is more than three times larger than the MSE of $\check{\theta}_{exp}$ for $T \leq 5$. The SOB corrected estimator $\hat{\theta}$ works well, also when measured by the performance of its first-order counterpart $\check{\theta}_{exp}$. In both models, the bias, standard deviation and MSE of $\hat{\theta}$ are substantially lower than the respective numbers for $\check{\theta}_{exp}$. The first- and second-order corrected estimators based on EQ also appear to outperform their split-panel jackknife counterparts introduced in Dhaene and Jochmans (2015). While particularly $\hat{\theta}_{SPJ}^{SO}$ does well in reducing the bias when $n = 1,000$ and $T \geq 10$, it does not exist in short panels ($T \leq 5$) and generally suffers from a comparatively high variance potentially due to the panel splitting.[19] Specific to static logit, $\hat{\theta}$ is only beaten (narrowly for $T \geq 5$) by the fixed-T consistent benchmark $\hat{\theta}_{CMLE}$. Further comparing our logit results with the numbers reported for Design 3 in DS shows that second-order corrections based on EQ and SA perform similarly well for $T \in \{5, 10, 20\}$. This may be explained by the fact that in the static logit example simplifications arise that reduce differences between both bias correction methods (see Section 3.2). In contrast, $\hat{\theta}$ greatly outperforms all competing estimators for any value of T in the static probit design. Moreover, comparing our results in Table B.2 with Design 3 in Table 2 of DS suggests that the EQ SOB correction performs considerably better than the SA second-order correction. While $\hat{\theta}$ has an absolute bias of (less than) 1% for $T \in \{5, 10, 20\}$, the estimator of DS still exhibits a substantially larger bias of 14.2% when $T = 5$. In their Table 4, DS further report biases of 32%, 14.4%, and 7.8% for $T \in \{3, 4, 5\}$

---

[18]The phenomenon that in small samples bias reduction can lead to a reduction in variance is well-documented in the literature (e.g., Hahn and Newey, 2004 or Schumann et al., 2021b).

[19]Nevertheless, it may still be a computationally attractive alternative when the panel is sufficiently long.

for their second-order corrected estimator, which is again considerably larger than the biases reported in Table B.1 for the second-order corrected estimator based on EQ.

A similar ranking also arises for the empirical coverage rates based on the Wald statistic: while the MLE is unreliable even for large values of T, the coverage rates of FOB corrected estimators are much closer to the nominal coverage level, in particular when the bias correction is based on EQ. The Wald confidence intervals based on the SOB corrected estimator $\hat{\theta}$ appears to be much more reliable as compared to the first-order corrected estimators. Especially in static probit where the CMLE does not exist, the Wald confidence interval based on the SOB corrected estimator $\hat{\theta}$ clearly outperforms all other estimators considered here. Notice that for $n = 100$ and $T = 5$, the empirical coverage rate of the Wald statistic based on $\hat{\theta}$ practically coincides with the nominal coverage level in both logit and probit. In comparison, DS still report an over-rejection of 13% of the LR test based their SA second-order corrected likelihood in the static probit model. As in DS, we observe a worsening of the empirical coverage rates when n gets larger.[20] However, while their rejection rate increases rather steeply to 45% for $T = 4$, the empirical coverage rate of the Wald test based on $\hat{\theta}$ remains close to 95%.

Finally, Table B.3, which shows the simulated means of the estimators considered here under misspecification (i.e., the data are generated following a probit model and estimated based on the logit likelihood), illustrates that EQ estimators do not necessarily suffer from a larger misspecification bias than SA estimators. Both the SA and EQ estimates appear to be upward biased but approach the CMLE as T increases, which itself is close to the usual "rule of thumb" value of 1.6 (see Wooldridge, 2015, Chap. 17).

6. CONCLUSION

We have constructed a second-order bias corrected objective function for estimation and inference in nonlinear panel data models with fixed effects based on expected quantities. To do so, we have derived an explicit form of the second-order profile likelihood bias and demonstrated how to construct feasible approximations that make full use of the known likelihood. The maximizer of our adjusted profile likelihood is shown to be asymptotically normal and correctly centered if $n, T \to \infty$ such that $n/T^5 \to 0$. A simulation study suggests excellent small sample properties, in particular in very short panels in which first-order corrected estimators still exhibit a substantial bias. We have further pointed out some fundamental differences between bias corrections based on EQ and SA and provided some intuition that helps explaining why EQ estimators appear to outperform SA estimators.

---

[20]This is in line with Lemma 6.1 in Schumann et al. (2021b).

Appendix A. Derivatives of the First-Order Bias Used to Construct (2.3)

In this section, we display the explicit derivatives of the FOB used in the expansion leading to the refined approximation of the FOB $\tilde{\mathcal{B}}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i}$. The detailed derivations are provided in Appendix C.2 of the Online Supplementary Material.

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

$$\mathcal{B}_{i\gamma}^{(1)}(\theta, \beta_{i0}) = -\frac{\sqrt{T}\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)l_{10}]}{2T\lambda_{i02}(\theta)} + \frac{\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)]\mathbb{E}_{\tau_0}[l_{i02}(\theta)l_{10}]}{2T\lambda_{i02}^2(\theta)},$$

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

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

$$- \frac{\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)]\lambda_{i03}^2(\theta)}{T\lambda_{i02}^3(\theta)} + \frac{\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)]\lambda_{i04}(\theta)}{2T\lambda_{i02}^2(\theta)},$$

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

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

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

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

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

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

Appendix B. Simulation tables

TABLE B.1. Simulation Results for Static Logit (Columns 3–6) and Static Probit (Columns 7–10) with $n = 100$.

T	Estimator	Bias	STD	MSE	Coverage	Bias	STD	MSE	Coverage
3	$\tilde{\theta}$	0.6346	0.3164	0.5028	0.2526	0.7057	0.2772	0.5749	0.0423
	$\check{\theta}_{avg}$	0.2791	0.2441	0.1375	0.7346	0.5945	0.2712	0.4270	0.1106
	$\check{\theta}_{exp}$	0.1780	0.2152	0.0780	0.8653	0.2224	0.1871	0.0844	0.6922
	$\hat{\theta}$	0.1130	0.2271	0.0643	0.8888	0.0320	0.1429	0.0214	0.9434
	$\tilde{\theta}_{CMLE}$	0.0189	0.1804	0.0329	0.9585	–	–	–	–
4	$\tilde{\theta}$	0.4213	0.2302	0.2305	0.3975	0.4979	0.2136	0.2935	0.1161
	$\check{\theta}_{avg}$	0.1612	0.1857	0.0605	0.8376	0.3552	0.2009	0.1665	0.3397
	$\check{\theta}_{exp}$	0.0995	0.1688	0.0384	0.9213	0.1344	0.1470	0.0397	0.7961
	$\hat{\theta}$	0.0419	0.1631	0.0284	0.9474	0.0133	0.1233	0.0154	0.9374
	$\hat{\theta}_{SPJ}^{FO}$	−0.2736	0.3520	0.1987	0.4852	−0.2330	0.2659	0.1249	0.3918
	$\tilde{\theta}_{CMLE}$	0.0112	0.1514	0.0230	0.9635	–	–	–	–
5	$\tilde{\theta}$	0.3028	0.1863	0.1264	0.5348	0.3662	0.1649	0.1613	0.1942
	$\check{\theta}_{avg}$	0.0951	0.1528	0.0324	0.9045	0.2210	0.1491	0.0711	0.5330
	$\check{\theta}_{exp}$	0.0577	0.1444	0.0242	0.9305	0.0864	0.1190	0.0216	0.8694
	$\hat{\theta}$	0.0166	0.1383	0.0194	0.9525	0.0034	0.1038	0.0108	0.9415
	$\hat{\theta}_{SPJ}^{FO}$	−0.2274	0.2180	0.0992	0.5333	−0.2330	0.2659	0.1249	0.3918
	$\tilde{\theta}_{CMLE}$	0.0037	0.1345	0.0181	0.9575	–	–	–	–
6	$\tilde{\theta}$	0.2432	0.1479	0.0810	0.5840	0.2836	0.1300	0.0973	0.2694
	$\check{\theta}_{avg}$	0.0698	0.1243	0.0203	0.9320	0.1475	0.1153	0.0351	0.6904
	$\check{\theta}_{exp}$	0.0436	0.1195	0.0162	0.9460	0.0589	0.0985	0.0132	0.9132
	$\hat{\theta}$	0.0139	0.1154	0.0135	0.9625	0.0005	0.0885	0.0078	0.9513
	$\hat{\theta}_{SPJ}^{FO}$	−0.1590	0.1702	0.0542	0.6400	−0.2229	0.2236	0.0996	0.3823
	$\hat{\theta}_{SPJ}^{SO}$	−0.0694	0.4693	0.2250	0.4110	−0.2953	0.6994	0.5761	0.2203
	$\tilde{\theta}_{CMLE}$	0.0073	0.1141	0.0131	0.9655	–	–	–	–
10	$\tilde{\theta}$	0.1280	0.1048	0.0274	0.7360	0.1415	0.0818	0.0267	0.5315
	$\check{\theta}_{avg}$	0.0237	0.0938	0.0093	0.9535	0.0462	0.0730	0.0075	0.9065
	$\check{\theta}_{exp}$	0.0145	0.0925	0.0088	0.9580	0.0182	0.0692	0.0051	0.9485
	$\hat{\theta}$	0.0039	0.0912	0.0083	0.9560	0.0001	0.0667	0.0045	0.9495
	$\hat{\theta}_{SPJ}^{FO}$	−0.0524	0.0967	0.0121	0.8810	−0.0803	0.0801	0.0129	0.6900
	$\hat{\theta}_{SPJ}^{SO}$	0.0201	0.1713	0.0297	0.7205	0.0208	0.1928	0.0376	0.5365
	$\tilde{\theta}_{CMLE}$	0.0028	0.0910	0.0083	0.9560	–	–	–	–

TABLE B.2. Simulation Results for Static Logit (Columns 3–6) and Static Probit (Columns 7–10) with $n = 1,000$.

T	Estimator	Bias	STD	MSE	Coverage	Bias	STD	MSE	Coverage
5	$\tilde{\theta}$	0.2924	0.0622	0.0893	0.0000	0.3436	0.0489	0.1204	0.0000
	$\check{\theta}_{avg}$	0.0882	0.0513	0.0104	0.4920	0.2002	0.0440	0.0420	0.0000
	$\check{\theta}_{exp}$	0.0513	0.0484	0.0050	0.7720	0.0715	0.0357	0.0064	0.4426
	$\hat{\theta}$	0.0103	0.0462	0.0022	0.9280	−0.0091	0.0312	0.0011	0.9277
	$\hat{\theta}_{SPJ}^{FO}$	−0.1733	0.0638	0.0341	0.0880	−0.0937	0.0616	0.0125	0.2240
	$\tilde{\theta}_{CMLE}$	0.0011	0.0432	0.0019	0.9480	–	–	–	–
10	$\tilde{\theta}$	0.1238	0.0343	0.0165	0.0240	0.1358	0.0259	0.0191	0.0000
	$\check{\theta}_{avg}$	0.0202	0.0305	0.0013	0.8840	0.2002	0.0440	0.0420	0.0000
	$\check{\theta}_{exp}$	0.0111	0.0303	0.0010	0.9320	0.0138	0.0219	0.0007	0.9160
	$\hat{\theta}$	0.0005	0.0298	0.0009	0.9480	−0.0041	0.0211	0.0005	0.9320
	$\hat{\theta}_{SPJ}^{FO}$	−0.0480	0.0315	0.0033	0.5600	−0.0719	0.0258	0.0058	0.1200
	$\hat{\theta}_{SPJ}^{SO}$	0.0093	0.0533	0.0029	0.6840	−0.0353	0.0498	0.0037	0.4720
	$\tilde{\theta}_{CMLE}$	0.0010	0.0295	0.0009	0.9480	–	–	–	–
20	$\tilde{\theta}$	0.0579	0.0202	0.0038	0.1880	0.0607	0.0164	0.0039	0.0120
	$\check{\theta}_{avg}$	0.0053	0.0190	0.0004	0.9560	0.0093	0.0153	0.0003	0.8880
	$\check{\theta}_{exp}$	0.0031	0.0190	0.0004	0.9640	0.0030	0.0151	0.0002	0.9320
	$\hat{\theta}$	0.0006	0.0189	0.0004	0.9640	−0.0005	0.0150	0.0002	0.9520
	$\hat{\theta}_{SPJ}^{FO}$	−0.0097	0.0198	0.0005	0.9080	−0.0164	0.0159	0.0005	0.7680
	$\hat{\theta}_{SPJ}^{SO}$	0.0028	0.0251	0.0006	0.8840	0.0023	0.0213	0.0005	0.8080
	$\tilde{\theta}_{CMLE}$	0.0009	0.0194	0.0004	0.9620	–	–	–	–

TABLE B.3. Means Under Misspecification ($n = 100$).

T	$\tilde{\theta}$	$\check{\theta}_{avg}$	$\check{\theta}_{exp}$	$\hat{\theta}$	$\tilde{\theta}_{CMLE}$	$\hat{\theta}_{SPJ}^{FO}$	$\hat{\theta}_{SPJ}^{SO}$
3	3.0052	2.3117	2.0734	2.0712	1.7612	–	–
4	2.6403	2.1176	1.9472	1.9331	1.7511	–	–
5	2.4015	1.9817	1.8695	1.8387	1.7431	1.3593	–
6	2.2455	1.8935	1.8204	1.7842	1.7346	1.4622	1.6122
10	1.9790	1.7663	1.7466	1.7267	1.7190	1.5625	1.6418

Notes: All estimators are based on the static logit model while the data are generated using the static probit model.

Appendix C. Details and Proofs for Section 4

C.1. The Profile Likelihood Bias Terms

We now derive the first- and second-order profile likelihood bias. Notice that for $(\theta, \alpha_i) \in \Theta \times \mathcal{J}$, $k \in \{0, 1, 2, 3, 4\}$ and $m \in \{1, \ldots, 9\}$, Assumption 4.3(i) implies that

(component-wise) $\lambda_{ikm}(\theta,\alpha_i) = O_p(1)$. Moreover, again for each component, it follows from Assumption 4.3(i) and time-independence that $l_{ikm}(\theta)$ is of order $O_p(1)$, as $\mathbb{E}_{\tau_0}[l_{ikm}^2(\theta)] = T^{-1}\sum_{t=1}^T \mathbb{E}_{\tau_0}[l_{itkm}(\theta)] = O_p(1)$ while $\mathbb{E}_{\tau_0}[l_{ikm}(\theta)] = 0$ by definition.

