Estimation and inference in dynamic nonlinear fixed effects panel data models by projection*

Andrew Adrian Yu Pua†

October 26, 2015

Abstract

I develop a bias reduction method for the estimators of structural parameters of a panel data model (whether linear or nonlinear) using a projection argument. A corrected score is calculated by projecting the score vector for the structural parameters onto the orthogonal complement of a space characterized by incidental parameter fluctuations. Assuming that the individual-specific effect could take on almost any finite value and that the densities for the data are correctly specified, I show that the asymptotic distribution of the structural parameters is normal and centered at zero under the same set of conditions considered in this literature. Furthermore, the construction of the projected score extends itself to situations where there are multiple fixed effects. Numerical experiments show that the finite sample performance of projected scores is at least as good or better than existing competitors when the number of time periods is around 3 or 4.

Keywords: Bias correction, dynamic nonlinear panel data models, score adjustments
JEL Classification: C13, C18, C23

1 Introduction

Neyman and Scott (1948) show that the method of maximum likelihood may fail to produce consistent and asymptotically efficient estimators when there are incidental parameters. Lancaster (2000) documents some of the developments after the publication of their paper.

*I thank Maurice Bun, Geert Dhaene, and Kees Jan van Garderen for their extensive comments on an earlier draft. I also thank participants of the 6th CFE-CSDA Conference, 2013 Amsterdam Workshop on Panel Data, 2013 KULeuven Workshop on Incidental Parameters, 2013 Netherlands Econometric Study group, and seminar participants at the Tinbergen Institute, the University of Bielefeld, and CORE. Finally, parts of the paper were written while I was being financially supported by the European Commission in the framework of the European Doctorate in Economics – Erasmus Mundus (EDEEM). All errors remain mine. Updated versions, if available, can be found in http://andrew-pua.ghost.io.

†Faculty of Economics and Business, University of Amsterdam, andrewypua@gmail.com

Roughly, these developments can be classified into two classes of solutions to the incidental parameter problem: solutions that exploit the structure of the model and solutions that involve orthogonal reparametrization. The latter has been explored more fully in Lancaster (2002) and Woutersen (2003; 2011). Most of the solutions that have been documented are called fixed-T solutions. If one would choose to use an asymptotic scheme where the number of cross-sectional units n grow large, leaving the number of time periods T fixed, then one has to use choose procedures that ensure that the estimating function is both functionally and stochastically independent of the incidental parameters.

Since incidental parameters in panel data models are represented as time-invariant parameters that appear in only a finite number of probability distributions, estimating these parameters induces finite sample bias in the time series dimension. This phenomenon allows us to reconsider the choice of asymptotic scheme. Research by Waterman (1993), Li, Lindsay, and Waterman (2003), and Hahn and Newey (2004) has paved the way for these large-T bias corrections. Arellano and Hahn (2007) primarily survey these developments for static panel data models with strictly exogenous regressors. They also document the three related ways of constructing these corrections – correcting the objective function, the moment equation, or the estimator itself. Although one can find consistent estimators of the common parameters, their asymptotic distributions are incorrectly centered. Under this asymptotic scheme, the nonzero center can be estimated when both the number of cross-sectional units and time periods grow at a particular rate (say $n/T \to c \in (0, \infty)$). As a result, one can construct an estimator with a correctly centered asymptotic distribution.

In this paper, I adjust the score or some suitably chosen moment function for the common parameter so that a consistent root of the adjusted score has a correctly centered asymptotic distribution. Furthermore, there are cases for which the adjustment produces a fixed-T consistent estimator. The score or some moment function is the most natural object to adjust because they are the starting points for proofs of consistency and asymptotic normality under regularity conditions. Depending on how one sees the multiple root problem, an issue with score-based adjustments is root selection.[1] In addition, when the common parameter is vector-valued, reconstructing a corrected objective function from the adjusted score or adjusted moment function may no longer be possible. Despite these issues, I discuss some of the advantages of using this score-based adjustment.

First, the computation of the large-T bias-corrected estimator typically requires the user to select an integer bandwidth whenever a model with some dynamics is being considered. This is true even for the case of a model with lagged dependent variables and strictly exogenous regressors (see for example, Bester and Hansen (2009) and Hahn and Kuersteiner (2011)) or a static binary choice model with predetermined regressors (see Fernandez-Val (2009)). Arellano and Hahn (2006) modify the objective function which also requires bandwidth selection. The proposed adjustment would not require bandwidth selection just

[1] Small, Wang, and Yang (2000) survey some existing methods for dealing with multiple roots.

like other score-based corrections (see for example, Woutersen (2003), Carro (2007), and Dhaene and Jochmans (2015b)). One can consider this as an improvement because score-based adjustments exploit the model structure fully in order to create the correction. As a result, finite sample performance may improve, especially in short panels. In a sense, one can "loosen" the use of the model structure as T becomes very large, so that choosing a bandwidth would suffice.

Second, the approach can accommodate multiple *individual-specific* fixed effects. Multiple fixed effects may arise when the thresholds in ordered choice models are individual-specific in addition to accounting for individual-specific effects in the linear predictor (see Bester and Hansen (2009) and Carro and Traferri (2012)). They also arise when a model explicitly allows for a vector of individual-specific effects. For example, Hausman and Pinkovskiy (2013) approximate a dynamic nonlinear model with general predetermined regressors and a scalar individual-specific effect by a Taylor series expansion around an estimator for the scalar individual-specific effect. They show that the transformed model is an affine function of a vector of fixed effects. The elements of this vector are the positive integer powers of the deviation of the scalar individual-specific effect from its estimator. Multiple fixed effects also arise when a model contains time dummies. I do not consider this case but Fernandez-Val and Weidner (2013) have recently proposed and justified the large-T bias corrections in this context.

Third, the approach can accommodate predetermined regressors aside from lagged dependent variables. The approach considered in this paper can accommodate predetermined regressors provided that the feedback process is specified to some degree. The feedback process can either be structural or be some flexible reduced form in the spirit of the Mundlak-Chamberlain device. The specification of the feedback process is partly a matter of interpretation. The Mundlak-Chamberlain device is a correlated random effects approach where the individual-specific fixed effect is usually expressed as a linear projection of the individual-specific fixed effect on the observable characteristics of the cross-sectional unit and a residual (see Mundlak (1978) and Chamberlain (1984)). As proposed by Wooldridge (2000) and applied by Moral-Benito (2013; 2014), the Mundlak-Chamberlain device can be used to flexibly specify the feedback process. In contrast to Wooldridge (2000), we do not specify reduced forms for the individual-specific fixed effect. Corrections that allow for general predetermined regressors without resorting to the device include work by Woutersen (2003), Fernandez-Val (2009), and Fernandez-Val and Weidner (2013).

I give details on the projection approach and its properties in Section 2. I also discuss some examples where analytical results are available. In Section 3, I present the results of two small-scale Monte Carlo simulations where I compare the projected score to the corrections proposed by Woutersen (2003), Carro (2007), Fernandez-Val (2009), and Hahn and Kuersteiner (2011). Other corrections that were not implemented include the corrections based on (i) modifying the likelihood (see Arellano and Hahn (2006) and Bartolucci et al.

(2014)) or integrating the likelihood (see Arellano and Bonhomme (2009) and De Bin, Sartori, and Severini (2015)) and (ii) simulation (see Kim and Sun (2009) and Dhaene and Jochmans (2015b)). I conclude in Section 4 and include a technical appendix for osme of the calculations and proofs.

2 The projection approach

2.1 Concept

Suppose we draw a random sample $\{y_i = (y_{i1}, \ldots, y_{iT}) : i = 1, \ldots, n\}$ from some known density $f(y_i; \theta_0, \alpha_{i0})$, where θ_0 is the true value for the common parameter and α_{i0} is the true value for the incidental parameter. Note that these parameters may be vector-valued but I assume that these are scalars for the purposes of illustration. Denote $\mathbb{E}[\cdot; \theta_0, \alpha_{i0}]$ to be the expectation at the true values of the parameters. Denote $\partial^k_{\alpha_i}$ to be the kth order partial derivative with respect to α_i.

To construct consistent estimators for θ_0 in the presence of unknown α_{i0} that have to be estimated, we need a concept that will quantify reduced sensitivity to perturbations of the true value of the incidental parameter, denoted by α_i', holding θ_0 fixed. This means that aside from searching for unbiased estimating functions $g(\theta, \alpha_i; y_i)$ that have zero expectation at the true value, i.e.

$$\mathbb{E}[g(\theta_0, \alpha_{i0}; y_i); \theta_0, \alpha_{i0}] = 0,$$

we have to further narrow the search to classes of estimating functions that satisfy either of the following conditions:

1. Global ancillarity, where the expectation of the estimating function does not depend on the perturbed value α_i':

$$\mathbb{E}\left[g(\theta_0, \alpha_{i0}; y_i); \theta_0, \alpha_i'\right] = 0, \ \forall \alpha_i' \neq \alpha_{i0}, \tag{1}$$

2. rth-order local $\mathbb{E}$-ancillarity, where the expectation of the estimating function does not depend on the perturbed value α_i' within some neighborhood of α_{i0}:

$$\partial^k_{\alpha_i'} \mathbb{E}\left[g(\theta_0, \alpha_{i0}; y_i); \theta_0, \alpha_i'\right]\Big|_{\alpha_i' = \alpha_{i0}} = 0, \quad \text{for } k = 1, \ldots, r \tag{2}$$

Moment functions satisfying (1) are difficult to construct. Bonhomme (2012) provides a theory that characterizes such moment functions using functional differencing, which is motivated by the theory of orthogonal projections. He also shows that fixed-T consistent estimation is possible in fully parametric and static and some dynamic panel data settings under some conditions on the distribution of the incidental parameters. Global ancillarity

is also equivalent to what Cox and Reid (1987) call global orthogonality. Tibshirani and Wasserman (1994) call this exact orthogonality in expectation. Woutersen (2011) calls this a zero-score property which holds not just at the true value α_{i0}. Therefore, a sample analog of the score will produce a consistent root regardless of the value plugged in for the incidental parameter.

A more attainable goal is to consider (2) so that (1) holds in a smaller region of the parameter space. To further motivate this condition, I expand, up to the second order, the density f in the left hand side of (1), i.e.,

$$\begin{aligned}
&\mathbb{E}\left[g\left(\theta_0,\alpha_{i0};y_i\right);\theta_0,\alpha_i'\right]\\
=&\int g\left(\theta_0,\alpha_{i0};y_i\right)f\left(y_i;\theta_0,\alpha_i'\right)dy_i\\
=&\int g\left(\theta_0,\alpha_{i0};y_i\right)f\left(y_i;\theta_0,\alpha_{i0}\right)dy_i+\int g\left(\theta_0,\alpha_{i0};y_i\right)\partial_{\alpha_i'}f\left(y_i;\theta_0,\alpha_i'\right)\Big|_{\alpha_i'=\alpha_{i0}}\left(\alpha_i'-\alpha_{i0}\right)dy_i\\
&+\frac{1}{2}\int g\left(\theta_0,\alpha_{i0};y_i\right)\partial_{\alpha_i'}^2f\left(y_i;\theta_0,\alpha_i'\right)\Big|_{\alpha_i'=\alpha_{i0}}\left(\alpha_i'-\alpha_{i0}\right)^2dy_i\\
=&\underbrace{\mathbb{E}\left[g\left(\theta_0,\alpha_{i0};y_i\right);\theta_0,\alpha_{i0}\right]}_{(a)}+\underbrace{\partial_{\alpha_i'}\mathbb{E}\left[g\left(\theta_0,\alpha_{i0};y_i\right);\theta_0,\alpha_i'\right]\Big|_{\alpha_i'=\alpha_{i0}}}_{(b)}\left(\alpha_i'-\alpha_{i0}\right)\\
&+\frac{1}{2}\partial_{\alpha_i'}^2\mathbb{E}\left[g\left(\theta_0,\alpha_{i0};y_i\right);\theta_0,\alpha_i'\right]\Big|_{\alpha_i'=\widetilde{\alpha}_i}\left(\alpha_i'-\alpha_{i0}\right)^2,
\end{aligned}$$

where $\widetilde{\alpha}_i$ is in between α_i' and α_{i0}. Since g is an unbiased estimating function, the term (a) in the preceding derivation is equal to zero. Under first-order local $\mathbb{E}$-ancillarity, the term (b) is also equal to zero. As a result, we have

$$\mathbb{E}\left[g\left(\theta_0,\alpha_{i0};y_i\right);\theta_0,\alpha_i'\right]=o\left(\alpha_i'-\alpha_{i0}\right).$$

Obviously, the extension to rth-order local $\mathbb{E}$-ancillarity will allow us to conclude that

$$\mathbb{E}\left[g\left(\theta_0,\alpha_{i0};y_i\right);\theta_0,\alpha_i'\right]=o\left(\alpha_i'-\alpha_{i0}\right)^r.$$

Notice that more and more smoothness would be required as one increases r.[2]

First-order local $\mathbb{E}$-ancillarity is what Cox and Reid (1987) call information orthogonality or local orthogonality when applied to the likelihood setting. They suggest finding a reparametrization so that θ and α_i are information orthogonal. They call the required transformation an orthogonal reparametrization, which means that, up to a certain order, estimating α_i will have minimal impact on consistently estimating θ. Lancaster (2002)

[2]Nonsmooth objective functions, especially those that arise in quantile regressions, are not covered by these ancillarity conditions. It is unclear how smoothing these objective functions will affect the bias-reducing properties of these ancillarity conditions.

and Woutersen (2011) derive orthogonal reparametrizations for common panel data models such as the static single index model with strictly exogenous regressors and the linear AR(1) dynamic panel data model. Unfortunately, finding an orthogonal reparametrization requires finding a solution (which may not exist) to a system of partial differential equations.