Proof of (4.1). Let $\delta_i(\theta) = \hat\alpha_i(\theta) - \alpha_i^*(\theta)$. Expanding $\ell_{i01}(\theta,\hat\alpha_i(\theta))$ around $\alpha_i^*(\theta)$ yields

$$0 = \ell_{i01}(\theta,\hat\alpha_i(\theta)) = \ell_{i01}(\theta) + \ell_{i02}(\theta,\bar\alpha(\theta))\delta_i(\theta) \Leftrightarrow \delta_i(\theta) = \frac{\ell_{i01}(\theta)}{-\sqrt{T}\ell_{i02}(\theta,\bar\alpha(\theta))},$$

where $\bar\alpha(\theta)$ lies between $\hat\alpha_i(\theta)$ and $\alpha_i^*(\theta)$. Let $d \in \{1,\ldots,6\}$. By Assumption 4.3(iii),

$$\sup_{\theta\in\Theta} |\delta_i^d(\theta)| = \sup_{\theta\in\Theta}\left|\frac{\ell_{i01}^d(\theta)}{-T^{d/2}\ell_{i02}^d(\theta,\bar\alpha(\theta))}\right| \le \frac{(\sup_{(\theta,\alpha_i)\in\Theta\times\mathcal{J}} |\ell_{i01}(\theta,\alpha_i)|)^d}{T^{d/2}D_2^d} = O_p(T^{-d/2}).$$

Thus, by the Markov inequality and Assumption 4.3(i),

$$\sup_{\theta\in\Theta} |\delta_i^d(\theta)| = O_p(T^{-d/2}). \tag{C.1}$$

Again expanding $\ell_{i01}(\theta,\hat\alpha_i(\theta))$ around $\alpha_i^*(\theta)$ and centering terms (i.e., writing $\ell_{i0k}(\theta) = l_{i0k}(\theta)/\sqrt{T} + \lambda_{i0k}(\theta)$) yields

$$0 = \ell_{i01}(\theta,\hat\alpha_i(\theta)) = l_{i01}(\theta) + \sum_{k=1}^5 \frac{1}{k!}\frac{l_{i0k+1}(\theta)}{\sqrt{T}}\delta_i^k(\theta)$$

$$+ \sum_{k=1}^5 \frac{1}{k!}\lambda_{i0k+1}(\theta)\delta_i^k(\theta) + \ell_{i07}(\theta,\bar\alpha(\theta))\delta_i^6(\theta). \tag{C.2}$$

Here, we have further used that $\lambda_{i01}(\theta) = 0$ by the first-order condition of the target value $\alpha_i^*(\theta)$. Writing $\tilde\delta(\theta) := \sqrt{T}\delta_i(\theta)$, which is of order $O_p(1)$ by (C.1), and rearranging this equation,

$$\delta_i(\theta) = -\lambda_{i02}^{-1}(\theta)\left(\sum_{k=0}^5 \frac{1}{k!}\frac{l_{i0k+1}(\theta)}{T^{(k+1)/2}}\tilde\delta^k(\theta) + \sum_{k=1}^5 \frac{1}{k!}\frac{\lambda_{i0k+1}(\theta)}{T^{k/2}}\tilde\delta^k(\theta)\right)$$

$$- \lambda_{i02}^{-1}(\theta)\ell_{i07}(\theta,\bar\alpha(\theta))\delta_i^6(\theta). \tag{C.3}$$

By Assumption 4.3(i) and the Markov inequality, $\sup_{\theta\in\Theta} |\ell_{i07}(\theta,\bar\alpha(\theta))| = O_p(1)$. Moreover, Assumption 4.3(ii) implies that there exists some constant $D_1 > 0$ such that $\inf_{\theta\in\Theta} |\lambda_{i02}^{-1}(\theta)| > D_1$ for T large enough. Together with (C.1), this shows that the supremum (over Θ) of the remainder term $\lambda_{i02}^{-1}(\theta)\ell_{i07}(\theta,\bar\alpha(\theta))\delta_i^6(\theta)$ in (C.3) is of order $O_p(T^{-3})$. As in Rilstone, Srivastava, and Ullah (1996), we notice that (C.3) is of the form

$$\delta_i(\theta) = a_{-1/2} + a_{-1} + a_{-3/2} + a_{-2} + a_{-5/2} + O_p(T^{-3}),$$

where $a_{-d/2}$ collects all terms of order $O_p(T^{-d/2})$, $d = 1,\ldots,5$. Next, we plug this expression for $\delta_i(\theta)$ into (C.2) and collect terms of the same order. Since the left-hand side of (C.2) is zero, we can then iteratively solve for $a_{-d/2}$. For instance, $T^{-1/2}(l_{i01}(\theta) + a_{-1/2}\lambda_{i02}(\theta)) = 0$ yields

$$a_{-1/2} = a_i(\theta,\beta_{i0}) := -\frac{l_{i01}(\theta)}{\lambda_{i02}(\theta)}. \tag{C.4}$$

Using this expression, we can then obtain

$$a_{-1} = -\lambda_{i02}^{-1}(\theta) l_{i02}(\theta) a_{-1/2} - \frac{1}{2}\lambda_{i02}^{-1}\lambda_{i03} a_{-1/2}^2,$$

which implies that

$$a_{-1} = -\frac{l_{i01}^2(\theta)\lambda_{i03}(\theta)}{2\lambda_{i02}^3(\theta)} + \frac{l_{i01}(\theta)l_{i02}(\theta)}{\lambda_{i02}^2(\theta)}. \tag{C.5}$$

Similarly,

$$a_{-3/2} = -\lambda_{i02}^{-1}(\theta) l_{i02}(\theta) a_{-1} - \frac{1}{2}\lambda_{i02}^{-1}(\theta)\lambda_{i03}(\theta)(a_{-1/2}^2 + a_{-1/2}a_{-1})$$
$$- \frac{1}{2}\lambda_{i02}^{-1}(\theta)l_{i03}(\theta) a_{-1/2}^2 - \frac{1}{6}\lambda_{i02}^{-1}(\theta)\lambda_{i04}(\theta) a_{-1/2}^3,$$

whereas

$$a_{-2} = -\lambda_{i02}^{-1}(\theta) l_{i02}(\theta) a_{-3/2} - \frac{1}{2}\lambda_{i02}^{-1}(\theta)\lambda_{i03}(\theta) a_{-1}^2 - \frac{1}{2}\lambda_{i02}^{-1}(\theta)l_{i03}(\theta) a_{-1/2}a_{-1}$$
$$- \frac{1}{6}\lambda_{i02}^{-1}(\theta)\lambda_{i04}(\theta) a_{-1/2}^2 a_{-1} - \frac{1}{6}\lambda_{i02}^{-1}(\theta)l_{i04}(\theta) a_{-1/2}^3$$

and

$$a_{-5/2} = -\lambda_{i02}^{-1}(\theta) l_{i02}(\theta) a_{-2} - \frac{1}{2}\lambda_{i02}^{-1}(\theta)\lambda_{i03}(\theta) a_{-1/2}a_{-2}$$
$$- \frac{1}{2}\lambda_{i02}^{-1}(\theta)l_{i03}(\theta) (a_{-1}^2 + a_{-1/2}a_{-3/2})$$
$$- \frac{1}{6}\lambda_{i02}^{-1}(\theta)\lambda_{i04}(\theta) a_{-1/2}a_{-1}^2 - \frac{1}{6}\lambda_{i02}^{-1}(\theta)l_{i04}(\theta) a_{-1/2}^2 a_{-1}$$
$$- \frac{1}{24}\lambda_{i02}^{-1}(\theta)\lambda_{i05}(\theta) a_{-1/2}^3 a_{-1} - \frac{1}{24}\lambda_{i02}^{-1}(\theta)l_{i04}(\theta) a_{-1/2}^4.$$

Explicit expressions for $a_{-3/2}, \ldots, a_{-5/2}$ are provided in Appendix H. Note that a_{-2} and $a_{-5/2}$ are not provided in Rilstone et al. (1996), so that we have extended their result for the maximum likelihood case. Moreover, these terms differ from those in Ferrari et al. (1996) or Shenton and Bowman (1977) because we cannot use Bartlett-identities (in α_i) to simplify moments of $l_{i01}(\theta)$. For example, in general $\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)] \neq -\lambda_{i02}(\theta)$. Next, expanding the profile likelihood,

$$\ell_i(\theta, \hat{\alpha}_i(\theta)) - \ell_i(\theta, \alpha_i^*(\theta)) = \sum_{k=1}^4 \frac{1}{k!} \frac{l_{i0k}(\theta)}{\sqrt{T}} \delta_i^k(\theta) + \sum_{m=2}^5 \frac{1}{m!} \lambda_{i0m}(\theta)\delta_i^m(\theta) + O_p(T^{-3}). \tag{C.6}$$

Since now explicit expressions for (powers of) $\delta_i(\theta)$ are available, we get an expansion of the profile likelihood around the target likelihood up to the desired order as

$$\ell_i(\theta, \hat{\alpha}_i(\theta)) - \ell_i(\theta, \alpha_i^*(\theta)) = \frac{B_i^{(1)}(\theta)}{T} + \frac{B_i^{(2,1)}(\theta)}{T^{3/2}} + \frac{B_i^{(2,2)}(\theta)}{T^2} + \frac{B_i^{(3)}(\theta)}{T^{5/2}} + O_p(T^{-3}), \tag{C.7}$$

where the explicit terms for $B_i^{(1)}(\theta), \ldots B_i^{(3)}(\theta)$ can be derived as follows:

(1)

$$B_i^{(1)}(\theta) := -\frac{l_{i01}^2(\theta)}{2\lambda_{i02}(\theta)},$$

(2)

$$B_i^{(2,1)}(\theta) := \frac{l_{i01}^2(\theta)l_{i02}(\theta)}{2\lambda_{i02}^2(\theta)} - \frac{l_{i01}^3(\theta)\lambda_{i03}(\theta)}{6\lambda_{i02}^3(\theta)},$$

(3)

$$B_i^{(2,2)}(\theta) := -\frac{l_{i01}^2(\theta)l_{i02}^2(\theta)}{2\lambda_{i02}^3(\theta)} - \frac{l_{i01}^3(\theta)l_{i03}(\theta)}{6\lambda_{i02}^3(\theta)} + \frac{l_{i01}^3(\theta)l_{i02}(\theta)\lambda_{i03}(\theta)}{2\lambda_{i02}^4(\theta)}$$
$$- \frac{l_{i01}^4(\theta)\lambda_{i03}^2(\theta)}{8\lambda_{i02}^5(\theta)} + \frac{l_{i01}^4(\theta)\lambda_{i04}(\theta)}{24\lambda_{i02}^4(\theta)},$$

(4) and

$$B_i^{(3)}(\theta) := \frac{l_{i01}^2(\theta)l_{i02}^3(\theta)}{2\lambda_{i02}^4(\theta)} + \frac{l_{i01}^3(\theta)l_{i02}(\theta)l_{i03}(\theta)}{2\lambda_{i02}^4(\theta)} + \frac{l_{i01}^4(\theta)l_{i04}(\theta)}{24\lambda_{i02}^4(\theta)} - \frac{l_{i01}^3(\theta)l_{i02}^2(\theta)\lambda_{i03}(\theta)}{\lambda_{i02}^5(\theta)}$$
$$- \frac{l_{i01}^4(\theta)l_{i03}(\theta)\lambda_{i03}(\theta)}{4\lambda_{i02}^5(\theta)} + \frac{5l_{i01}^4(\theta)l_{i02}(\theta)\lambda_{i03}^2(\theta)}{8\lambda_{i02}^6(\theta)} - \frac{l_{i01}^5(\theta)\lambda_{i03}^3(\theta)}{8\lambda_{i02}^7(\theta)}$$
$$- \frac{l_{i01}^4(\theta)l_{i02}(\theta)\lambda_{i04}(\theta)}{6\lambda_{i02}^5(\theta)} + \frac{l_{i01}^5(\theta)\lambda_{i03}(\theta)\lambda_{i04}(\theta)}{12\lambda_{i02}^6(\theta)} - \frac{l_{i01}^5(\theta)\lambda_{i05}(\theta)}{120\lambda_{i02}^5(\theta)}.$$

By (E.3) we have $\sup_{\theta\in\Theta}\mathbb{E}_{\tau_0}[B_i^{(3)}(\theta)] = T^{-1/2}O_p(1)$. Thus, by taking expectations conditional on covariates $\mathbb{E}_{\tau_0}$, we arrive at (4.1). Finally, $\text{Rem}^{(0)}(\theta)$ is a combination of $T^{-3}\sqrt{T}B_i^{(3)}(\theta)$ with $\mathbb{E}_{\tau_0}[\sqrt{T}B_i^{(3)}(\theta)] = O_p(1)$ by Lemma E.1, products of likelihood derivatives with $a_{-1/2}, \ldots, a_{-5/2}$ and

$$\frac{1}{5!}\frac{l_{i0k}(\theta, \bar{\alpha}(\theta))}{\sqrt{T}}\delta_i^5(\theta) + \frac{1}{6!}\lambda_{i06}(\theta)\delta_i^6(\theta),$$

which is the remainder in (C.6). Thus, $\sup_{\theta\in\Theta}\text{Rem}^{(0)}(\theta) = O_p(T^{-3})$ follows from Assumption 4.3 and Lemma E.1. □

Remark C.1. Taking the first and second derivatives with respect to θ of the remainder term $\text{Rem}^{(0)}(\theta)$ does not affect its stochastic order. Given the form of the remainder term specified in the proof of (4.1), the first and second derivatives of $\sqrt{T}\mathbb{E}_{\tau_0}[B_i^{(3)}(\theta)]$ with respect to θ consist of products of 8 and 9 likelihood derivatives with bounded expectation by Lemma E.1. Moreover, the first and second derivatives with respect to θ of $\delta_i(\theta)$ can be shown to be of order $O_p(T^{-1/2})$ by following the argument leading to (C.1), as all derivatives of centered likelihood derivatives covered by Assumption 4.3 are of order $O_p(T^{-1/2})$.

Remark C.2. While the stochastic order of $\delta_i(\theta)$ is given by (C.1), expectations of odd powers of $\delta_i(\theta)$ converge at a faster rate. To see this, we again use $\delta_i(\theta) = \ell_{i01}(\theta)/(-\sqrt{T}\ell_{i02}(\theta, \bar{\alpha}(\theta)))$, where the denominator is bounded away from zero by Assumption 4.3(iii). Thus,

$$\mathbb{E}_{\tau_0}[\sup_{\theta \in \Theta} \delta_i^d(\theta)] \leq cT^{-d/2}\mathbb{E}_{\tau_0}[\sup_{\theta \in \Theta} l_{i01}^d(\theta)] = \begin{cases} O_{\mathrm{p}}(T^{-d/2}), & d \text{ even,} \\ O_{\mathrm{p}}(T^{-(d+1)/2}), & d \text{ odd.} \end{cases} \tag{C.8}$$

Here, c denotes a constant and the last equation follows from Lemma E.1.

C.2. Feasible Bias Estimation

In an initial step, we show that

$$T^{-1}\mathbb{E}_{\tau_0}[\mathrm{Rem}^{(1)}(\theta)] = O_{\mathrm{p}}(T^{-3}), \tag{C.9}$$

where $\mathrm{Rem}^{(1)}(\theta)$ is the remainder term in (4.3). For the proof, we introduce some notation for partial derivatives. For $k, k_1, \ldots, k_4 \in \{1, \ldots, p+2\}$, let derivatives of $\mathcal{B}_i^{(1)}(\theta, \beta_i)$ with respect to components of β_i be denoted by corresponding indices, i.e., $\mathcal{B}_{i;k}^{(1)}(\theta, \beta_i) := \partial_{\beta_{ik}}\mathcal{B}_i^{(1)}(\theta, \beta_i), \mathcal{B}_{i;k_1,k_2}^{(1)}(\theta, \beta_i) := \partial_{\beta_{ik_1}\beta_{ik_2}}^2 \mathcal{B}_i^{(1)}(\theta, \beta_i)$ and so on. Further, we let $(\hat{\beta}_i - \beta_{i0})_k$, $k \in \{1, \ldots, p+2\}$ denote the kth components of $(\hat{\beta}_i - \beta_{i0})$.

Proof of (C.9). We can, ignoring constants, write the remainder term as

$$\mathrm{Rem}^{(1)}(\theta) = \mathrm{Rem}^{(1,a)}(\theta) + \mathrm{Rem}^{(1,b)}(\theta), \tag{C.10}$$

where

$$\mathrm{Rem}^{(1,a)}(\theta) := \sum_{k_1=1}^{p+2}\sum_{k_2=1}^{p+2}\sum_{k_3=1}^{p+2} \mathcal{B}_{i;k_1,k_2,k_3}^{(1)}(\theta, \beta_i)|_{\beta_i=\beta_{i0}}(\hat{\beta}_i - \beta_{i0})_{k_1}(\hat{\beta}_i - \beta_{i0})_{k_2}(\hat{\beta}_i - \beta_{i0})_{k_3}$$

and

$$\mathrm{Rem}^{(1,b)}(\theta) := \sum_{k_1=1}^{p+2}\sum_{k_2=1}^{p+2}\sum_{k_3=1}^{p+2}\sum_{k_4=1}^{p+2}$$

$$\mathcal{B}_{i;k_1,k_2,k_3,k_4}^{(1)}(\theta, \beta_i)|_{\beta_i=\bar{\beta}_i}(\hat{\beta}_i - \beta_{i0})_{k_1}(\hat{\beta}_i - \beta_{i0})_{k_2}(\hat{\beta}_i - \beta_{i0})_{k_3}(\hat{\beta}_i - \beta_{i0})_{k_4}$$

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

and $\bar{\beta}_i$ lies between $\hat{\beta}_i$ and β_{i0}. By (C.1), (I.4), and Assumption 4.4(ii) and (iii),

$$(\hat{\beta}_i - \beta_{i0})_k = \begin{cases} O_{\mathrm{p}}(T^{-1/2}), & k \in \{1, p+2\}, \\ O_{\mathrm{p}}(T^{-1}), & k \in \{2, \ldots, p+1\}. \end{cases} \tag{C.11}$$

Moreover, as $\hat{\beta}_i \overset{p}{\to} \beta_{i0}$ as $T \to \infty$, the second- and fourth-order derivatives of $\mathcal{B}_i^{(1)}(\theta, \beta_i)$ are of order $O_{\mathrm{p}}(1)$ when evaluated at $\beta_i = \bar{\beta}_i$ by Lemma E.2. Thus, $\mathrm{Rem}^{(1,b)}(\theta) = O_{\mathrm{p}}(T^{-2})$ by Lemma E.1. Consequently, $T^{-1}\mathbb{E}_{\tau_0}[\mathrm{Rem}^{(1,b)}(\theta)] = O_{\mathrm{p}}(T^{-3})$. Further notice that the same

argument also holds for derivatives of $\mathrm{Rem}^{(1,b)}(\theta)$ with respect to θ, so that the stochastic order is not affected by taking derivatives. Next, we notice that the derivatives of $\mathcal{B}_i^{(1)}(\theta, \beta_{i0})$ do not depend on outcome data, so that

$$\mathbb{E}_{\tau_0}[\mathcal{B}_{i;k_1,k_2,k_3}^{(1)}(\theta, \beta_i)|_{\beta_i=\beta_{i0}}(\hat{\beta}_i - \beta_{i0})_{k_1}(\hat{\beta}_i - \beta_{i0})_{k_2}(\hat{\beta}_i - \beta_{i0})_{k_3}]$$
$$= \mathcal{B}_{i;k_1,k_2,k_3}^{(1)}(\theta, \beta_i)|_{\beta_i=\beta_{i0}}\mathbb{E}_{\tau_0}[(\hat{\beta}_i - \beta_{i0})_{k_1}(\hat{\beta}_i - \beta_{i0})_{k_2}(\hat{\beta}_i - \beta_{i0})_{k_3}].$$

Moreover, (C.11) shows that $(\hat{\beta}_i - \beta_{i0})_{k_1}(\hat{\beta}_i - \beta_{i0})_{k_2}(\hat{\beta}_i - \beta_{i0})_{k_3}$ is of order $O_\mathrm{p}(T^{-2})$ or lower if at least one index in $\{k_1, k_2, k_3\}$ is an element of $\{1, \ldots, p\}$. In this case, Lemma E.1 implies that the expectation of $(\hat{\beta}_i - \beta_{i0})_{k_1}(\hat{\beta}_i - \beta_{i0})_{k_2}(\hat{\beta}_i - \beta_{i0})_{k_3}$ is of sufficiently low order. We thus need to consider

$$\mathbb{E}_{\tau_0}[(\hat{\beta}_i - \beta_{i0})_{k_1}(\hat{\beta}_i - \beta_{i0})_{k_2}(\hat{\beta}_i - \beta_{i0})_{k_3}] \text{ for } k_1, k_2, k_3 \in \{p+1, p+2\}.$$

We first observe that $\mathbb{E}_{\tau_0}[\delta_i^3(\theta)] = O_\mathrm{p}(T^{-2})$ by (C.8). Moreover, $\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0} = \delta_i(\tilde{\theta}) + O_\mathrm{p}(T^{-1})$ by (I.4). Thus,

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

where we have again used (C.8). Using similar arguments, $\mathbb{E}_{\tau_0}[\delta_i(\theta)(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0})^2] = O_\mathrm{p}(T^{-2})$ and $\mathbb{E}_{\tau_0}[(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0})^3] = O_\mathrm{p}(T^{-2})$, as $\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0}$ can be approximated with $\delta_i(\theta_0)$ up to a bias of sufficiently low order (see the proof of (4.4)). In total, $T^{-1}\mathrm{Rem}^{(1,a)}(\theta) = O_\mathrm{p}(T^{-3})$. □

Remark C.3. In the same way, one can show that derivatives of $\mathcal{B}_i^{(1)}(\theta, \beta_i)$ with respect to β_i can be approximated by replacing β_{i0} with β_{i0}, i.e., $\partial_{\beta_i}\mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i} = \partial_{\beta_i}\mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i=\beta_{i0}} + \mathrm{Rem}^{(2)}(\theta)$ with

$$\mathrm{Rem}^{(2)}(\theta) = O_\mathrm{p}(T^{-1/2}) \text{ and } \mathbb{E}_{\tau_0}[\mathrm{Rem}^{(2)}(\theta)] = O_\mathrm{p}(T^{-1}). \tag{C.12}$$

Moreover, also the proof of (4.2) can be carried out with analogous arguments. The full proofs are given in Appendix I.

In the proofs of (4.4) and (4.5), we again denote partial derivatives by parameters in the index and use a plug-in notation so that for instance $A_{\alpha_i}(\theta_0, \beta_{i0}) := \partial_{\alpha_i}A(\theta, \beta_i)|_{\theta=\theta_0, \beta_i=\beta_{i0}}$.

Proof of (4.4). In an initial step, the full proof of (4.4) requires us to show the first equation in

$$\mathbb{E}_{\tau_0}[\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0}] = \mathbb{E}_{\tau_0}[\delta(\theta_0)] + O_\mathrm{p}(T^{-2}),$$
$$\mathbb{E}_{\tau_0}[\delta_{iT}(\theta)(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0})] = \mathbb{E}_{\tau_0}[\delta_{iT}(\theta)\delta_{iT}(\theta_0)] + O_\mathrm{p}(T^{-2}), \quad \text{and}$$
$$\mathbb{E}_{\tau_0}[(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0})^2] = \mathbb{E}_{\tau_0}[\delta_{iT}^2(\theta_0)] + O_\mathrm{p}(T^{-2}), \tag{C.13}$$

which is shown in Appendix I. Note first that by (C.5) and Lemma E.1,

$$\mathbb{E}_{\tau_0}[\delta_i(\theta)] = -\frac{\mathbb{E}_{\tau_0}[l_{i01}^2(\theta)]\lambda_{i03}(\theta)}{2T\lambda_{i02}^3(\theta)} + \frac{\mathbb{E}_{\tau_0}[l_{i01}(\theta)l_{i02}(\theta)]}{T\lambda_{i02}^2(\theta)} + O_\mathrm{p}(T^{-2}) = O_\mathrm{p}(T^{-1}). \tag{C.14}$$

Therefore, $A(\theta, \beta_{i0}) = \mathbb{E}_{\tau_0}[\hat{\alpha}_i(\theta) - \alpha_i^*(\theta)] + O_p(T^{-2})$. We first use an expansion (ignoring constants) to see

$$
A(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i} = A(\theta, \beta_{i0}) + A_{\alpha_i}(\theta, \beta_{i0})\delta_i(\theta) + A_\gamma(\theta, \beta_{i0})'(\tilde{\theta} - \theta_0) + A_{\phi_i}(\theta, \beta_{i0})(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0})
$$

$$
+ A_{\alpha_i\alpha_i}(\theta, \beta_i)|_{\beta_i=\bar{\beta}_i}\delta_i^2(\theta) + A_{\phi_i\phi_i}(\theta, \beta_i)|_{\beta_i=\bar{\beta}_i}(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0})^2
$$

$$
+ A_{\alpha_i\gamma}(\theta, \beta_i)'|_{\beta_i=\bar{\beta}_i}(\tilde{\theta} - \theta_0)\delta_i(\theta) + A_{\phi_i\gamma}(\theta, \beta_i)'|_{\beta_i=\bar{\beta}_i}(\tilde{\theta} - \theta_0)(\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0}),
$$

<div align="right">(C.15)</div>

where $\bar{\beta}_i$ again lies between $\hat{\beta}_i$ and β_{i0}. In order to show (4.4), we need to demonstrate that derivatives of $A(\theta, \beta_i)$ are of order $O_p(T^{-1})$. To do so, one can follow the steps in the proof of Lemma E.2. First, notice that for each of the derivatives the denominator consists of (powers of) $\mathbb{E}[\ell_{i02}(\theta, \alpha_i)]$, with $\beta_i = (\tau_i', \alpha_i)' = \beta_{i0}$ or $\beta_i = \bar{\beta}_i$, where $\bar{\beta}_i$ lies between $\hat{\beta}_i$ and β_{i0} so that $\bar{\beta}_i$ lies in $\mathcal{M}$ specified in Assumption 4.3(ii) with probability approaching one. Therefore, for T large enough, the denominator of each of the derivatives is bounded away from zero. Moreover, the numerator consists of derivatives of $\mathbb{E}[l_{i01}^2(\theta, \alpha_i)]$, $\mathbb{E}_{\tau_0}[\ell_{i03}(\theta, \alpha_i)]$ and $\mathbb{E}_{\tau_0}[\ell_{i02}(\theta, \alpha_i)]$, which can be shown to be bounded under Assumption 4.3(i) using the Cauchy–Schwarz inequality and the Jensen inequality, as in the proof of Lemma E.2. Due to the prefactor T^{-1}, the derivatives of $A(\theta, \beta_i)$ when evaluates at $\beta_i = \beta_{i0}$ or $\beta_i = \bar{\beta}_i$ are of order $O_p(T^{-1})$. Combining this with (C.1), (I.4), and Assumption 4.4, we see that the right hand side of (C.15) is of order $O_p(T^{-3/2})$. Moreover, the derivatives of $A(\theta, \beta_i)$ do not depend on outcome data when evaluated at β_{i0}. This implies for example that

$$
\mathbb{E}_{\tau_0}[A_{\alpha_i}(\theta, \beta_{i0})\delta_i(\theta)] = A_{\alpha_i}(\theta, \beta_{i0})\mathbb{E}_{\tau_0}[\delta_i(\theta)],
$$

which is of order $O_p(T^{-2})$ by (C.14). Similarly, $A_\gamma(\theta, \beta_{i0})'\mathbb{E}_{\tau_0}[(\tilde{\theta} - \theta_0)] = O_p(T^{-2})$ by Assumption 4.4. Further, $A_{\phi_i}(\theta, \beta_{i0})\mathbb{E}_{\tau_0}[\hat{\alpha}_i(\tilde{\theta}) - \alpha_{i0}] = O_p(T^{-2})$ by (C.13) and (C.8). Finally, we note that all terms involving second derivatives of $A(\theta, \beta_i)$ are of order $O_p(T^{-2})$ by (C.1), (I.4), and Assumption 4.4.   □

An explicit expression for the FOB of $\tilde{\theta}$. Notice that θ and γ are the same for each individual i in the definition of $\hat{F}_i(\theta, \beta_i)$ and $\partial_\theta \mathcal{B}_i^{(1)}(\theta, \beta_i)$. Furthermore, there is not need for a notation that differentiates between α_i and ϕ_i here. Thus, let $\zeta := (\theta', \alpha_1, \ldots, \alpha_n)' \in \Theta \times \mathcal{J} \times \cdots \times \mathcal{J}$, and further write $\zeta_0 := (\theta_0', \alpha_{i0}, \ldots, \alpha_{n0})'$ and $\hat{\zeta} := (\tilde{\theta}', \hat{\alpha}_1(\tilde{\theta}), \ldots, \hat{\alpha}_n(\tilde{\theta}))'$. Finally, define

$$
\mathcal{T}(\zeta) := \left[-\frac{1}{n}\sum_{i=1}^n \hat{F}_i(\theta, \beta_i)|_{\beta_i=(\theta', \alpha_i, \alpha_i)'} \right]^{-1}
$$

$$
\frac{1}{nT}\sum_{i=1}^n \partial_\theta \mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i=(\theta', \alpha_i, \alpha_i)'} = O_p(T^{-1}).
$$

<div align="right">(C.16)</div>

Then $\hat{\mathcal{T}} = \mathcal{T}(\hat{\zeta})$ and the FOB of $\tilde{\theta}$ can be expressed as $\mathcal{T}(\zeta)|_{\zeta=\zeta_0}$. The latter can be seen by first considering equation (F.1) together with the remarks in the proof of (6.7) in Schumann et al. (2021b) and taking conditional expectations $\mathbb{E}_{\tau_0}$ on both sides (notice that the conditional expectation operator $\mathbb{E}_{\tau_0}$ corresponds to their $\mathbb{E}_0$). Since, in their notation,

$l_{\dot{T}k}(\theta_0) = O_p(\frac{1}{\sqrt{nT}})$ while $\mathbb{E}_0[l_{\dot{T}k}(\theta_0)] = 0$ and $\tilde{\theta} - \theta_0 = O_p(\frac{1}{\sqrt{nT}}) + O_p(T^{-1})$,

$$\mathbb{E}_{\tau_0}[\tilde{\theta}] - \theta_0 = \left(n^{-1} \sum_{i=1}^{n} \mathbb{E}_{\tau_0}[\partial^2_{\theta\theta'}\ell_i(\theta, \hat{\alpha}_i(\theta))|_{\theta=\theta_0}] \right)^{-1}$$

$$\partial_\theta \mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\theta=\theta_0, \beta_i=\beta_{i0}} + O_p(T^{-2}).$$

Since the profile likelihood is "first-order Hessian biased,"

$$\mathbb{E}_{\tau_0}[\partial^2_{\theta\theta'}\ell_i(\theta, \hat{\alpha}_i(\theta))|_{\theta=\theta_0}] = \mathbb{E}_{\tau_0}[\partial^2_{\theta\theta'}\ell_i(\theta, \alpha_i^*(\theta))|_{\theta=\theta_0}] + O_p(T^{-1}).$$

Using an expansion and the fact that $\partial_\theta \mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\theta=\theta_0, \beta_i=\beta_{i0}} = O_p(T^{-1})$ together with $n/T \not\to 0$ then yields (C.16).[21] $\qquad\square$

In the next step, we show that our estimator for $\mathbb{E}_{\tau_0}[\tilde{\theta}] - \theta_0$ is first-order unbiased, i.e.,

$$\mathbb{E}_{\tau_0}[\hat{\mathcal{T}}] = \mathbb{E}_{\tau_0}[\mathcal{T}(\hat{\zeta})] = \mathbb{E}_{\tau_0}[\tilde{\theta}] - \theta_0 + O_p(T^{-2}). \tag{C.17}$$

Proof of (C.17). Recall $\zeta = (\theta', \alpha_1, \ldots, \alpha_n)' \in \Theta \times \mathcal{J} \times \cdots \times \mathcal{J}$ so that $\dim(\zeta) = p+n$. Let $k_1, k_2 \in \{1, \ldots, p+n\}$. As in the proof of (C.9) we denote partial derivatives with respect to components of ζ in the index, i.e.,

$$\mathcal{T}_k(\zeta) := \frac{\partial \mathcal{T}(\zeta)}{\partial \zeta_k} \quad \text{and} \quad \mathcal{T}_{k_1,k_2}(\zeta) := \frac{\partial^2 \mathcal{T}(\zeta)}{\partial \zeta_{k_1} \partial \zeta_{k_2}}.$$