2.2 Implications

Instead of finding solutions to the system of partial differential equations and applying the reparametrization, we can determine how g should be restricted so that g will satisfy (2). Notice that rth-order local $\mathbb{E}$-ancillarity is equivalent to searching for g such that the following set of moment conditions will hold:

$$\mathbb{E}\left[g\left(\theta_0,\alpha_{i0};y_i\right)V_i^{(k)}\left(\theta_0,\alpha_{i0}\right)\right]=0,\ \text{for } k=1,\ldots,r, \tag{3}$$

where

$$V_i^{(k)}\left(\theta_0,\alpha_{i0}\right)=\frac{\partial_{\alpha_i}^k f\left(y_i;\theta_0,\alpha_{i0}\right)}{f\left(y_i;\theta_0,\alpha_{i0}\right)} \tag{4}$$

is the kth element of the so-called Bhattacharyya basis (see the pioneering works by Bhattacharyya (1946; 1947; 1948)). To show the equivalence, write the left hand side of (3) as

$$\begin{aligned}\mathbb{E}\left[g\left(\theta_0,\alpha_{i0};y_i\right)V_i^{(k)}\left(\theta_0,\alpha_{i0}\right)\right] &= \int g\left(\theta_0,\alpha_{i0};y_i\right)\partial_{\alpha_i'}^k f\left(y_i;\theta_0,\alpha_i'\right)\Big|_{\alpha_i'=\alpha_{i0}} dy_i \\ &= \partial_{\alpha_i'}^k\int g\left(\theta_0,\alpha_{i0};y_i\right)f\left(y_i;\theta_0,\alpha_i'\right)dy_i\Big|_{\alpha_i'=\alpha_{i0}} \\ &= \partial_{\alpha_i'}^k\mathbb{E}\left[g\left(\theta_0,\alpha_{i0};y_i\right);\theta_0,\alpha_i'\right]\Big|_{\alpha_i'=\alpha_{i0}},\end{aligned}$$

where the last expression is equal to zero by (2). Note that whenever an estimating function g satisfies rth-order local $\mathbb{E}$-ancillarity, it also satisfies kth-order local $\mathbb{E}$-ancillarity for all $k=1,\ldots,r-1$.

At this point, I will reduce notation by suppressing the arguments (θ_0,α_{i0}). I now show some of the consequences of (3) when $r=2$. First, note that

$$\mathbb{E}\left[\partial_{\alpha_{i0}}g\right] = \partial_{\alpha_{i0}}\mathbb{E}[g]-\mathbb{E}\left[gV_i^{(1)}\right]=0, \tag{5}$$

which follows from the requirement that g be an unbiased estimating function and (3) when $r=1$. Furthermore, another consequence of (3) when $r=2$ is

$$Cov\left(V_i^{(1)},\partial_{\alpha_{i0}}g\right)=\mathbb{E}\left[V_i^{(1)}\partial_{\alpha_{i0}}g\right]-\mathbb{E}\left[V_i^{(1)}\right]\mathbb{E}\left[\partial_{\alpha_{i0}}g\right]=\mathbb{E}\left[V_i^{(1)}\partial_{\alpha_{i0}}g\right]=0. \tag{6}$$

This zero covariance property follows from calculating the derivative of (3) with respect to α_{i0}:

$$\partial_{\alpha_{i0}}\mathbb{E}\left[gV_i^{(1)}\right]=\mathbb{E}\left[gV_i^{(2)}\right]-\mathbb{E}\left[V_i^{(1)}\partial_{\alpha_{i0}}g\right]. \tag{7}$$

Since g satisfies first-order local $\mathbb{E}$-ancillarity, the expression $\mathbb{E}\left[gV_i^{(1)}\right]$ on the left hand side is equal to zero. Since g satisfies second-order local $\mathbb{E}$-ancillarity, the first term in the right hand side of (7) is equal to zero. As a result, the covariance between $V_i^{(1)}$ and $\partial_{\alpha_{i0}}g$ is zero whenever g satisfies second-order local $\mathbb{E}$-ancillarity. Finally,

$$\mathbb{E}\left[\partial^2_{\alpha_{i0}}g\right] \;=\; \partial_{\alpha_{i0}}\mathbb{E}\left[\partial_{\alpha_{i0}}g\right]-\mathbb{E}\left[V_i^{(1)}\partial_{\alpha_{i0}}g\right]=0, \tag{8}$$

which follows from (5) and (6).

It is exactly this zero covariance property (6), along with the consequences of second-order local $\mathbb{E}$-ancillarity (5) and (8), that mimics the bias reduction that has already been developed in the literature. Estimator-based corrections in the spirit of Hahn and Newey (2004) and Hahn and Kuersteiner (2011) trace the source of the bias in the estimator to the $O\left(T^{-1}\right)$ bias in the unadjusted score or moment function. To illustrate how their work relates to the projected score, I reproduce their calculation of the bias of some moment function u_{it} for the common parameter θ. Note that v_{it} is the moment function for the incidental parameter α_i. In the context of a static panel data model, the bias of u_{it} is given by

$$\mathbb{E}[u_{it}(\theta,\widehat{\alpha_i})]=\frac{1}{T}\left\{\mathbb{E}\left[\partial_{\alpha_i}u_{it}\right]\beta_i+\mathbb{E}\left[\psi_{it}\partial_{\alpha_i}u_{it}\right]+\frac{1}{2}\mathbb{E}\left[\partial^2_{\alpha_i}u_{it}\right]\mathbb{E}\left[\psi^2_{it}\right]\right\}+o\left(T^{-1}\right),$$

where ψ_{it} and β_i are components of the higher-order asymptotic expansion for $\widehat{\alpha_i}$, i.e.,

$$\psi_{it}=-\mathbb{E}\left[\partial_{\alpha_i}v_{it}\right]^{-1}v_{it},\ \beta_i=-\mathbb{E}\left[\partial_{\alpha_i}v_{it}\right]^{-1}\left\{\mathbb{E}\left[\psi_{it}\partial_{\alpha_i}v_{it}\right]+\frac{1}{2}\mathbb{E}\left[\partial^2_{\alpha_i}v_{it}\right]\mathbb{E}\left[\psi^2_{it}\right]\right\}$$

Notice that if we chose a moment function u_{it} such that

$$\mathbb{E}\left[\partial_{\alpha_i}u_{it}\right]=0,\ \mathbb{E}\left[\psi_{it}\partial_{\alpha_i}u_{it}\right]=0,\ \mathbb{E}\left[\partial^2_{\alpha_i}u_{it}\right]=0,$$

the $O\left(T^{-1}\right)$ bias disappears. These three equations are exactly the consequences of first-order local $\mathbb{E}$-ancillarity, the zero-covariance property in (6), and second-order local $\mathbb{E}$-ancillarity, respectively. It is in this sense that starting from local $\mathbb{E}$-ancillarity may be more transparent and intuitive when considering bias corrections.