Further recall $\hat{\zeta} = (\tilde{\theta}', \hat{\alpha}_1(\tilde{\theta}), \ldots, \hat{\alpha}_n(\tilde{\theta}))'$ and $\zeta_0 = (\theta_0', \alpha_{i0}, \ldots, \alpha_{n0})'$, and let $(\hat{\zeta} - \zeta_0)_k$ denote the kth component of $(\hat{\zeta} - \zeta_0)$. Using an expansion (ignoring constants),

$$\mathcal{T}(\zeta)|_{\zeta=\hat{\zeta}} = \mathcal{T}(\zeta_0) + \sum_{k=1}^{p+n} \mathcal{T}_k(\zeta)|_{\zeta=\zeta_0}(\hat{\zeta} - \zeta_0)_k$$

$$+ \sum_{k_1=1}^{p+n} \sum_{k_2=1}^{p+n} \mathcal{T}_{k_1,k_2}(\zeta)|_{\zeta=\bar{\zeta}}(\hat{\zeta} - \zeta_0)_{k_1}(\hat{\zeta} - \zeta_0)_{k_2},$$

where $\bar{\zeta}$ lies between $\hat{\zeta}$ and ζ_0. Taking expectations yields

$$\mathbb{E}_{\tau_0}[\mathcal{T}(\zeta)|_{\zeta=\hat{\zeta}}] - \mathcal{T}(\zeta_0) = \sum_{k=1}^{p+n} \mathcal{T}_k(\zeta)|_{\zeta=\zeta_0} \mathbb{E}_{\tau_0}[(\hat{\zeta} - \zeta_0)_k]$$

$$+ \sum_{k_1=1}^{p+n} \sum_{k_2=1}^{p+n} \mathbb{E}_{\tau_0}[\mathcal{T}_{k_1,k_2}(\zeta)|_{\zeta=\bar{\zeta}}(\hat{\zeta} - \zeta_0)_{k_1}(\hat{\zeta} - \zeta_0)_{k_2}],$$

where we have used that $\mathcal{T}_k(\zeta)|_{\zeta=\zeta_0}$ does not depend on outcome data. Clearly, $\mathbb{E}_{\tau_0}[(\hat{\zeta} - \zeta_0)_k] = O_p(T^{-1})$ for every $k \in \{1, \ldots, p+n\}$ by (C.1), (I.4), and Assumption 4.4(ii) and (iii). Moreover, $(\hat{\zeta} - \zeta_0)_k$ is of order $O_p(T^{-1/2})$ or lower and $(\hat{\zeta} - \zeta_0)_{k_1}(\hat{\zeta} - \zeta_0)_{k_2}$ is of order $O_p(T^{-1})$ or lower again by (C.1), (I.4), and Assumption 4.4(ii) and (iii). It thus remains to

---

[21] A more detailed argument is provided in Appendix I.

justify that derivatives of $\mathcal{T}(\zeta)$ of first and second orders are of order $O_p(T^{-1})$. To see this, notice that the derivatives of $\mathcal{T}(\zeta)$ evaluated at ζ_0 or $\bar{\zeta}$ consist of fractions with (powers of) $-n^{-1}\sum_{i=1}^n \hat{F}_i(\theta, \beta_i)$ evaluated at (θ_0, β_{i0}) or $(\bar{\theta}, \bar{\beta}_i)$ in the denominator. Since $\bar{\theta} \overset{p}{\to} \theta_0$ and $\bar{\beta}_i \overset{p}{\to} \beta_{i0}$, Assumption 4.3(v) implies that the denominators are bounded away from zero. The numerators contain first- and second-order derivatives of $\hat{F}_i(\theta, \beta_i)$, which can be shown to be bounded in a neighborhood of the true values θ_0 and β_{i0} by Assumption 4.3(i) and (ii) as in the proof of Lemma E.2. Moreover, the numerator contains derivatives up to third order of $\mathcal{B}_i^{(1)}(\theta, \beta_i)$, which are bounded in a neighborhood of the true values by Lemma E.2. Thus, since stochastic orders are not altered by taking averages across individuals by Assumption 4.1(iii), the first and second derivatives of $\mathcal{T}(\zeta)$ are of order $O_p(T^{-1})$ in a neighborhood of the true values due to the factor T^{-1} in the definition of $\mathcal{T}(\zeta)$. $\square$

Proof of (4.6). Recall that by (2.3),

$$\tilde{\mathcal{B}}_i^{(1)}(\theta, \beta_i) := \mathcal{B}_i^{(1)}(\theta, \beta_i) - \mathcal{B}_{i\alpha_i}^{(1)}(\theta, \beta_i)A(\theta, \beta_i) - \frac{1}{2}\mathcal{B}_{i\alpha_i\alpha_i}^{(1)}(\theta, \beta_i)V(\theta, \beta_i) - \mathcal{B}_{i\gamma}^{(1)}(\theta, \beta_i)'\hat{\mathcal{T}}$$

$$- \mathcal{B}_{i\phi_i}^{(1)}(\theta, \beta_i)A(\theta, \beta_i)|_{\theta=\tilde{\theta}} - \frac{1}{2}\mathcal{B}_{i\phi_i\phi_i}^{(1)}(\theta, \beta_i)V(\theta, \beta_i)|_{\theta=\tilde{\theta}} - \mathcal{B}_{i\alpha_i\phi_i}^{(1)}(\theta, \beta_i)\tilde{V}(\theta, \beta_i).$$

Using (C.13), we therefore need to show

$$\mathbb{E}_{\tau_0}[\mathcal{B}_{i\alpha_i}^{(1)}(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}A(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}] = \mathcal{B}_{i\alpha_i}^{(1)}(\theta, \beta_{i0})\mathbb{E}_{\tau_0}[\delta_i(\theta)] + O_p(T^{-2}), \tag{1}$$

$$\mathbb{E}_{\tau_0}[\mathcal{B}_{i\alpha_i\alpha_i}^{(1)}(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}V(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}] = \mathcal{B}_{i\alpha_i\alpha_i}^{(1)}(\theta, \beta_{i0})\mathbb{E}_{\tau_0}[\delta_i^2(\theta)] + O_p(T^{-2}), \tag{2}$$

$$\mathbb{E}_{\tau_0}[\mathcal{B}_{i\gamma}^{(1)}(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}'\hat{\mathcal{T}}] = \mathcal{B}_{i\gamma}^{(1)}(\theta, \beta_{i0})'\mathbb{E}_{\tau_0}[\tilde{\theta}] - \theta_0 + O_p(T^{-2}), \tag{3}$$

$$\mathbb{E}_{\tau_0}[\mathcal{B}_{i\phi_i}^{(1)}(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}A(\theta, \beta_i)|_{\theta=\tilde{\theta}, \beta_i=\hat{\beta}_i}] = \mathcal{B}_{i\phi_i}^{(1)}(\theta, \beta_{i0})\mathbb{E}_{\tau_0}[\delta_i(\theta_0)] + O_p(T^{-2}), \tag{4}$$

$$\mathbb{E}_{\tau_0}[\mathcal{B}_{i\phi_i\phi_i}^{(1)}(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}V(\theta, \beta_i)|_{\theta=\tilde{\theta}, \beta_i=\hat{\beta}_i}] = \mathcal{B}_{i\phi_i\phi_i}^{(1)}(\theta, \beta_{i0})\mathbb{E}_{\tau_0}[\delta_i^2(\theta_0)] + O_p(T^{-2}), \quad \text{and} \tag{5}$$

$$\mathbb{E}_{\tau_0}[\mathcal{B}_{i\alpha_i\phi_i}^{(1)}(\theta, \beta_{i0})\tilde{V}(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}] = \mathcal{B}_{i\alpha_i\phi_i}^{(1)}(\theta, \beta_{i0})\mathbb{E}_{\tau_0}[\delta_i(\theta)\delta_i(\theta_0)] + O_p(T^{-2}). \tag{6}$$

We first analyze (1) using $g(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}$, where

$$g(\theta, \beta_i) := \mathcal{B}_{i\alpha_i}^{(1)}(\theta, \beta_i)A(\theta, \beta_i).$$

Expanding $g(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}$ around β_{i0} now yields

$$g(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i} = g(\theta, \beta_{i0}) + \partial_{\beta_i'}g(\theta, \beta_i)|_{\beta_i=\beta_{i0}}(\hat{\beta}_i - \beta_{i0})$$

$$+ \frac{1}{2}(\hat{\beta}_i - \beta_{i0})'\partial_{\beta_i\beta_i'}g(\theta, \beta_i)|_{\beta_i=\bar{\beta}_{i1}}(\hat{\beta}_i - \beta_{i0}),$$

where $\bar{\beta}_{i1}$ lies between $\hat{\beta}_i$ and β_{i0}. Next, we notice that the first and second derivatives of $g(\theta, \beta_i)$ consist of first and second derivatives of $\mathcal{B}_{i\alpha_i}^{(1)}(\theta, \beta_i)$ and $A(\theta, \beta_i)$. Since $\bar{\beta}_i$ lies in any neighborhood of β_{i0} with probability approaching one as $T \to \infty$, Lemma E.2 implies that derivatives of $\mathcal{B}_{i\alpha_i}^{(1)}(\theta, \beta_i)$ are $O_p(1)$ uniformly across β_i in the neighborhood specified

in Lemma E.2. Further, as shown in the proof of (4.4), the first and second derivative of $A(\theta, \beta_i)$ are of order $O_p(T^{-1})$, again since $\bar{\beta}_i \xrightarrow{p} \beta_{i0}$. Thus, the first and second derivative of $g(\theta, \beta_i)$ are of order $O_p(T^{-1})$. Moreover, $\partial_{\beta_i} g(\theta, \beta_i)|_{\beta_i = \beta_{i0}}$ does not depend on outcome data, so that

$$\mathbb{E}_{\tau_0}[\partial_{\beta_i'} g(\theta, \beta_i)|_{\beta_i = \beta_{i0}} (\hat{\beta}_i - \beta_{i0})] = \partial_{\beta_i'} g(\theta, \beta_i)|_{\beta_i = \beta_{i0}} \mathbb{E}_{\tau_0}[(\hat{\beta}_i - \beta_{i0})] = O_p(T^{-2}),$$

where we have used (C.14), (C.8), and (C.16). Since $\mathbb{E}_{\tau_0}[(\hat{\beta}_i - \beta_{i0})'(\hat{\beta}_i - \beta_{i0})] = O_p(T^{-2})$,

$$\mathbb{E}_{\tau_0}[(\hat{\beta}_i - \beta_{i0})' \partial_{\beta_i \beta_i'} g(\theta, \beta_i)|_{\beta_i = \bar{\beta}_{i1}} (\hat{\beta}_i - \beta_{i0})] = O_p(T^{-3}).$$

In total, we have therefore shown that

$$T^{-1} \mathbb{E}_{\tau_0}[\mathcal{B}_{i\alpha_i}^{(1)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i} A(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i}] = T^{-1} \mathcal{B}_{i\alpha_i}^{(1)}(\theta, \beta_{i0}) A(\theta, \beta_{i0}) + O_p(T^{-3}).$$

Similarly, again using Lemma E.2 and the arguments outlined in the proof of (4.5) showing that the first and second derivatives of $V(\theta, \beta_i)$ with respect to the components of β_i are of order $O_p(T^{-1})$, we obtain (2). To show (3) we assume that $\dim(\theta) = 1$ in order to avoid tensor notation. Since we need to show (3) for each component, the same argument can be carried out component-wise when $\dim(\theta) > 1$. Recall that $\hat{\mathcal{T}} = \mathcal{T}(\zeta)|_{\zeta = \hat{\zeta}}$. We then note that

$$\mathcal{B}_{i\gamma}^{(1)}(\theta, \beta_i)|_{\beta_i = \hat{\beta}_i} \mathcal{T}(\zeta)|_{\zeta = \hat{\zeta}}| = [\mathcal{B}_{i\gamma}^{(1)}(\theta, \beta_{i0}) + \partial_{\beta_i'} \mathcal{B}_{i\gamma}^{(1)}(\theta, \beta_i)|_{\beta_i = \beta_{i0}} (\hat{\beta}_i - \beta_{i0})$$
$$+ \frac{1}{2}(\hat{\beta}_i - \beta_{i0})' \partial_{\beta_i \beta_i'} \mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \bar{\beta}_{i2}} (\hat{\beta}_i - \beta_{i0})][\mathcal{T}(\zeta_0) + Rem],$$

where $\bar{\beta}_{i2}$ lies again between $\hat{\beta}_i$ and β_{i0} and $Rem = O_p(T^{-3/2})$ and $\mathbb{E}_{\tau_0}[Rem] = O_p(T^{-2})$ by the proof of (C.17). Now, $\partial_{\beta_i} \mathcal{B}_{i\gamma}^{(1)}(\theta, \beta_i)|_{\beta_i = \beta_{i0}}$ and $\mathcal{T}(\zeta_0)$ do not depend on outcome data. Therefore,

$$\mathbb{E}_{\tau_0}[\partial_{\beta_i'} \mathcal{B}_{i\gamma}^{(1)}(\theta, \beta_i)|_{\beta_i = \beta_{i0}} (\hat{\beta}_i - \beta_{i0}) \mathcal{T}(\zeta_0)] = \partial_{\beta_i'} \mathcal{B}_{i\gamma}^{(1)}(\theta, \beta_i)|_{\beta_i = \beta_{i0}} \mathbb{E}_{\tau_0}[\hat{\beta}_i - \beta_{i0}] \mathcal{T}(\zeta_0)$$
$$= O_p(T^{-2}),$$

since $\mathbb{E}_{\tau_0}[\hat{\beta}_i] - \beta_{i0} = O_p(T^{-1})$ component-wise by (C.14), (C.8), and Assumption 4.4(ii) and (iii) and since $\mathcal{T}(\zeta_0) = O_p(T^{-1})$. Moreover,

$$\mathbb{E}_{\tau_0}[\mathcal{B}_{i\gamma}^{(1)}(\theta, \beta_{i0}) Rem] = \mathcal{B}_{i\gamma}^{(1)}(\theta, \beta_{i0}) \mathbb{E}_{\tau_0}[Rem] = O_p(T^{-2})$$

by (C.16) as $\mathcal{T}(\zeta_0) = O_p(T^{-1})$. Furthermore,

$$\mathbb{E}_{\tau_0}[(\hat{\beta}_i - \beta_{i0})' \partial_{\beta_i \beta_i'} \mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i = \bar{\beta}_{i2}} (\hat{\beta}_i - \beta_{i0}) \mathcal{T}(\zeta_0)] = O_p(T^{-2})$$

by (C.8), (C.16) and Assumption 4.4(ii) and (iii). The expectation of the remaining terms can similarly be shown to be of sufficiently low order by combining (C.1), (I.4), (C.16), and Assumption 4.4(ii) and (iii).

To show (4), we again use an expansion (ignoring constants) to obtain

$$\mathcal{B}_{i\phi_i}^{(1)}(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i} = \mathcal{B}_{i\phi_i}^{(1)}(\theta, \beta_{i0}) + \partial_{\beta_i'}\mathcal{B}_{i\phi_i}^{(1)}(\theta, \beta_i)|_{\beta_i=\beta_{i0}}(\hat{\beta}_i - \beta_{i0})$$

$$+ (\hat{\beta}_i - \beta_{i0})'\partial_{\beta_i\beta_i'}\mathcal{B}_{i\phi_i}^{(1)}(\theta, \beta_i)|_{\beta_i=\bar{\beta}_{i3}}(\hat{\beta}_i - \beta_{i0}), \tag{C.18}$$

where again $\bar{\beta}_{i3}$ is some value between $\hat{\beta}_i$ and β_{i0}, and

$$A(\theta, \beta_i)|_{\theta=\tilde{\theta}, \beta_i=\hat{\beta}_i}$$

$$= A(\theta_0, \beta_{i0}) + \partial_{\theta'}A(\theta, \beta_i)|_{\theta=\theta_0, \beta_i=\beta_{i0}}(\tilde{\theta} - \theta_0) + (\tilde{\theta} - \theta_0)'\partial_{\theta\theta'}A(\theta, \beta_i)|_{\theta=\bar{\theta}, \beta_i=\beta_{i0}}(\tilde{\theta} - \theta_0)$$

$$+ \partial_{\beta_i'}A(\theta, \beta_i)|_{\theta=\theta_0, \beta_i=\beta_{i0}}(\hat{\beta}_i - \beta_{i0}) + (\hat{\beta}_i - \beta_{i0})'\partial_{\beta_i\beta_i'}A(\theta, \beta_i)|_{\theta=\theta_0, \beta_i=\bar{\beta}_{i4}}(\hat{\beta}_i - \beta_{i0})$$

$$+ (\tilde{\theta} - \theta_0)'\partial_{\beta_i\theta'}A(\theta, \beta_i)|_{\theta=\bar{\theta}, \beta_i=\bar{\beta}_{i4}}(\hat{\beta}_i - \beta_{i0}), \tag{C.19}$$

where $\bar{\theta}$ lies between $\hat{\theta}$ and θ_0 and $\bar{\beta}_{i4}$ lies between $\hat{\beta}_i$ and β_{i0}. Recall that $A(\theta_0, \beta_{i0}) = \mathbb{E}_{\tau_0}[\delta_i(\theta)]$ by (4.4). As argued in the proof of (4.4), the first and second derivatives of $A(\theta, \beta_i)$ are of order $O_p(T^{-1})$ uniformly over θ and over β_i in a neighborhood of β_{i0}. Hence, by Assumption 4.4(ii) and (iii), (C.14), and (C.8),

$$\mathbb{E}_{\tau_0}[(\tilde{\theta} - \theta_0)'\partial_{\beta_i\theta'}A(\theta, \beta_i)|_{\theta=\bar{\theta}, \beta_i=\bar{\beta}_{i4}}(\hat{\beta}_i - \beta_{i0})] = O_p(T^{-5/2}).$$

Similarly, using that $\partial_{\beta_i\beta_i'}\mathcal{B}_{i\phi_i}^{(1)}(\theta, \beta_i)|_{\beta_i=\bar{\beta}_{i3}} = O_p(1)$ by Lemma E.2,

$$\mathbb{E}_{\tau_0}[(\hat{\beta}_i - \beta_{i0})'\partial_{\beta_i\beta_i'}\mathcal{B}_{i\phi_i}^{(1)}(\theta, \beta_i)|_{\beta_i=\bar{\beta}_{i3}}(\hat{\beta}_i - \beta_{i0})] = O_p(T^{-2})$$

and

$$\mathbb{E}_{\tau_0}[(\tilde{\theta} - \theta_0)'\partial_{\theta\theta'}A(\theta, \beta_i)|_{\theta=\bar{\theta}, \beta_i=\beta_{i0}}(\tilde{\theta} - \theta_0)] = O_p(T^{-3}).$$

Next, we notice that the first derivatives in (C.18) and (C.19) do not depend on outcome data. Since in addition $\mathbb{E}_{\tau_0}[\hat{\beta}_i] - \beta_{i0} = O_p(T^{-1})$ and $\mathbb{E}_{\tau_0}[\tilde{\theta}] - \theta_0 = O_p(T^{-1})$ component-wise, multiplying (C.18) and (C.19) yields (4). Parts (5) and (6) can be shown using analogous arguments. $\square$

Appendix D. Proofs for Section 4.3

This section contains the proof for asymptotic normality of $\hat{\theta}$.

Proof of (4.9). Let $\mathcal{B}$ denote an open ball centered at θ_0 and let $\overline{\mathcal{B}}$ denote its closure. Now, the first and second derivative of $\mathcal{B}_i^{(1)}(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}$ and $\mathcal{B}_i^{(2)}(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i}$ are bounded in probability uniformly in θ on $\overline{\mathcal{B}}$ by Lemma E.1 in conjunction with Assumption 4.3(ii). Moreover, by (C.7) we can expand the profile likelihood around the target likelihood up to an error term that is $o_p(1)$ uniformly on Θ. Since the same arguments hold for the first and second derivatives with respect to θ, we can write

$$\partial_{\theta\theta'}\ell_i^*(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i} = \partial_{\theta\theta'}\ell_i(\theta) + R_i^*(\theta) \tag{D.1}$$

with $\sup_{\theta\in\mathcal{B}}|\mathfrak{s}'R_i^*(\theta)\mathfrak{s}| = o_p(1)$. Thus, $n^{-1}\sum_{i=1}^n \sup_{\theta\in\mathcal{B}}|\mathfrak{s}'R_i^*(\theta)\mathfrak{s}| = o_p(1)$ by Assumption 4.1(iii). It is further shown subsequently that the target likelihood satisfies the second

Bartlett identity, i.e.,

$$\mathbb{E}_{\tau_0}[\partial_\theta \ell_i(\theta) \partial_{\theta'} \ell_i(\theta)]|_{\theta=\theta_0} = -\mathbb{E}_{\tau_0}[\partial_{\theta\theta'} \ell_i(\theta)]|_{\theta=\theta_0}. \tag{D.2}$$

Thus, by (D.1), (D.2), and the definition of F_i,

$$\frac{1}{n}\sum_{i=1}^{n} -\partial_{\theta\theta'}\ell_i^*(\theta,\beta_i)|_{\theta=\bar{\theta}, \beta_i=\hat{\beta}_i} - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[F_i]$$

$$= \frac{1}{n}\sum_{i=1}^{n} -\partial_{\theta\theta'}\ell_i(\theta)|_{\theta=\bar{\theta}} - \frac{1}{n}\sum_{i=1}^{n} -\mathbb{E}[\partial_{\theta\theta'}\ell_i(\theta)|_{\theta=\bar{\theta}}]$$

$$+ \frac{1}{n}\sum_{i=1}^{n} -\mathbb{E}[\partial_{\theta\theta'}\ell_i(\theta)|_{\theta=\bar{\theta}}] - \frac{1}{n}\sum_{i=1}^{n} -\mathbb{E}[\partial_{\theta\theta'}\ell_i(\theta)|_{\theta=\theta_0}] - \frac{1}{n}\sum_{i=1}^{n} R_i^*(\theta)|_{\theta=\bar{\theta}}. \tag{D.3}$$

Notice that for $\mathfrak{s} \in \mathbb{S} := \{s \in \mathbb{R}^p : \|s\| = 1\}$, we have

$$|\frac{1}{n}\sum_{i=1}^{n}\mathfrak{s}'R_i^*(\theta)|_{\theta=\bar{\theta}}\, \mathfrak{s}| \leq \frac{1}{n}\sum_{i=1}^{n}\sup_{\theta\in\mathcal{B}}|\mathfrak{s}'R_i^*(\theta)\mathfrak{s}| = o_{\mathrm{p}}(1). \tag{D.4}$$

Next, from (D.10), (I3), and Assumption 4.3(i) and (ii), the Cauchy–Schwarz inequality and

$$\partial_\theta \alpha_i^*(\theta) = -\lambda_{i11}(\theta)/\lambda_{i02}(\theta), \tag{D.5}$$

it follows that $\mathfrak{s}'\mathrm{var}[\partial_{\theta\theta'}\ell_i(\theta)] \leq \mathbb{E}[(\partial_{\theta\theta'}\ell_i(\theta))^2]\mathfrak{s} = O(1)$. Moreover, with the same arguments it can be shown that the expectation of the third derivative with respect to θ of $\ell_i(\theta)$ is bounded. By Corollary 2.2 in Newey (1991) and the discussion following it (p. 1163), we thus have that $n^{-1}\sum_{i=1}^{n}\mathbb{E}[\partial_{\theta\theta'}\ell_i(\theta)]$ is equicontinuous on $\overline{\mathcal{B}}$ and

$$\sup_{\theta\in\overline{\mathcal{B}}}|\frac{1}{n}\sum_{i=1}^{n}\mathfrak{s}'[\partial_{\theta\theta'}\ell_i(\theta) - \mathbb{E}[\partial_{\theta\theta'}\ell_i(\theta)]]\mathfrak{s}| = o_{\mathrm{p}}(1), \quad \text{as } n \to \infty.$$

Consequently, since $\bar{\theta} \in \overline{\mathcal{B}}$ w.p.a.1,

$$\left|\frac{1}{n}\sum_{i=1}^{n} -\mathfrak{s}'\partial_{\theta\theta'}\ell_i(\theta)_{\theta=\bar{\theta}}\, \mathfrak{s} - \frac{1}{n}\sum_{i=1}^{n} -\mathbb{E}[\mathfrak{s}'\partial_{\theta\theta'}\ell_i(\theta)_{\theta=\bar{\theta}}\mathfrak{s}]\right|$$

$$\leq \sup_{\theta\in\overline{\mathcal{B}}}\left|\frac{1}{n}\sum_{i=1}^{n} -\mathfrak{s}'\partial_{\theta\theta'}\ell_i(\theta)\mathfrak{s} - \frac{1}{n}\sum_{i=1}^{n} -\mathbb{E}[\mathfrak{s}'\partial_{\theta\theta'}\ell_i(\theta)]\mathfrak{s}\right| = o_{\mathrm{p}}(1). \tag{D.6}$$

Moreover, since the sequence $(\bar{Q}_n)$ is equicontinuous on $\overline{\mathcal{B}}$, for each $\epsilon > 0$ there exist $\eta > 0$ and $n_0 \in \mathbb{N}$ such that $n > n_0 \implies \sup_{\{\theta_1\in\overline{\mathcal{B}}:\|\theta_1-\theta_0\|<\eta\}}|\bar{Q}_n(\theta_1) - \bar{Q}_n(\theta_0)| < \epsilon$. Hence, since $\bar{\theta} \in \overline{\mathcal{B}(\theta_0, \delta \wedge \eta)}$ w.p.a.1, where δ is the radius of $\mathcal{B}$,

$$\frac{1}{n}\sum_{i=1}^{n} -\mathfrak{s}'\mathbb{E}[\partial_{\theta\theta'}\ell_i(\theta)|_{\theta=\bar{\theta}}]\mathfrak{s} - \frac{1}{n}\sum_{i=1}^{n} -\mathfrak{s}'\mathbb{E}[\partial_{\theta\theta'}\ell_i(\theta)|_{\theta=\theta_0}]\mathfrak{s} = o(1). \tag{D.7}$$

By (D.3), (D.4), (D.6), and (D.7),

$$\frac{1}{n}\sum_{i=1}^{n} -\partial_{\theta\theta'}\ell_i^*(\theta,\beta_i)|_{\theta=\bar{\theta},\,\beta_i=\hat{\beta}_i} - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[F_i] = o_p(1).$$

Hence, (4.9) follows from Assumption 4.3(v). □

Proof of (D.2). First, notice that

$$\partial_\theta \ell_i(\theta,\alpha_i^*(\theta)) = \ell_{i10}(\theta) + \ell_{i01}(\theta)\partial_\theta\alpha_i^*(\theta) = \ell_{i10}(\theta) - \ell_{i01}(\theta)\lambda_{i02}^{-1}(\theta)\lambda_{i11}(\theta), \qquad \textbf{(D.8)}$$

where we have used $\partial_\theta\alpha_i^*(\theta) = \lambda_{i02}^{-1}(\theta)\lambda_{i11}(\theta)$ in the last equation. Therefore,

$$[\partial_\theta\ell_i(\theta)\partial_{\theta'}\ell_i(\theta)]_{\theta=\theta_0} = \ell_{i10}\ell_{i10}' - \ell_{i10}\ell_{i01}\lambda_{i02}^{-1}\lambda_{i11} - \ell_{i01}\lambda_{i02}^{-1}\lambda_{i11}\ell_{i10}' + \ell_{i01}^2\lambda_{i02}^{-1}\lambda_{i11}\lambda_{i11}'.$$

Therefore, using the Bartlett identities $T\mathbb{E}_{\tau_0}[\ell_{i10}\ell_{i10}'] = -\lambda_{i20}$, $T\mathbb{E}_{\tau_0}[\ell_{i01}^2(\theta,\alpha_i)] = -\mathbb{E}_{\tau_0}[\ell_{i02}(\theta,\alpha_i)]$ and $T\mathbb{E}_{\tau_0}[\ell_{i01}\ell_{i10}] = -\lambda_{i11}$, we get

$$T\mathbb{E}_{\tau_0}[\partial_\theta\ell_i(\theta)\partial_{\theta'}\ell_i(\theta)]_{\theta=\theta_0} = -\lambda_{i20} + \lambda_{i02}^{-1}\lambda_{i11}\lambda_{i11}'. \qquad \textbf{(D.9)}$$

Next, differentiating (D.8) with respect to θ,

$$\partial_{\theta\theta'}\ell_i(\theta) = \ell_{i20}(\theta) + \ell_{i11}(\theta)\partial_{\theta'}\alpha_i^*(\theta) + \partial_\theta\alpha_i^*(\theta)[\ell_{i11}'(\theta) + \ell_{i02}(\theta)\partial_\theta\alpha_i^*(\theta)]$$
$$+ \ell_{i\alpha}(\theta)\partial_{\theta\theta'}\alpha_i^*(\theta).$$

Hence,

$$\mathbb{E}_{\tau_0}[\partial_{\theta\theta'}\ell_i(\theta)] = \lambda_{i20}(\theta) + \lambda_{i11}(\theta)\partial_\theta\alpha_i^*(\theta) + \partial_\theta\alpha_i^*(\theta)[\lambda_{i11}'(\theta) + \lambda_{i02}(\theta)\partial_{\theta'}\alpha_i^*(\theta)],$$
$$\textbf{(D.10)}$$

where we have used that by definition of the target value $\lambda_{i01}(\theta) = 0$. Again using (D.5) now yields

$$\mathbb{E}_{\tau_0}[\partial_{\theta\theta'}\ell_i(\theta)]|_{\theta=\theta_0} = \lambda_{i20} - \lambda_{i02}^{-1}\lambda_{i11}\lambda_{i11}'. \qquad \textbf{(D.11)}$$