Let us now consider the case of dynamic nonlinear panel data models. In their motivation for their bias correction procedure, Hahn and Kuersteiner (2011) show that the

nonzero center of the asymptotic distribution of the uncorrected MLE is

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\mathbb{E}\left[\partial_{\alpha_i} v_{it}\right]}\mathbb{E}\left[\frac{1}{T}\left(\sum_{t=1}^{T} v_{it}\right)\left(\sum_{t=1}^{T}\partial_{\alpha_i} u_{it}\right)\right]-\frac{\mathbb{E}\left[\partial_{\alpha_i}^2 u_{it}\right]}{2\left(\mathbb{E}\left[\partial_{\alpha_i} v_{it}\right]\right)^2}\mathbb{E}\left[\frac{1}{T}\left(\sum_{t=1}^{T} v_{it}\right)^2\right].$$

Once again notice that if we choose a moment function for the common parameters u_{it} that satisfies second-order local $\mathbb{E}$-ancillarity, this nonzero center disappears.

In addition to the preceding discussion, the criterion of second-order local $\mathbb{E}$-ancillarity is also constructive because we can interpret (3) in Hilbert space terms, where the expectation operator is the inner product. We can think of (3) as finding g that is orthogonal to a linear subspace spanned by $\left(V_i^{(1)},\ldots,V_i^{(r)}\right)$. This linear subspace represents local effects of incidental parameter fluctuations. An analogous idea appears in linear regression settings so that we can interpret the desired estimating function g as a residual orthogonal to the explanatory variables $\left(V_i^{(1)},\ldots,V_i^{(r)}\right)$. This residual is called the rth-order projected score. In principle, one can construct the rth-order projected score but a lot of the benefits in terms of bias correction can already be reaped at the second order as seen in the preceding discussions.

2.3 Computation

Let us consider the situation where one has a complete specification of a likelihood for the data. For every $i=1,\ldots,n$, let $z_i=(y_{i0},y_{i1},\ldots,y_{iT},x_{i1},\ldots,x_{iT})$ be the data for the ith unit and $z=(z_1,\ldots,z_n)$ be the full data. Let $f(z_{it};\theta,\alpha_i)$ be the density of the data where $\theta\in\mathbb{R}^p$ and $\alpha_i\in\mathbb{R}^q$. Assume the cross-sectional units are independent of each other. The joint density of the observables is given by

$$f(z;\theta,\alpha)=\prod_{i=1}^{n} f(z_i;\theta,\alpha_i).$$

Note that the density $f(z_i;\theta,\alpha_i)$ is specified such that predetermined regressors can be accommodated. For example, if we let $x_i^t=(x_{i1},\ldots,x_{it})$ and $y_i^t=(y_{i0},y_{i1},\ldots,y_{it})$, we can write $f(z_i;\theta,\alpha_i)$ as

$$\begin{aligned} f(z_i;\theta,\alpha_i) &= f(y_{iT}|x_i^T,y_i^{T-1};\theta,\alpha_i)\times f(x_{iT}|y_i^{T-1},x_i^{T-1})\times\ldots\times f(y_{i2}|x_i^2,y_i^1;\theta,\alpha_i) \\ &\quad\times f(x_{i2}|y_i^1,x_{i1})\times f(y_{i1}|x_{i1},y_{i0};\theta,\alpha_i)\times f(y_{i0},x_{i1}) \end{aligned}$$

We usually specify parametric models for $f\left(y_{it}|x_i^t,y_i^{t-1};\theta,\alpha_i\right)$ and treat these models as structural. Flexible reduced forms can then be used to specify the feedback proceeses $f\left(x_{it}|y_i^{t-1},x_i^{t-1}\right)$. These flexible reduced forms can introduce further individual-specific fixed effects different from α_i. Examples can be found in Moral-Benito (2013; 2014). Note

that the distribution of the initial values $f(y_{i0}, x_{i1})$ can be specified or be left unspecified. If left unspecified, I condition on initial values.

The θ-score and α_i-score be can be written as

$$\begin{aligned} U_{i,0}(\theta, \alpha_i; z_i) &= \partial_\theta \log f(z_i; \theta, \alpha_i), \\ V_i(\theta, \alpha_i; z_i) &= \partial_{\alpha_i} \log f(z_i; \theta, \alpha_i). \end{aligned}$$

Observe that the α_i-score only uses the time-series observations for the ith cross-sectional unit and is a function of α_i and not of α_j for $j \neq i$.

When we set $k = 1$ in (4), $V_i^{(1)}$ coincides with the α_i-score so that $V_i^{(1)} = V_i$. The second-order terms $V_i^{(2)}$ can be written as

$$V_i^{(2)} = \partial_{\alpha_i^{\mathrm{T}}} V_i + V_i V_i^{\mathrm{T}}. \tag{9}$$

The preceding recurrence relation, which can be generalized to the rth order, is a consequence of

$$\partial_{\alpha_i^{\mathrm{T}}} V_i = \partial_{\alpha_i^{\mathrm{T}}} \left(\frac{\partial_{\alpha_i} f}{f} \right) = \frac{f \times \partial_{\alpha_i, \alpha_i^{\mathrm{T}}} f - \partial_{\alpha_i} f \times \partial_{\alpha_i^{\mathrm{T}}} f}{f^2} = \frac{\partial_{\alpha_i, \alpha_i^{\mathrm{T}}} f}{f} - \frac{\partial_{\alpha_i} f}{f} \frac{\partial_{\alpha_i^{\mathrm{T}}} f}{f} = V_i^{(2)} - V_i V_i^{\mathrm{T}},$$

which follows from the quotient rule for derivatives. Note that (9) is a recurrence relation because one can generate $V_i^{(r)}$ from $V_i^{(r-1)}$. Define the second-order extended information matrix as

$$M_{i,2} = \mathbb{E}\left[\begin{pmatrix} U_{i,0} \\ V_i \\ vec\left[V_i^{(2)}\right] \end{pmatrix} \begin{pmatrix} U_{i,0}^{\mathrm{T}} & V_i^{\mathrm{T}} & vec\left[V_i^{(2)}\right]^{\mathrm{T}} \end{pmatrix} \right] = \begin{pmatrix} M_{11,i} & M_{12,i} \\ M_{21,i} & M_{22,i} \end{pmatrix},$$

where the submatrices are defined as follows:

$$\begin{aligned} M_{11,i} &= \mathbb{E}\left[U_{i,0} U_{i,0}^{\mathrm{T}} \right], \\ M_{12,i} &= M_{21,i}^{\mathrm{T}} = \mathbb{E}\left[\begin{pmatrix} U_{i,0} V_i^{\mathrm{T}} & U_{i,0} vec\left[V_i^{(2)}\right]^{\mathrm{T}} \end{pmatrix} \right], \\ M_{22,i} &= \mathbb{E}\left[\begin{pmatrix} V_i \\ vec\left[V_i^{(2)}\right] \end{pmatrix} \begin{pmatrix} V_i^{\mathrm{T}} & vec\left[V_i^{(2)}\right]^{\mathrm{T}} \end{pmatrix} \right]. \end{aligned}$$

The second-order projected score and its information matrix for the ith unit could be expressed as

$$U_{i,2} = U_{i,0} - M_{12,i} \left(M_{22,i}\right)^{-} \begin{pmatrix} V_i \\ vec\left[V_i^{(2)}\right] \end{pmatrix}, \tag{10}$$

$$I_{i,2} = M_{11,i} - M_{12,i} \left(M_{22,i}\right)^{-} M_{21,i}. \tag{11}$$

where $\left(M_{22,i}\right)^{-}$ is the Moore-Penrose inverse of $M_{22,i}$. As discussed in the previous subsection, the second-order projected score is really the residual orthogonal to the linear subspace spanned by $\left(V_i^{(1)}, V_i^{(2)}\right)$. Thus, the second-order projected score $U_{i,2}$ makes the θ-score $U_{i,0}$ less sensitive to the presence of the incidental parameters. The second-order projected score and its associated information matrix for the full data can then be computed by summing up n components of the form (10) and (11).

As a result of all the preceding discussions, I present the following lemma and a more formal proof in the appendix.

Lemma 1. The second-order projected score $U_{i,2}$ is an unbiased estimating equation that satisfies second-order local $\mathbb{E}$-ancillarity (2).

In general, the projected score may depend on both θ and α_i. Thus, we have to substitute an estimator for α_i to form a plug-in projected score. The first-order projected score for the ith unit can be written as

$$U_{i,1} = U_{i,0} - \mathbb{E}\left(U_{i,0}V_i^{\mathrm{T}}\right)\left[\mathbb{E}\left(V_iV_i^{\mathrm{T}}\right)\right]^{-} V_i.$$

Solving $V_i = 0$ gives an estimator for α_i given θ, denoted by $\widehat{\alpha_i}(\theta)$. The plug-in first-order projected score $\sum_i \widehat{U_{i,1}}$ coincides with the profile score for θ. Dhaene and Jochmans (2015b) show that the panel Poisson model and panel exponential duration model have profile scores that have zero expectation. Therefore, the plug-in first-order projected score mimics the behavior of the profile score when applied to these models.

On the other hand, the second-order projected score is given by

$$\begin{aligned} U_{i,2} \;=\; & U_{i,0} - \mathbb{E}\left(U_{i,0}V_i^{\mathrm{T}}\right)\left[\mathbb{E}\left(V_iV_i^{\mathrm{T}}\right)\right]^{-} V_i \\ & -\mathbb{E}\left(U_{i,0}vec\left[V_i^{(2)}\right]^{\mathrm{T}}\right)\left[\mathbb{E}\left(vec\left[V_i^{(2)}\right]vec\left[V_i^{(2)}\right]^{\mathrm{T}}\right)\right]^{-} vec\left[V_i^{(2)}\right]. \end{aligned} \tag{12}$$

The next two propositions show that the plug-in second-order projected score matches the properties of existing bias corrections.

Proposition 1. *Assume that the conditions for (17), the conditions of Lemma 1, and the central limit theorems for* V_i , V_i^2, $\partial_{\alpha_i}U_{i,2}$, and $\partial_{\alpha_i}^2U_{i,2}$ *hold. Then,*

$$\mathbb{E}\left(\widehat{U_{i,2}}(\theta_0) - U_{i,2}\right) = \mathbb{E}\left(\widehat{U_{i,2}}(\theta_0) - U_{i,0}\right) = O\left(T^{-1}\right). \tag{13}$$

Since the second-order projected score $U_{i,2}$ satisfies second-order local $\mathbb{E}$-ancillarity by Lemma 1, we have (a) $\mathbb{E}\left[\partial_{\alpha_i}U_{i,2}\right] = 0$, (b) zero covariance between $\partial_{\alpha_i}U_{i,2}$ and V_i, and (c) $\mathbb{E}\left[\partial_{\alpha_i}^2U_{i,2}\right] = 0$. These three implications of second-order local $\mathbb{E}$-ancillarity are the most crucial reasons why $U_{i,2}$ already provides much of the bias reduction that existing methods aim to provide, just as sketched in the preceding subsection.

Assume that the system of equations implied by the plug-in second-order projected score has a solution in some neighborhood of the true value θ_0. We denote this solution by $\widehat{\theta^c}$ and it satisfies $\sum_i \widehat{U_{i,2}}\left(\widehat{\theta^c}\right) = 0$. This solution has an asymptotic distribution that is exactly the asymptotic distribution of the MLE.

Proposition 2. *Under the asymptotic scheme where $n, T \to \infty$, $n/T \to c \in (0, \infty)$, and $n/T^3 \to 0$, we have*

$$\sqrt{nT}\left(\widehat{\theta^c} - \theta_0\right) \xrightarrow{d} N\left(0, \left(\lim_{n,T\to\infty} \frac{1}{nT}\sum_{i=1}^{n} \mathbb{E}\left[U_{i,0}U_{i,0}^{\mathrm{T}}\right]\right)^{-1}\right). \tag{14}$$

2.4 Examples

Consider the following examples to demonstrate the calculations and some of the complications (and virtues) that may arise for the plug-in second-order projected score.