Comparing (D.9) and (D.11), the desired result follows. □

Justification for Assumption 4.5. We need to justify that each component of $\partial_\theta R_i(\theta)$ is of order $O_p(T^{-3})$ when evaluated at $\theta = \theta_0$. In order to keep the notation simple, we assume here that $\dim(\theta) = 1$ and write $r_i(\theta) := T^3 R_i(\theta)$ and $r_i'(\theta_0) := \partial_\theta r_i(\theta)|_{\theta=\theta_0}$. By (4.7), $r_i(\theta) = O_p(1)$. Hence, since the derivative of $r_i(\theta)$ exists at $\theta = \theta_0$ by Assumption 4.3(i), we need to show that for each $\delta > 0$ there exists some constant $M < \infty$ and $T_0 \in \mathbb{N}$ such that $T \geq T_0$ implies

$$\mathbb{P}(|r_i'(\theta)|_{\theta=\theta_0}| > M) < \delta.$$

Notice that the argument in r_i and r_i' does not drift with n or T (i.e., θ is not allowed to involve a multiplicative factor of n or T) as this violates the compactness assumption in Assumption 4.1. Since the derivative exists and θ_0 lies in the interior of Θ, there is some $\tau > 0$ such that $0 < h < \tau$ implies $\theta_0 + h \in \Theta$ and

$$\left| \frac{r_i(\theta_0+h)}{h} - \frac{r_i(\theta_0)}{h} - r_i'(\theta_0) \right| < 1,$$

which implies that

$$|r_i'(\theta_0)| < 1 + |r_i(\theta_0 + h)/h| + |r_i(\theta_0)/h|.$$

Fix h with $0 < h < \tau$. Since for each $\theta \in \Theta$ we have shown that $r_i(\theta) = O_p(1)$, there exists some constants M_1 such that

$$\mathbb{P}(|r_i(\theta_0 + h)| > M_1) < \delta/2$$

and

$$\mathbb{P}(|r_i(\theta_0)| > M_1) < \delta/2.$$

Therefore,

$$\mathbb{P}(|r_i'(\theta_0)| > M) \leq \mathbb{P}(1 + |r_i(\theta_0 + h)|/h + |r_i(\theta_0)|/h > M)$$

$$\leq \mathbb{P}(|r_i(\theta_0 + h)| > h(M-1)/2) + \mathbb{P}(|r_i(\theta_0)| > h(M-1)/2) < \delta,$$

if M is chosen such that $M_1 < h(M-1)/2$. □

Proof of (4.10). Recall that by (2.5) and (C.7)

$$\ell_i^*(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i} = \ell_i(\theta) + \mathcal{B}_i^*(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i} + \mathrm{Rem}^{(0)}(\theta),$$

where

$$\mathcal{B}_i^*(\theta, \beta_i) := \frac{B_i^{(1)}(\theta)}{T} + \frac{B_i^{(2,1)}(\theta)}{T^{3/2}} + \frac{B_i^{(2,2)}(\theta)}{T^2} - \frac{\tilde{\mathcal{B}}_i^{(1)}(\theta, \beta_i)}{T} - \frac{\tilde{\mathcal{B}}_i^{(2)}(\theta, \beta_i)}{T^2},$$

and $\mathbb{E}_{\tau_0}[\mathrm{Rem}^{(0)}(\theta)] = O_p(T^{-3})$. As in the proof of (4.2) and (4.6), we can use a second-order expansion around β_{i0} to show that $\mathcal{B}_i^*(\theta, \beta_i)|_{\beta_i=\hat{\beta}_i} = \mathcal{B}_i^*(\theta, \beta_{i0}) + R_i^*(\theta)$ with $\mathbb{E}_{\tau_0}[\mathcal{B}_i^*(\theta, \beta_{i0})] = 0$ for every $\theta \in \Theta$ and where $R_i^*(\theta)$ consists of the remainder terms in (4.2) and (4.6) and thus $\mathbb{E}_{\tau_0}[R_i^*(\theta)] = O_p(T^{-3})$. Moreover, $\mathrm{Rem}^{(0)}(\theta) + R_i^*(\theta) = R_i(\theta)$, where $R_i(\theta)$ is the remainder term in (4.7) and Assumption 4.5. Since differentiation and integration is interchangeable, this in particular implies that

$$\mathbb{E}_{\tau_0}[\partial_\theta \mathcal{B}_i^*(\theta, \beta_{i0})]|_{\theta=\theta_0} = \partial_\theta \mathbb{E}_{\tau_0}[\mathcal{B}_i^*(\theta, \beta_{i0})]|_{\theta=\theta_0} = 0. \tag{D.12}$$

Now,

$$\sqrt{\frac{T}{n}} \sum_{i=1}^n \partial_\theta \ell_i^*(\theta, \beta_i)|_{\theta=\theta_0, \beta_i=\hat{\beta}_i} = \sqrt{\frac{T}{n}} \sum_{i=1}^n \partial_\theta \ell_i(\theta)|_{\theta=\theta_0}$$

$$+ \sqrt{\frac{T}{n}} \sum_{i=1}^n \partial_\theta \mathcal{B}_i^*(\theta, \beta_{i0})|_{\theta=\theta_0} + \sqrt{\frac{T}{n}} \sum_{i=1}^n \partial_\theta R_i(\theta)|_{\theta=\theta_0}. \tag{D.13}$$

First, we consider the stochastic properties of the last term on the right-hand side of (D.13). By Assumption 4.5, we immediately have

$$\sqrt{\frac{T}{n}} \sum_{i=1}^n \partial_\theta R_i(\theta)|_{\theta=\theta_0} = \sqrt{nT} O_p(T^{-3}) = O_p\left(\sqrt{\frac{n}{T^5}}\right).$$

Since (D.12) together with iterated expectations implies $\mathbb{E}[\partial_\theta \mathcal{B}_i^*(\theta, \beta_i)]|_{\theta=\theta_0, \beta_i=\hat{\beta}_i} = 0$ for every i, we only need to consider the variance of the second term on the right-hand side of (D.13) in order to determine its stochastic properties. Notice that the leading term in $\partial_\theta \mathcal{B}_i^*(\theta, \beta_{i0})$ is of the form $T^{-1}f(\theta)$, where $\mathbb{E}[(f(\theta_0))^2] = O(1)$. Thus, $\text{var}[\partial_\theta \mathcal{B}_i^*(\theta, \beta_{i0})|_{\theta=\theta_0}] \leq \mathbb{E}[(\partial_\theta \mathcal{B}_i^*(\theta, \beta_{i0})|_{\theta=\theta_0})^2] = O(T^{-2})$. Therefore, 4.1(iii) implies

$$\text{var}[\sqrt{\frac{T}{n}} \sum_{i=1}^n \partial_\theta \mathcal{B}_i(\theta, \beta_{i0})|_{\theta=\theta_0}] = T\,\text{var}[\partial_\theta \mathcal{B}_i(\theta, \beta_{i0})|_{\theta=\theta_0}] = O(T^{-1}),$$

and hence,

$$\sqrt{\frac{T}{n}} \sum_{i=1}^n \partial_\theta \mathcal{B}_i^*(\theta, \beta_{i0})|_{\theta=\theta_0} = O_p(T^{-1/2}).$$

In total,

$$\sqrt{\frac{T}{n}} \sum_{i=1}^n \partial_\theta \mathcal{B}_i^*(\theta, \beta_{i0})|_{\theta=\theta_0} + \sqrt{\frac{T}{n}} \sum_{i=1}^n \partial_\theta R_i(\theta)|_{\theta=\theta_0} = O_p\left(\frac{1}{\sqrt{T}}\right) + O_p\left(\sqrt{\frac{n}{T^5}}\right),$$

which in combination with (D.13) shows the desired result. $\qquad\square$

Proof of (4.11). Let $\mathfrak{s} \in \mathbb{S} := \{s \in \mathbb{R}^p : \|s\| = 1\}$. We begin by showing that

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}[(\mathfrak{s}' \partial_\theta \ell_i(\theta)|_{\theta=\theta_0})^4] = O(T^{-2}) \tag{D.14}$$

for any $\mathfrak{s} \in \mathbb{S}$. Notice that $\partial_\theta \ell_i(\theta)|_{\theta=\theta_0} = \ell_{i10} - \ell_{i01}\lambda_{i02}^{-1}\lambda_{i11}$ by (D.8), which after centering yields

$$\partial_\theta \ell_i(\theta)|_{\theta=\theta_0} = T^{-1/2}l_{i10} - T^{-1/2}l_{i01}\lambda_{i02}^{-1}\lambda_{i11},$$

where we have used that $\mathbb{E}_{\tau_0}[\ell_{i10}] = \mathbb{E}_{\tau_0}[\ell_{i01}] = 0$ by score unbiasedness of the likelihood at the true values. Using Assumptions 4.3(i) together with the Jensen inequality and the Cauchy–Schwarz inequality implies that $\mathbb{E}[(\mathfrak{s}' \partial_\theta \ell_i(\theta)|_{\theta=\theta_0})^4] = O(T^{-2})$ as $T \to \infty$. Thus, (D.14) follows. Now, let $\mathfrak{s} \in \mathbb{S}$, and observe that we can write $\sqrt{T/n} \sum_{i=1}^n \mathfrak{s}' \partial_\theta \ell_i(\theta)|_{\theta=\theta_0} = \sum_{i=1}^n W_{ni}$, where $W_{ki} := \sqrt{T_k/n}\mathfrak{s}' \partial_\theta \ell_{ik}(\theta)|_{\theta=\theta_0}$ is a triangular array of random variables for $1 \leq i \leq n$ and $1 \leq k \leq i$. Let $\epsilon > 0$. Then,

$$\sum_{i=1}^n \mathbb{E}[W_{ni}^2 \mathbb{1}(|W_{ni}| > \epsilon)] \leq \frac{1}{\epsilon^2} \sum_{i=1}^n \mathbb{E}[W_{ni}^4] = \frac{1}{\epsilon^2} \frac{T^2}{n^2} \sum_{i=1}^n \mathbb{E}[(\mathfrak{s}' \partial_\theta \ell_i(\theta)_{\theta=\theta_0})^4].$$

By (D.14), $\sum_{i=1}^n \mathbb{E}[W_{ni}^2 \mathbb{1}(|W_{ni}| > \epsilon)] = O(n^{-1})$ as $n \to \infty$, i.e., the Lindeberg condition is satisfied. Moreover, since $\mathbb{E}_{\tau_0}[\partial_\theta \ell_i(\theta)|_{\theta=\theta_0}] = \mathbb{E}_{\tau_0}[\ell_{i10}] - \lambda_{i02}^{-1}\lambda_{11}\mathbb{E}_{\tau_0}[\ell_{i01}] = 0$ by score

unbiasedness, $\mathbb{E}[\partial_\theta \ell_i(\theta)|_{\theta=\theta_0}] = 0$ by iterated expectations. Thus,

$$
\sum_{i=1}^{n} \mathrm{var}[W_{ni}] = \frac{T}{n} \sum_{i=1}^{n} \mathbb{E}[\mathfrak{s}'[\partial_\theta \ell_i(\theta)\partial_{\theta'}\ell_i(\theta)]|_{\theta=\theta_0}\mathfrak{s}]
$$

$$
= \frac{T}{n} \sum_{i=1}^{n} \mathfrak{s}'\mathbb{E}[\mathbb{E}_{\tau_0}[\partial_\theta \ell_i(\theta)\partial_{\theta'}\ell_i(\theta)]|_{\theta=\theta_0}\mathfrak{s}]
$$

$$
= \frac{1}{n} \sum_{i=1}^{n} \mathfrak{s}'\mathbb{E}[F_i]\mathfrak{s},
$$

where by definition $F_i = \mathbb{E}_{\tau_0}[\partial_\theta \ell_i(\theta)\partial_{\theta'}\ell_i(\theta)]|_{\theta=\theta_0}$. Hence, by the Lindeberg–Feller CLT for triangular arrays (Durrett, 1991, p. 98) and Assumption 4.3(v), $\sum_{i=1}^{n} W_{ni} \xrightarrow{\mathrm{d}} \mathcal{N}(0, \mathfrak{s}'F\mathfrak{s})$ as $n \to \infty$. Consequently, by the Crámer–Wold device,

$$
\sqrt{\frac{T}{n}} \sum_{i=1}^{n} \partial_\theta \ell_i(\theta_0) \xrightarrow{\mathrm{d}} \mathcal{N}(0, F) \quad \text{as } n \to \infty.
$$

$\square$

Appendix E. Auxiliary Results

In this section, we derive the stochastic orders of products of likelihood derivatives. Moreover, we show that derivatives of $\mathcal{B}_i^{(1)}(\theta, \beta_i)$ with respect to β_i are bounded in probability.

LEMMA E.1. *Let* $h \in \{1, \ldots, 18\}$, $v := (v_1, \ldots, v_p)$ *be a vector of nonnegative integers,* $|v| = \sum_{s=1}^{p} v_s$ *with* $|v| \le 4$ *and let* $0 \le \mu \le 9$. *Under Assumption 4.3(i),*

$$
\mathbb{E}[\sup_{\tau_i \in \Theta \times \mathcal{J}} \mathbb{E}_{\tau_i}[\sup_{(\theta, \alpha_i) \in \Theta \times \mathcal{J}} \prod_{j=1}^{h} D^{v_j + \mu_j} l_i(\theta, \alpha_i; \tau_i)]] = \begin{cases} O(1), & h \text{ even,} \\ O(T^{-1/2}), & h \text{ odd,} \end{cases}
$$

where

$$
D^{v+\mu} l_i(\theta, \alpha_i; \tau_i) = \frac{\partial^{|v|+\mu} l_i(\theta, \alpha_i; \tau_i)}{\partial \theta_1^{v_1} \cdots \partial \theta_p^{v_p} \partial \alpha_i^{\mu}}.
$$

Proof of Lemma E.1. First notice that by compactness of the parameter spaces and the continuity of $\mathbb{E}_{\tau_i}[D^{v+\mu} l_i(\theta, \alpha_i)]$ in τ_i, θ and α_i, for each panel length T there exist some $\tilde{\tau}_i \in \Theta \times \mathcal{J}$, $\tilde{\theta} \in \Theta$ and $\tilde{\alpha}_i \in \mathcal{J}$ such that

$$
\sup_{\tau_i \in \Theta \times \mathcal{J}} \mathbb{E}_{\tau_i}[\sup_{(\theta, \alpha_i) \in \Theta \times \mathcal{J}} D^{v+\mu} l_i(\theta, \alpha_i; \tau_i)] = \mathbb{E}_{\tilde{\tau}_i}[D^{v+\mu} l_i(\tilde{\theta}, \tilde{\alpha}_i; \tilde{\tau}_i)] = 0.
$$

Thus, the mean-zero property of centralized likelihood derivatives is not affected by taking the supremum in each parameter. Next,

$$\mathbb{E}[\sup_{\tau_i \in \Theta \times \mathcal{J}} \mathbb{E}_{\tau_i}[\sup_{(\theta, \alpha_i) \in \Theta \times \mathcal{J}} \prod_{j=1}^{h} D^{\nu_j + \mu_j} l_i(\theta, \alpha_i; \tau_i)]]$$

$$= T^{-h/2} \sum_{k_1=1}^{T} \cdots \sum_{k_h=1}^{T} \mathbb{E}[\sup_{\tau_i \in \Theta \times \mathcal{J}} \mathbb{E}_{\tau_i}[\sup_{(\theta, \alpha_i) \in \Theta \times \mathcal{J}} \prod_{j=1}^{h} D^{\nu_j + \mu_j} l_{ik_j}(\theta, \alpha_i; \tau_i)]].$$

$$\textbf{(E.1)}$$

Now, to bound the summands,

$$|\mathbb{E}[\sup_{\tau_i \in \Theta \times \mathcal{J}} \mathbb{E}_{\tau_i}[\sup_{(\theta, \alpha_i) \in \Theta \times \mathcal{J}} \prod_{j=1}^{h} D^{\nu_j + \mu_j} l_{ik_j}(\theta, \alpha_i; \tau_i)]]|$$

$$\leq \mathbb{E}[\sup_{\tau_i \in \Theta \times \mathcal{J}} \mathbb{E}_{\tau_i}[\prod_{j=1}^{h} \sup_{(\theta, \alpha_i) \in \Theta \times \mathcal{J}} |D^{\nu_j + \mu_j} l_{ik_j}(\theta, \alpha_i; \tau_i)|]]. \qquad \textbf{(E.2)}$$

Repeated use of the Cauchy–Schwarz inequality and the Jensen inequality together with Assumption 4.3(i) now shows that the right-hand side of (E.2), and thus every summand on the right hand side of (E.1), is bounded. For instance, for $h = 2$, applying the Cauchy–Schwarz inequality on the right-hand side of (E.2) yields the upper bound

$$\left(\tilde{E} \left[\left(\sup_{(\theta, \alpha_i) \in \Theta \times \mathcal{J}} |D^{\nu_1 + \mu_1} l_{ik_1}(\theta, \alpha_i; \tau_i)| \right)^2 \right] \right)^{1/2}$$

$$\left(\tilde{E} \left[\left(\sup_{(\theta, \alpha_i) \in \Theta \times \mathcal{J}} |D^{\nu_2 + \mu_2} l_{ik_2}(\theta, \alpha_i; \tau_i)| \right)^2 \right] \right),$$

where $\tilde{E}$ denotes integration with respect to the unconditional density evaluated at $\tilde{\tau}_i$ corresponding to $\mathbb{E}[\sup_{\tau_i \in \Theta \times \mathcal{J}} \mathbb{E}_{\tau_i}[\cdot]]$. By the Jensen inequality and Assumption 4.3(i),

$$\tilde{E} \left[\left(\sup_{(\theta, \alpha_i) \in \Theta \times \mathcal{J}} |D^{\nu_1 + \mu_1} l_{ik_1}(\theta, \alpha_i; \tau_i)| \right)^2 \right]$$

$$\leq c \left(\tilde{E}[M_{ik_1}^2] + \mathbb{E} \left[\left(\sup_{\tau_i \in \Theta \times \mathcal{J}} \mathbb{E}_{\tau_i}[M_{ik_1}] \right)^2 \right] \right) = O(1),$$

where c denotes a constant. Next, we note that by time-independence and because $\mathbb{E}_{\tau_i}[l_{it}(\theta, \alpha_i; \tau_i)] = 0$ by definition, $\mathbb{E}_{\tilde{\tau}_i}[\prod_{j=1}^{h} D^{\nu_j + \mu_j} l_{ik_j}(\tilde{\theta}, \tilde{\alpha}_i; \tilde{\tau}_i)] = 0$ if there exists some $j^* \in \{1, \ldots, h\}$ such that time period k_{j^*} differs from k_j for all $j \neq j^*$, since this implies

that

$$
\mathbb{E}_{\tilde{\tau}_i}\left[\prod_{j=1}^{h} D^{\nu_j+\mu_j} l_{ik_j}(\tilde{\theta},\tilde{\alpha}_i;\tilde{\tau}_i)\right] = \mathbb{E}_{\tilde{\tau}_i}\left[\prod_{j=1,j\neq j^*}^{h} D^{\nu_j+\mu_j} l_{ik_j}(\tilde{\theta},\tilde{\alpha}_i;\tilde{\tau}_i)\right]
$$

$$
\mathbb{E}_{\tilde{\tau}_i}\left[D^{\nu_{j^*}+\mu_{j^*}} l_{ik_{j^*}}(\tilde{\theta},\tilde{\alpha}_i;\tilde{\tau}_i)\right] = 0.
$$

Consequently, any summation over an index that is not paired with others drops out. Therefore, if h is even, there are at most $h/2$ pairs of indices and thus at most $h/2$ sums over absolutely bounded summands scaled by a factor $T^{-h/2}$ on the right-hand side of (C.8). The total expression is thus of order $O(1)$. If h is odd, the maximum number of summations is reached with $(h-3)/2$ pairs of indices and a single triplet of indices. Therefore, we have $(h-3)/2+1 = (h-1)/2$ summations over bounded summands, scaled by the factor $T^{-h/2}$. Consequently, the total expression is of order $O(T^{-1/2})$. $\qquad\square$

As the variance is bounded by the second moment, Lemma E.1 in particular implies

$$
\sup_{\tau_i\in\Theta\times\mathcal{J}} \mathbb{E}_{\tau_i}\left[\sup_{(\theta,\alpha_i)\in\Theta\times\mathcal{J}}\prod_{j=1}^{h} D^{\nu_j+\mu_j} l_i(\theta,\alpha_i;\tau_i)\right] = \begin{cases} O_{\mathrm{p}}(1), & h \text{ even}, \\ O_{\mathrm{p}}(T^{-1/2}), & h \text{ odd}, \end{cases} \tag{E.3}
$$

for $h = 1,\ldots,9$. The next Lemma shows that sufficient derivatives of $\mathcal{B}_i^{(1)}(\theta,\beta_i)$ are bounded. The proof proceeds by taking derivatives explicitly before showing boundedness using Assumption 4.3 together with repeated applications of the Cauchy–Schwarz inequality and the Jensen inequality.

LEMMA E.2. *Let* $\beta_i = (\theta',\phi_i,\alpha_i)' \in \Upsilon$ *with* $\dim(\beta_i) = p+2$. *Further, let* $\nu := (\nu_1,\ldots,\nu_{p+2})$ *be a vector of nonnegative integers* ν_s *with* $|\nu| := \sum_{s=1}^{p+2}\nu_s$ *and*

$$
D_{\beta_i}^{\nu}\mathcal{B}_i^{(1)}(\theta,\beta_i) = \frac{\partial^{|\nu|}\mathcal{B}_i^{(1)}(\theta,\beta_i)}{\partial\beta_{i1}^{\nu_1}\cdots\partial\beta_{ip+2}^{\nu_{p+2}}}.
$$

Then, there exists an open ball B *centered at* $\beta_{i0} = (\theta_0',\alpha_{i0},\alpha_i^*(\theta))'$ *such that for* T *large enough*

$$
\sup_{\beta_i\in B}\sup_{\theta\in\Theta} |D_{\beta_i}^{\nu}\mathcal{B}_i^{(1)}(\theta,\beta_i)| = O_{\mathrm{p}}(1)
$$

for all ν *with* $|\nu| \leq 4$.

Proof of Lemma E.2. Recall from (2.1)

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

where we used time-independence and $\mathbb{E}_{\tau_i}[l_{it01}(\theta,\alpha_i;\tau_i)] = 0$ in the second equation. Thus, derivatives of $\mathcal{B}_i^{(1)}(\theta,\beta_i)$ consist of terms with derivatives of $T^{-1}\sum_{t=1}^{T}\mathbb{E}_{\tau_i}[l_{it01}^2(\theta,\alpha_i;\tau_i)]$ and $T^{-1}\sum_{t=1}^{T}\mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)]$ in the numerator and $\mathbb{E}_{\tau_i}[\ell_{i02}(\theta,\alpha_i)]$ in the denominator. First, we will consider the latter term in the denominator. By Assumption 4.3(ii), there

exists an open ball $\mathcal{M}$ centered at β_{i0} and a constant $D > 0$ such that for all T large enough

$$\inf_{\beta_i \in \mathcal{M}} \inf_{\theta \in \Theta} |\mathbb{E}_{\tau_i}[\ell_{i02}(\theta, \alpha_i)]| > D.$$

Let from now on B be an open ball centered at β_{i0} such that $B \subset \mathcal{M}$. The denominator is then uniformly bounded away from zero over Θ and B. Next, we consider the terms in the numerator. Here, it follows from Assumption 4.3(i) that the derivatives with respect to α are bounded. For example,

$$\frac{1}{T} \sum_{t=1}^{T} |\partial_{\alpha_i}^4 \mathbb{E}_{\tau_i}[\ell_{it02}(\theta, \alpha_i)]| = \frac{1}{T} \sum_{t=1}^{T} |\mathbb{E}_{\tau_i}[\ell_{it06}(\theta, \alpha_i)]| \leq \frac{1}{T} \sum_{t=1}^{T} M_{it} = O_{\mathrm{p}}(1).$$

Moreover,

(1)

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

(2)

$$\partial_{\alpha_i}^2 \mathbb{E}_{\tau_i}[l_{i01}^2(\theta, \alpha_i; \tau_i)] = 2\mathbb{E}_{\tau_i}[l_{i01}(\theta, \alpha_i; \tau_i)l_{i03}(\theta, \alpha_i; \tau_i)] + 2\mathbb{E}_{\tau_i}[l_{i02}^2(\theta, \alpha_i; \tau_i)],$$

(3)

$$\partial_{\alpha_i}^3 \mathbb{E}_{\tau_i}[l_{i01}^2(\theta, \alpha_i; \tau_i)] = 6\mathbb{E}_{\tau_i}[l_{i02}(\theta, \alpha_i; \tau_i)l_{i03}(\theta, \alpha_i; \tau_i)] + 2\mathbb{E}_{\tau_i}[l_{i01}(\theta, \alpha_i; \tau_i)l_{i04}(\theta, \alpha_i; \tau_i)],$$

(4) and

$$\begin{aligned} \partial_{\alpha_i}^4 \mathbb{E}_{\tau_i}[l_{i01}^2(\theta, \alpha_i; \tau_i)] = {} & 8\mathbb{E}_{\tau_i}[l_{i02}(\theta, \alpha_i; \tau_i)l_{i04}(\theta, \alpha_i; \tau_i)] + 6\mathbb{E}_{\tau_i}[l_{i03}^2(\theta, \alpha_i; \tau_i)] \\ & + 2\mathbb{E}_{\tau_i}[l_{i02}(\theta, \alpha_i; \tau_i)l_{i04}(\theta, \alpha_i; \tau_i)] + 2\mathbb{E}_{\tau_i}[l_{i01}(\theta, \alpha_i; \tau_i)l_{i05}(\theta, \alpha_i; \tau_i)], \end{aligned}$$

so that each of the terms in (1)–(4) consists of the expectation of a product of two centered likelihood terms, where the expectation is taken with respect to the parameters γ and ϕ_i. As in the proof of (E.3), we can use the Cauchy–Schwarz inequality and the Jensen inequality to show that all terms on the right hand side are bounded uniformly over the parameter space and across T. To illustrate this, notice that since $\mathbb{E}_{\tau_i}[l_{it0m}(\theta, \alpha_i)] = 0$ by definition, independence over time implies

$$\sup_{\beta_i \in B} \sup_{\theta \in \Theta} |\mathbb{E}_{\tau_i}[l_{i0m_1}(\theta, \alpha_i; \tau_i)l_{i0m_2}(\theta, \alpha_i; \tau_i)]|$$

$$\leq \frac{1}{T} \sum_{t=1}^{T} \sup_{\beta_i \in B} \sup_{\theta \in \Theta} |\mathbb{E}_{\tau_i}[l_{it0m_1}(\theta, \alpha_i; \tau_i)l_{it0m_2}(\theta, \alpha_i; \tau_i)]|.$$

Applying the Cauchy–Schwarz inequality,

$$\sup_{\beta_i \in B} \sup_{\theta \in \Theta} |\mathbb{E}_{\tau_i}[l_{it0m_1}(\theta, \alpha_i; \tau_i)l_{it0m_2}(\theta, \alpha_i; \tau_i)]|$$

$$\leq \sqrt{\sup_{\beta_i \in B} \sup_{\theta \in \Theta} \mathbb{E}_{\tau_i}[l_{it0m_1}^2(\theta, \alpha_i; \tau_i)]} \sqrt{\sup_{\beta_i \in B} \sup_{\theta \in \Theta} \mathbb{E}_{\tau_i}[l_{it0m_2}^2(\theta, \alpha_i; \tau_i)]}.$$

Now, by the Jensen inequality,

$$\mathbb{E}_{\tau_i}[l_{it0m_1}^2(\theta, \alpha_i; \tau_i)] \leq 2(\mathbb{E}_{\tau_i}[\ell_{it0m_1}^2(\theta, \alpha_i)] + \mathbb{E}_{\tau_i}[\ell_{it0m_1}(\theta, \alpha_i)]^2)$$

$$\leq 2(\sup_{\tau_i} \mathbb{E}_{\tau_i}[M_{it}] + (\sup_{\tau_i} \mathbb{E}_{\tau_i}[M_{it}])^2)$$

so that, by Assumption 4.3 and the Jensen inequality,

$$\mathbb{E}[\sqrt{\sup_{\tau_i} \sup_{\theta, \alpha} \mathbb{E}_{\tau_i}[l_{it0m_1}^2(\theta, \alpha_i)]}] \leq \sqrt{2(\mathbb{E}[\sup_{\tau_i} \mathbb{E}_{\tau_i}[M_{it}]] + \mathbb{E}[(\sup_{\tau_i} \mathbb{E}_{\tau_i}[M_{it}])^2)]} = O_p(1).$$

Hence,

$$\frac{1}{T} \sum_{t=1}^{T} \sup_{\beta_i \in B} \sup_{\theta \in \Theta} |\mathbb{E}_{\tau_i}[l_{it0m_1}(\theta, \alpha_i; \tau_i)l_{it0m_2}(\theta, \alpha_i; \tau_i)]| = O_p(1),$$

which shows that

$$\sup_{\theta \in \Theta} \sup_{\beta_i \in B} |\partial_{\alpha_i}^4 \mathbb{E}_{\tau_i}[l_{i01}^2(\theta, \alpha_i; \tau_i)]| = O_p(1).$$

The remaining derivatives with respect to α_i can be handled in a similar way. Next, we consider derivatives with respect to ϕ_i. Using that

$$\partial_{\phi_i} f_{Y_{it}|\mathcal{X}_i, \gamma, \phi_i} = \ell_{it01}(\gamma, \phi_i) f_{Y_{it}|\mathcal{X}_i, \gamma, \phi_i},$$

we get

(1)

$$\partial_{\phi_i} \mathbb{E}_{\tau_i}[\ell_{it02}(\theta, \alpha_i)] = \mathbb{E}_{\tau_i}[\ell_{it02}(\theta, \alpha_i)\ell_{it01}(\gamma, \phi_i)],$$

(2)

$$\partial_{\phi_i}^2 \mathbb{E}_{\tau_i}[\ell_{it02}(\theta, \alpha_i)] = \mathbb{E}_{\tau_i}[\ell_{it02}(\theta, \alpha_i)\ell_{it02}(\gamma, \phi_i)] + \mathbb{E}_{\tau_i}[\ell_{it02}(\theta, \alpha_i)\ell_{it01}^2(\gamma, \phi_i)],$$

(3)

$$\partial_{\phi_i}^3 \mathbb{E}_{\tau_i}[\ell_{it02}(\theta, \alpha_i)] = \mathbb{E}_{\tau_i}[\ell_{it02}(\theta, \alpha_i)\ell_{it03}(\gamma, \phi_i)]$$

$$+ 3\mathbb{E}_{\tau_i}[\ell_{it02}(\theta, \alpha_i)\ell_{it012}(\gamma, \phi_i)\ell_{it01}(\gamma, \phi_i)] + \mathbb{E}_{\tau_i}[\ell_{it02}(\theta, \alpha_i)\ell_{it01}^3(\gamma, \phi_i)],$$

(4) and

$$\partial_{\phi_i}^4 \mathbb{E}_{\tau_i}[\ell_{it02}(\theta, \alpha_i)] = 4\mathbb{E}_{\tau_i}[\ell_{it02}(\theta, \alpha_i)\ell_{it03}(\gamma, \phi_i)\ell_{it01}(\gamma, \phi_i)]$$

$$+ 3\mathbb{E}_{\tau_i}[\ell_{it02}(\theta, \alpha_i)\ell_{it02}^2(\gamma, \phi_i)]$$

$$+ 6\mathbb{E}_{\tau_i}[\ell_{it02}(\theta, \alpha_i)\ell_{it02}(\gamma, \phi_i)\ell_{it01}^2(\gamma, \phi_i)] + \mathbb{E}_{\tau_i}[\ell_{it02}(\theta, \alpha_i)\ell_{it04}(\gamma, \phi_i)]$$

$$+ \mathbb{E}_{\tau_i}[\ell_{it02}(\theta, \alpha_i)\ell_{it01}^4(\gamma, \phi_i)].$$

Again, using the Cauchy–Schwarz inequality repeatedly, we can bound the supremum of each summand. For example,

$$|\mathbb{E}_{\tau_i}[\ell_{it02}(\theta, \alpha_i)\ell_{it01}^4(\gamma, \phi_i)]| \leq \sqrt{\mathbb{E}_{\tau_i}[\ell_{it02}^2(\theta, \alpha_i)]} \sqrt{\mathbb{E}_{\tau_0}[\ell_{it01}^8(\gamma, \phi_i)]}.$$

But,

$$|\ell_{it02}^2(\theta,\alpha_i)| \le \sup_{(\theta,\alpha_i)\in\Theta\times\mathcal{J}} |\ell_{it02}(\theta,\alpha_i)|^2 \le M_{it}^2$$

and

$$|\ell_{it01}^8(\theta,\alpha_i)| \le \sup_{(\theta,\alpha_i)\in\Theta\times\mathcal{J}} |\ell_{it01}(\theta,\alpha_i)|^8 \le M_{it}^8.$$