Example 1. (Linear AR(1) dynamic panel data model) Let $y_{it} = \alpha_i + \rho y_{i,t-1} + \varepsilon_{it}$ where $\varepsilon_{it} \sim \text{iid}\, N\left(0, \sigma^2\right)$ for all $i = 1, \ldots, N$ and $t = 1, 2$. Note that I do not restrict ρ so that y_{it} will be stationary. I condition on y_{i0} and assume that it is uncorrelated with future realizations of ε_{it}. The MLE for α_i given ρ and σ^2 is $\widehat{\alpha_i}\left(\rho, \sigma^2\right) = \bar{y}_i - \rho\,\bar{y}_{i,-1}$, where $\bar{y}_i = \left(y_{i1} + y_{i2}\right)/2$ and $\bar{y}_{i,-1} = \left(y_{i0} + y_{i1}\right)/2$. After calculating the second-order projected score for this case,[3] we substitute the MLE for $\widehat{\alpha_i}\left(\rho, \sigma^2\right)$ and obtain the following system of equations:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} \frac{\sigma^2 + \left(y_{i1} - y_{i0}\right)\left(y_{i2} - y_{i1} - \rho\left(y_{i1} - y_{i0}\right)\right)}{2\sigma^2} &= 0, \\
\frac{1}{n}\sum_{i=1}^{n} \frac{-2\sigma^2 + y_{i2} - y_{i1} - \rho\left(y_{i1} - y_{i0}\right)}{4\sigma^4} &= 0.
\end{aligned}$$

Eliminating σ^2 from the preceding system gives

$$\frac{2}{n}\sum_{i=1}^{n}\left(y_{i1} - y_{i0}\right)\left(y_{i2} - y_{i1} - \rho\left(y_{i1} - y_{i0}\right)\right) + \frac{1}{n}\sum_{i=1}^{n}\left(y_{i2} - y_{i1} - \rho\left(\left(y_{i1} - y_{i0}\right)\right)\right)^2 = 0.$$

Simplifying the equation above gives a quadratic equation in ρ of the form $A_n\rho^2 + B_n\rho + C_n = 0$ where

$$\begin{aligned}
A_n &= \frac{1}{n}\sum_{i=1}^{n}\left(y_{i1} - y_{i0}\right)^2, \\
B_n &= -\frac{2}{n}\sum_{i=1}^{n}\left[\left(y_{i2} - y_{i1}\right)\left(y_{i1} - y_{i0}\right) + \left(y_{i1} - y_{i0}\right)^2\right],
\end{aligned}$$

[3] Explicit calculations can be found in the appendix.

$$C_n = \frac{1}{n}\sum_{i=1}^{n}(y_{i2}-y_{i1})^2+\frac{2}{n}\sum_{i=1}^{n}(y_{i1}-y_{i0})(y_{i2}-y_{i1}).$$

I now show consistency of one of the roots of the quadratic equation. First, assume that $A_n \xrightarrow{p} A \neq 0$. Since $Cov(\epsilon_{i2}-\epsilon_{i1}, y_{i1}-y_{i0}) = -\sigma^2$, we must have $B_n \xrightarrow{p} -2\rho A + 2\sigma^2 - 2A$ and $C_n \xrightarrow{p} \rho^2 A - 2\rho\sigma^2 + 2\rho A$. By Slutsky's lemma, we also have $B_n^2 - 4A_nC_n \xrightarrow{p} 4\left(\sigma^2 - A\right)^2$. This means that the quadratic equation will always have real roots. As a result, we have

$$\widehat{\rho_n} = \frac{-B_n \pm \sqrt{B_n^2 - 4A_nC_n}}{2A_n} \xrightarrow{p} \rho - \left(\frac{\sigma^2}{A}-1\right) \pm \left|\frac{\sigma^2}{A}-1\right|,$$

where we either have $\widehat{\rho_n} \xrightarrow{p} \rho$ or $\widehat{\rho_n} \xrightarrow{p} \rho - 2\left(\sigma^2/A - 1\right)$. The estimator for $\widehat{\sigma_n^2}$ is given by

$$\widehat{\sigma_n^2} = -\frac{1}{n}\sum_{i=1}^{n}(y_{i1}-y_{i0})(y_{i2}-y_{i1}-\widehat{\rho_n}(y_{i1}-y_{i0})),$$

and will only be consistent if $\widehat{\rho_n}$ is consistent. Notice that the roots were obtained without resorting to an iterative procedure unlike the bias correction proposal by Bun and Carree (2005).

Which of the two roots should be chosen? To illustrate, consider the case where we have stationarity. Assume that y_{i0} is drawn from its stationary distribution where $\mathbb{E}(y_{i0}) = \alpha_i/(1-\rho)$ and $Var(y_{i0}) = \sigma^2/\left(1-\rho^2\right)$, where $|\rho| < 1$. In this case, $A_n \xrightarrow{p} 2\sigma^2/(1+\rho) \neq 0$. As a result, $\sigma^2/A - 1 < 0$. Thus, the consistent root is the smaller root of the quadratic equation. Now, consider the case where $\rho = 1$. Note that the large-n limit of A_n is such that $\sigma^2/A - 1 < 0$ since $y_{i1} - y_{i0} = \alpha_i + \varepsilon_{i1}$ implies that $\mathbb{E}(y_{i1}-y_{i0})^2 = \mathbb{E}(\alpha_i+\varepsilon_{i1})^2 = \mathbb{E}\left(\alpha_i^2\right) + \sigma^2 > \sigma^2$. As a result, the consistent root is still the smaller root of the quadratic equation.

Dhaene and Jochmans (2015a) extensively document the behavior of the resulting likelihood obtained after integrating the adjusted profile score. They have shown that the profile score has a bias that depends only on the common parameters and not on the incidental parameters. The adjusted profile score is then the difference between the profile score and its bias. They also propose a procedure to choose among the multiple critical points of the adjusted likelihood. Extensions of the model that allow for incidental trends can be found in Moon and Phillips (2004), where they also link the second-order projected score to their proposed moment condition.

Allowing for further lags should be straightforward for the projected score because a scalar pth order difference equation can be written as a vector first-order difference equation. Therefore, the quadratic equation derived for the AR(1) case is still going to be a quadratic equation with coefficients that are matrices. Allowing for regressors, whether strictly exogenous or predetermined, will not remove the multiple root problem and will

have to be examined on a case-by-case basis. ■

To explore the effect of including a predetermined regressor, consider an extension of the previous example that automatically allows for two individual-specific fixed effects.

Example 2. (Linear panel VAR(1) model) Consider the following structural model for the dynamics of two variables (y_{it}, x_{it}):

$$\begin{aligned} y_{it} &= \phi_{11} y_{i,t-1} + \phi_{12} x_{i,t-1} + \eta_{xi} + \varepsilon_{1it} \\ x_{it} &= \phi_{21} y_{i,t-1} + \phi_{22} x_{i,t-1} + \eta_{yi} + \varepsilon_{2it} \end{aligned}$$

where the idiosyncratic errors have the following distribution:

$$\begin{pmatrix} \varepsilon_{1it} \\ \varepsilon_{2it} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix} \right).$$

for $i = 1, \ldots, n$ and $t = 1, 2$. Assume that (i) Σ is positive definite, i.e., $\det(\Sigma) = \sigma_{11}\sigma_{22} - \sigma_{12}^2 > 0$, (ii) the initial observations (y_{i0}, x_{i0}) are available, and (iii) the distribution of the fixed effects and inital observations are left unspecified. The structural parameters are $\boldsymbol{\theta} = (\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \sigma_{11}, \sigma_{22}, \sigma_{12})$. The MLEs for η_{xi} and η_{yi} given the other parameters are

$$\begin{aligned} \widehat{\eta_{xi}} &= \frac{1}{2}(y_{i2} - y_{i1}(\phi_{11} - 1) - \phi_{11} y_{i0} - \phi_{12} x_{i0} - \phi_{12} x_{i1}) \\ \widehat{\eta_{yi}} &= \frac{1}{2}(x_{i2} - \phi_{21} y_{i0} - \phi_{21} y_{i1} - x_{i1}(\phi_{22} - 1) - \phi_{22} x_{i0}) \end{aligned}$$