By Assumption 4.3(i) we therefore have

$$\frac{1}{T}\sum_{t=1}^T \sup_{\beta_i\in B}\sup_{\theta\in\Theta} |\mathbb{E}_{\tau_i}[\ell_{i02}(\theta,\alpha_i)\ell_{i01}^4(\gamma,\phi_i)]|$$

$$\le \frac{1}{T}\sum_{t=1}^T \sqrt{\sup_{(\gamma,\phi_i)\in\Theta\times\mathcal{J}} \mathbb{E}_{\tau_i}[M_{it}^2]}\sqrt{\sup_{(\gamma,\phi_i)\in\Theta\times\mathcal{J}} \mathbb{E}_{\tau_i}[M_{it}^8]} = O_{\mathrm{p}}(1).$$

Therefore, in total

$$\partial_{\phi_i}^4 \mathbb{E}_{\tau_i}[\ell_{i02}(\theta,\alpha_i)] = O_{\mathrm{p}}(1).$$

A similar argument can be used to show that all fourth-order derivatives of $\mathbb{E}_{\tau_i}[\ell_{i02}(\theta,\alpha_i)]$ are bounded. As an example for derivatives with respect to components of γ, let $\psi_1,\ldots,\psi_4\in\{1,\ldots,p\}$ so that γ_ψ denotes the ψth component of γ. Further, let $\ell_{it1m}^{\psi_1}(\theta,\alpha_i)$ denote the ψth component of the vector $\ell_{it1m}(\theta,\alpha_i)$, $\ell_{it2m}^{\psi_1\psi_2}(\theta,\alpha_i)$ denote the (ψ_1,ψ_2)th element of the matrix $\ell_{it2m}(\theta,\alpha_i)$, $\ell_{it2m}^{\psi_1\psi_2\psi_3}(\theta,\alpha_i)$ denote the (ψ_1,ψ_2,ψ_3)th element of the three-dimensional array $\ell_{it3m}(\theta,\alpha_i)$ and so on, and consider

(1)

$$\partial_{\gamma_{\psi_1}}\mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)] = \mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it10}^{\psi_1}(\gamma,\phi_i)],$$

(2)

$$\partial_{\gamma_{\psi_1}\gamma_{\psi_2}}^2\mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)] = \mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it20}^{\psi_1\psi_2}(\gamma,\phi_i)]$$
$$+ \mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it10}^{\psi_1}(\gamma,\phi_i)\ell_{it10}^{\psi_2}(\gamma,\phi_i)],$$

(3)

$$\partial_{\gamma_{\psi_1}\gamma_{\psi_2}\gamma_{\psi_3}}^3\mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)] = \mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it30}^{\psi_1\psi_2\psi_3}(\gamma,\phi_i)]$$

$$+ \mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it20}^{\psi_1\psi_2}(\gamma,\phi_i)\ell_{it10}^{\psi_3}(\gamma,\phi_i)] + \mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it20}^{\psi_1\psi_3}(\gamma,\phi_i)\ell_{it10}^{\psi_2}(\gamma,\phi_i)]$$

$$+ \mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it10}^{\psi_1}(\gamma,\phi_i)\ell_{it10}^{\psi_2\psi_3}(\gamma,\phi_i)]$$

$$+ \mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it01}^{\psi_1}(\gamma,\phi_i)\ell_{it01}^{\psi_2}(\gamma,\phi_i)\ell_{it01}^{\psi_3}(\gamma,\phi_i)],$$

(4) and

$$\partial_{\gamma_{\psi_1}\gamma_{\psi_2}\gamma_{\psi_3}\gamma_{\psi_4}}^4\mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)] = \mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it40}^{\psi_1\psi_2\psi_3\psi_4}(\gamma,\phi_i)]$$

$$+ \mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it30}^{\psi_1\psi_2\psi_3}(\gamma,\phi_i)\ell_{it10}^{\psi_4}(\gamma,\phi_i)]$$

$$+\mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it30}^{\psi_1\psi_2\psi_4}(\gamma,\phi_i)\ell_{it10}^{\psi_3}(\gamma,\phi_i)]$$

$$+\mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it30}^{\psi_1\psi_3\psi_4}(\gamma,\phi_i)\ell_{it10}^{\psi_2}(\gamma,\phi_i)]$$

$$+\mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it30}^{\psi_2\psi_3\psi_4}(\gamma,\phi_i)\ell_{it10}^{\psi_1}(\gamma,\phi_i)]$$

$$+\mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it20}^{\psi_1\psi_2}(\gamma,\phi_i)\ell_{it20}^{\psi_3\psi_4}(\gamma,\phi_i)]$$

$$+\mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it20}^{\psi_1\psi_3}(\gamma,\phi_i)\ell_{it20}^{\psi_2\psi_4}(\gamma,\phi_i)]$$

$$+\mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it20}^{\psi_1\psi_4}(\gamma,\phi_i)\ell_{it20}^{\psi_2\psi_3}(\gamma,\phi_i)]$$

$$+\mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it20}^{\psi_1\psi_2}(\gamma,\phi_i)\ell_{it10}^{\psi_3}(\gamma,\phi_i)\ell_{it10}^{\psi_4}(\gamma,\phi_i)]$$

$$+\mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it20}^{\psi_1\psi_3}(\gamma,\phi_i)\ell_{it10}^{\psi_2}(\gamma,\phi_i)\ell_{it10}^{\psi_4}(\gamma,\phi_i)]$$

$$+\mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it20}^{\psi_1\psi_4}(\gamma,\phi_i)\ell_{it10}^{\psi_2}(\gamma,\phi_i)\ell_{it10}^{\psi_3}(\gamma,\phi_i)]$$

$$+\mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it20}^{\psi_2\psi_3}(\gamma,\phi_i)\ell_{it10}^{\psi_1}(\gamma,\phi_i)\ell_{it10}^{\psi_4}(\gamma,\phi_i)]$$

$$+\mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it20}^{\psi_2\psi_4}(\gamma,\phi_i)\ell_{it10}^{\psi_1}(\gamma,\phi_i)\ell_{it10}^{\psi_3}(\gamma,\phi_i)]$$

$$+\mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it20}^{\psi_3\psi_4}(\gamma,\phi_i)\ell_{it10}^{\psi_1}(\gamma,\phi_i)\ell_{it10}^{\psi_2}(\gamma,\phi_i)]$$

$$+\mathbb{E}_{\tau_i}[\ell_{it02}(\theta,\alpha_i)\ell_{it01}^{\psi_1}(\gamma,\phi_i)\ell_{it01}^{\psi_2}(\gamma,\phi_i)\ell_{it01}^{\psi_3}(\gamma,\phi_i)\ell_{it01}^{\psi_4}(\gamma,\phi_i)].$$

Since by Assumption 4.3 each individual likelihood contribution in time period t is uniformly bounded across the parameters $(\theta,\alpha_i) \in \Theta \times \mathcal{J}$ and since the expectations are uniformly bounded across $(\gamma,\phi_i) \in \Theta \times \mathcal{J}$ and across time, the supremum over $\theta \in \Theta$ and $\beta_i \in B$ of each of the terms in (1)–(4) can again be shown to be bounded by repeated use of the Cauchy–Schwarz inequality. Analogous arguments can be used for mixed derivatives.

Next, we consider derivatives of $T^{-1}\sum_{t=1}^T \mathbb{E}_{\tau_i}[l_{it01}^2(\theta,\alpha_i;\tau_i)]$. By definition

$$\mathbb{E}_{\tau_i}[l_{it01}(\theta,\alpha_i;\tau_i)] = \mathbb{E}_{\tau_i}[\ell_{it01}(\theta,\alpha_i) - \mathbb{E}_{\tau_i}[\ell_{it01}(\theta,\alpha_i)]] = 0,$$

which implies

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

$$= \frac{1}{T}\sum_{t=1}^T (\mathbb{E}_{\tau_i}[\ell_{it01}^2(\theta,\alpha_i)] - \mathbb{E}_{\tau_i}[\ell_{it01}(\theta,\alpha_i)]^2).$$

To show, for example, that $\sup_{\beta_i \in B}\sup_{\theta \in \Theta} \partial_{\phi_i}^4 \mathbb{E}_{\tau_i}[l_{i01}^2(\theta,\alpha_i;\tau_i)] = O_p(1)$, we need to compute

$$\partial_{\phi_i}^4 \mathbb{E}_{\tau_i}[\ell_{it01}^2(\theta,\alpha_i)] = \mathbb{E}_{\tau_i}[\ell_{it01}^2(\theta,\alpha_i)\ell_{it04}(\gamma,\phi_i)] + 4\mathbb{E}_{\tau_i}[\ell_{it01}^2(\theta,\alpha_i)\ell_{it03}(\gamma,\phi_i)\ell_{it01}(\gamma,\phi_i)]$$

$$+6\mathbb{E}_{\tau_i}[\ell_{it01}^2(\theta,\alpha_i)\ell_{it02}(\gamma,\phi_i)\ell_{it01}^2(\gamma,\phi_i)] + \mathbb{E}_{\tau_i}[\ell_{it01}^2(\theta,\alpha_i)\ell_{it01}^4(\gamma,\phi_i)]$$

and

$$\partial_{\phi_i}^4 \mathbb{E}_{\tau_i}[\ell_{it01}(\theta, \alpha_i)]^2$$

$$= 6\mathbb{E}_{\tau_i}[\ell_{it01}(\theta, \alpha_i)\ell_{it01}(\gamma, \phi_i)]\left(\mathbb{E}_{\tau_i}[\ell_{it01}(\theta, \alpha_i)\ell_{it02}(\gamma, \phi_i)] + \mathbb{E}_{\tau_i}[\ell_{it01}(\theta, \alpha_i)\ell_{it01}^2(\gamma, \phi_i)]\right)$$

$$+ 2\mathbb{E}_{\tau_i}[\ell_{it01}(\theta, \alpha_i)](\mathbb{E}_{\tau_i}[\ell_{it01}(\theta, \alpha_i)\ell_{it03}(\gamma, \phi_i)] + 3\mathbb{E}_{\tau_i}[\ell_{it01}(\theta, \alpha_i)\ell_{it02}(\gamma, \phi_i)\ell_{it01}(\gamma, \phi_i)]$$

$$+ \mathbb{E}_{\tau_i}[\ell_{it01}(\theta, \alpha_i)\ell_{it01}^3(\gamma, \phi_i)]).$$

Analogous to previous arguments, using the Cauchy–Schwarz inequality and Assumption 4.3(i), we can show that

$$\frac{1}{T}\sum_{t=1}^T |\partial_{\phi_i}^4 \mathbb{E}_{\tau_i}[\ell_{it01}^2(\theta, \alpha_i)]| = O_\mathrm{p}(1)$$

and

$$\frac{1}{T}\sum_{t=1}^T |\partial_{\phi_i}^4 \mathbb{E}_{\tau_i}[\ell_{it01}(\theta, \alpha_i)]^2| = O_\mathrm{p}(1).$$

This implies

$$\frac{1}{T}\sum_{t=1}^T \sup_{\beta_i \in \Upsilon} \sup_{\theta \in \Theta} |\partial_{\phi_i}^4 \mathbb{E}_{\tau_i}[l_{it01}^2(\theta, \alpha_i)]| = O_\mathrm{p}(1),$$

which shows that

$$\sup_{\beta_i \in \Upsilon} \sup_{\theta \in \Theta} |\partial_{\phi_i}^4 \mathbb{E}_{\tau_i}[l_{i01}^2(\theta, \alpha_i; \tau_i)]| = O_\mathrm{p}(1).$$

The proof for derivatives with respect to components of γ and for the remaining mixed derivatives can be carried out analogously. □

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit: https://doi.org/10.1017/S0266466622000160

REFERENCES

Alvarez, J. & M. Arellano (2003) The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71, 1121–1159.

Arellano, M. (2003) *Panel Data Econometrics*. Oxford University Press.

Arellano, M. & S. Bonhomme (2009) Robust priors in nonlinear panel data models. *Econometrica* 77, 489–536.

Arellano, M. & J. Hahn (2007) Understanding bias in nonlinear panel models: Some recent developments. In R. Blundel, W. Newey, & T. Persson (eds.) *Advances in Economics and Econometrics: Ninth World Congress*, vol. 3, pp. 381–409. Cambridge University Press.

Arellano, M. & J. Hahn (2016) A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects. *Global Economic Review* 45, 251–274.

Chamberlain, G. (1982) Multivariate regression models for panel data. *Journal of Econometrics* 18, 5–46.

Chamberlain, G. (1984) Panel data. In Z. Griliches & M.D. Intriligator (eds.), *Handbook of Econometrics*, vol. II, pp. 1247–1318. Elsevier Science B.V.

Chamberlain, G. (2010) Binary response models for panel data: Identification and information. *Econometrica* 78, 159–168.

Dhaene, G. & K. Jochmans (2015) Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies* 82, 991–1030.

Dhaene, G. & Y. Sun (2021) Second-order corrected likelihood for nonlinear panel models with fixed effects. *Journal of Econometrics* 220, 227–252. Annals Issue: Celebrating 40 Years of Panel Data Analysis: Past, Present and Future.

Durrett, R. (1991) *Probability: Theory and Examples*. Wadsworth & Brooks/Cole.

Fernández-Val, I. (2009) Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics* 150, 71–85.

Fernández-Val, I. & M. Weidner (2016) *Individual and time effects in nonlinear panel models with large N, T*. *Journal of Econometrics* 192, 291–312.

Fernández-Val, I. & M. Weidner (2018) Fixed effects estimation of large-T panel data models. *Annual Review of Economics* 10, 109–138.

Ferrari, S.L., D.A. Botter, G.M. Cordeiro, & F. Cribari-Neto (1996) Second-and third-order bias reduction for one-parameter family models. *Statistics & Probability Letters* 30, 339–345.

Hahn, J. & G. Kuersteiner (2002) Asymptotically unbiased inference for a dynamic panel model with fixed effects when both *n* and *T* are large. *Econometrica* 70, 1639–1657.

Hahn, J. & G. Kuersteiner (2011) Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory* 27, 1152–1191.

Hahn, J., G. Kuersteiner, & W. Newey (2004) *Higher order efficiency of bias corrections. Unpublished Manuscript,* MIT.

Hahn, J. & W.K. Newey (2004) Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72, 1295–1319.

Newey, W.K. (1991) Uniform convergence in probability and stochastic equicontinuity. *Econometrica* 59, 1161–1167.

Neyman, J. & E.L. Scott (1948) Consistent estimation from partially consistent observations. *Econometrica* 16, 1–32.

Pace, L. & A. Salvan (2006) Adjustments of the profile likelihood from a new perspective. *Journal of Statistical Planning and Inference* 136, 3554–3564.

Rilstone, P., V.K. Srivastava, & A. Ullah (1996) The second-order bias and mean squared error of nonlinear estimators. *Journal of Econometrics* 75, 369–395.

Schumann, M., T.A. Severini, & G. Tripathi (2021a) Integrated likelihood based inference for nonlinear panel data models with unobserved effects. *Journal of Econometrics* 223, 73–95.

Schumann, M., T.A. Severini, & G. Tripathi (2021b) The role of information unbiasedness in panel data likelihoods. In progress.

Severini, T.A. (2000) *Likelihood Methods in Statistics*. Oxford University Press.

Shenton, L.R. & K.O. Bowman (1977) *Maximum Likelihood Estimation in Small Samples*. C. Griffin.

Varrette, S., P. Bouvry, H. Cartiaux, & F. Georgatos (2014) Management of an academic HPC cluster: The UL experience. In *Proceedings of the 2014 International Conference on High Performance Computing and Simulation (HPCS, 2014),* Bologna, Italy. IEEE.

Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. MIT Press.

Wooldridge, J.M. (2015) *Introductory eEconometrics: A Modern Approach*. Nelson Education.