The explicit calculations for the projected score can be carried out in *Mathematica*. The expectation of the plug-in first-order projected score for the ith cross-sectional unit has nonzero bias, i.e.

$$\mathbb{E}\left(\widehat{\mathbf{U}_{i,1}}\right) = \left(-\frac{1}{2}, 0, 0, -\frac{1}{2}, -\frac{\sigma_{22}}{2\det(\Sigma)}, -\frac{\sigma_{11}}{2\det(\Sigma)}, -\frac{\sigma_{12}}{\det(\Sigma)} \right).$$

Notice that this nonzero bias does not depend on η_{xi} and η_{yi}. As a result, this fits into Case 2 of Dhaene and Jochmans (2015b), where the profile score has expectation free of the incidental parameters. Similarly, calculations in *Mathematica* show that the expectation of the plug-in second-order projected score for the ith cross-sectional unit has zero bias. ■

Next, I consider a nonlinear model where the score of some conditional likelihood for the model is an unbiased estimating equation.

Example 3. (Static logit model with strictly exogenous regressors) Suppose $y_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2} \sim$ Bernoulli(p_{it}) with probability of success $p_{it} = \mathbb{E}(y_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}) = F(\alpha_i + \mathbf{x}_{it}^{\mathrm{T}}\boldsymbol{\beta})$ for $i = 1, \ldots, n$ and $t = 1, 2$. Assume that F is the logistic cdf. For $j = 0, 1, 2$, define $N_j = \{i : y_{i1} + y_{i2} = j\}$.

The MLE for α_i for fixed $\boldsymbol{\beta}$ is given by $\widehat{\alpha_i} = -(\mathbf{x}_{i1} + \mathbf{x}_{i2})^{\mathrm{T}} \boldsymbol{\beta}/2$ if $i \in N_1$; $\widehat{\alpha_i} \to -\infty$ if $i \in N_0$; $\widehat{\alpha_i} \to \infty$ if $i \in N_2$. Following the computations in the appendix, the plug-in second-order projected score using all i can be computed as

$$\begin{aligned}\sum_{i=1}^{n} U_2^i &= \sum_{i \in N_0} y_{i2} + \sum_{i \in N_2} (y_{i2} - 1) + \sum_{i \in N_1} (\mathbf{x}_{i2} - \mathbf{x}_{i1})^{\mathrm{T}} \left[y_{i2} - \frac{1}{1 + e^{(\mathbf{x}_{i2} - \mathbf{x}_{i1})^{\mathrm{T}} \boldsymbol{\beta}}} \right] \\ &= \sum_{i \in N_1} (\mathbf{x}_{i2} - \mathbf{x}_{i1})^{\mathrm{T}} \left[y_{i2} - \frac{1}{1 + e^{(\mathbf{x}_{i2} - \mathbf{x}_{i1})^{\mathrm{T}} \boldsymbol{\beta}}} \right].\end{aligned}$$

Note that the individuals in N_0 and N_2 have zero contribution to the plug-in second-order projected score. Although the above expression is monotonically decreasing in β, there is no closed-form solution to the above estimating equation. Despite this, the plug-in second-order projected score can be shown that this coincides with score of the conditional likelihood formed from the units for which $y_{i1} + y_{i2} = 1$. Since Chamberlain (1980) shows that the conditional MLE is $\sqrt{n}$-consistent, the same goes for the root of the plug-in second-order projected score.

Arellano and Bonhomme (2009) derive a bias-reducing prior for this model for general T that removes the $O\left(T^{-1}\right)$ bias. Their Monte Carlo simulations include an estimator where the adjustment was iterated. The simulations indicate that the iterated adjustment will mimic the properties of the conditional score when n is fixed and T increased to around 20. In contrast, Dhaene and Jochmans (2015b), who also consider the case of $T = 2$, show that the conditional score can be obtained either by an infinite-order profile score adjustment or by rescaling the profile score by the total number of movers. It is unclear whether rescaling will extend to the case where $T > 2$. ■

3 Simulations

In this section, I show that the finite sample performance of the plug-in second-order projected score is as good as or sometimes better than some existing competitors. I focus on panels with a very small value of T for the following reasons. Panels obtained from developing countries or panels formed from small-scale experiments usually have single-digit T. In practice, applied researchers will also use a subset of the data, especially when there are structural breaks in the time series or when the data are unbalanced. Therefore, it seems appropriate to choose small values of T to gauge finite sample performance.

I implement the projected score method and other alternatives using *Mathematica*.[4] *Mathematica* allows us to calculate the symbolic representation of the projected score and to compute the roots using the `FindRoot` command. Thus, the user only needs to specify the likelihood function and modify the code for the situation he considers without recoding

[4]All *Mathematica* notebooks and `R` code are available upon request.

the actual expressions of the corrections.[5] Furthermore, the calculations become much more compact and organized. I use two starting points, namely, the MLEs for the pooled and fixed effects model, for the root-finding algorithm. I use the software `R` to generate the data for the Monte Carlo experiments and to compute the MLEs (using the routine `glm`) under the pooled and fixed effects model.[6] The draws for the individual-specific fixed effects α_i are fixed across 5000 replications.

The implementation exploits the comparative advantages of both `R` and *Mathematica*. `R` can be used to generate samples from a user-specified data generating process and to perform routine estimation procedures, while *Mathematica* can be used to symbolically calculate the adjusted score and find its roots. The coding style in the *Mathematica* notebook allows any end user to do the following:

1. Specify either an objective function or an estimating function based on some parametric model.

2. Use the built-in commands for differentiation and calculation of expectations to produce symbolic representations of the adjustment found in (12) .

3. Import data and estimation results. The data and estimation results can come from any statistical software capable of exporting its outputs to a text file.

4. Use the programmed functions to generate empirical counterparts of the symbolic representations, to calculate roots and produce output for diagnostics, and to generate routine estimation results such as standard errors.

The coding style almost creates the feeling of a built-in package which may attract more users. But the user only has to change the parametric model in the *Mathematica* notebook whenever the user contemplates changes in the model.

To construct the plug-in second-order projected score, I compute the projected score as discussed in (12) and use an estimator for α_i. Rather than recompute $\widehat{\alpha_i}(\theta)$ at every iteration of the root-finding algorithm, I use a linear approximation of $\widehat{\alpha_i}(\theta)$ around $\widehat{\alpha_i}$ suggested by Bellio and Sartori (2003), i.e.,

$$\widehat{\alpha_i}(\theta) = \widehat{\alpha_i} + j^{-1}_{\alpha_i\alpha_i}\left(\widehat{\theta}, \widehat{\alpha_i}\right) j_{\alpha_i\theta}\left(\widehat{\theta}, \widehat{\alpha_i}\right)\left(\widehat{\theta} - \theta\right),$$

where $j_{\alpha_i\alpha_i}$ and $j_{\alpha_i\theta}$ are the corresponding (α_i, α_i) and (α_i, θ) blocks of the observed information matrix

$$j(\theta, \alpha_i) = \begin{bmatrix} j_{\theta\theta} & j_{\theta\alpha_i} \\ j_{\alpha_i\theta} & j_{\alpha_i\alpha_i} \end{bmatrix},$$

[5] Coding the actual expressions would take an inordinate number of lines of code and would only be valid for a specific model.

[6] Whenever the MLE does not exist, I take notice of this and I increase the number of replications so that I could attain the target of 5000 replications.

respectively. Other alternatives may be possible, for instance, using penalized likelihood estimator proposed by Firth (1993) and Kosmidis and Firth (2010) or the EM-based estimator proposed by Chen (2014). The idea behind these estimators is to improve the quality and stability of the plug-in values for α_i. These alternatives may be helpful in models where the plug-in values for α_i are either extreme or even undefined.

The first data generating process I consider is the static probit model. I use the following design adapted from Fernandez-Val (2009) with some modifications. The original design included a stationary AR(1) model with a linear time trend for the exogenous regressor x_{it}. Omitting this feature leads to the following modified specification:

$$
\begin{aligned}
&y_{it}|x_{i1},\ldots,x_{iT},\alpha_i \sim Ber\left(p_{it}\right), \;\; p_{it} = \Phi\left(\alpha_i + \beta_0 x_{it}\right),\\
&x_{it} \sim \text{iid } N\left(0,1\right), \alpha_i \sim \text{iid } N\left(0,1\right), x_{it} \perp \alpha_i,\\
&n = 125, T = 4, \beta_0 = 0.5,
\end{aligned}
$$

where $\Phi(\cdot)$ is the standard normal CDF. An important thing to note is that the regressor is already independent of the fixed effects. I choose this design because I stripped it down to the simplest elements. I already explored the static logit case in an example found in the previous section.

I also compare the performance of the projected score to the uncorrected MLE, the corrected estimator by Fernandez-Val (2009), and the score corrections by Carro (2007) and Woutersen (2003). Table 1 contains simulation results for the static probit model based on 5000 replications. The results indicate good finite sample performance of the projected score relative to all the other corrections. The Monte Carlo estimate of the bias is almost reduced by 90% relative to the uncorrected MLE. As a result, taking higher-order projections may not be needed as the gains will be marginal relative to computational cost. Furthermore, the standard deviation of the estimator obtained from the projected score is comparable to the standard deviation of the other estimators. The results clearly indicate that score-based corrections may be preferable in terms of RMSE. Although the number of nonconvergent cases is very small relative to the number of replications, I recommend obtaining a log of the iterations produced by the root-finding algorithm when implementing score-based corrections.

The second data generating process is the first-order dynamic logit model. Once more, I adapt the design from Fernandez-Val (2009) with some modifications.

$$
\begin{aligned}
&y_{it}|y_{i,t-1},\ldots,y_{i0},x_{i0},x_{i1},\ldots,x_{iT},\alpha_i \sim Ber\left(p_{it}\right), \;\; p_{it} = F\left(\alpha_i + \rho_0 y_{i,t-1} + \beta_0 x_{it}\right),\\
&y_{i0}|x_{i0},x_{i1},\ldots,x_{iT},\alpha_i \sim Ber\left(p_{i0}\right), \;\; p_{i0} = F\left(\alpha_i + \beta_0 x_{i0}\right),\\
&x_{it} \sim \text{iid } L\left(0,1\right), \alpha_i \sim \text{iid } L\left(0,1\right), x_{it} \perp \alpha_i,\\
&n = 125, T = 3, \beta_0 = 1, \rho_0 = 0.5.
\end{aligned}
$$

Table 1: Finite sample performance of estimators of β_0

Estimator	Mean bias	Median bias	Standard deviation	Median AD	RMSE
Uncorrected MLE	0.210	0.203	0.135	0.089	0.723
Fernandez-Val (2009)	0.162	0.156	0.123	0.081	0.674
Woutersen (2003)	0.069	0.064	0.099	0.066	0.577
(12 cases nonconvergent)					
Carro (2007)	0.071	0.066	0.100	0.066	0.580
(13 cases nonconvergent)					
Projected score	0.030	0.025	0.095	0.063	0.538

Note: True value of β_0 is equal to 0.5. Results are based on 5000 replications.

In this design, $F(\cdot)$ is the logistic CDF and $L(0,1)$ is the logistic distribution with mean 0 and scale 1. The original design assumes that $x_{it} \sim N\left(0, \pi^2/3\right)$ and the individual-specific fixed effects were generated as an average of the four oldest values of x_{it}. I choose to use $L\left(0,1\right)$ because it is quite similar to $N\left(0, \pi^2/3\right)$ but with heavier tails. I condition on y_{i0} instead of using the information from the distribution $y_{i0}|x_{i0}, x_{i1}, \ldots, x_{iT}$ in the likelihood function. For this model, the alternatives are the fixed -T consistent estimator proposed by Honoré and Kyriazidou (2000), the corrected estimators by Fernandez-Val (2009) and Hahn and Kuersteiner (2011), and the score-based corrections by Carro (2007) and Woutersen (2003).

Recall that Hahn and Kuersteiner (2011) obtain a characterization of the nonzero center of the asymptotic distribution of the MLE as discussed in Example **??**. Estimator-based corrections will have to rely on an estimator of this nonzero center. This nonzero center depends on the cross-spectrum of the α_i-score and the derivative of the θ-score with respect to α_i at the zero frequency and the spectrum of the α_i-score at the zero frequency. Since the cross-spectrum and spectrum are infinite sums of cross-covariances and covariances, respectively, a feasible prcoedure would require some lag truncation. As a result, we would require an integer bandwidth of lower order than $T^{1/2}$ for trimming purposes and for the asymptotic theory to hold. Since $T = 3$, I set the bandwidth at values 0, 1, and 2.

In contrast, Honoré and Kyriazidou (2000) propose an estimator based on the maximizer of a likelihood conditioned on the subset of observations for which $\{x_{i2} = x_{i3}\}$. Since this set is a zero probability event given the DGP, a kernel with a corresponding bandwidth is used to give higher weight to observations where x_{i2} is close to x_{i3} and give lower weight to observations otherwise. I use a standard normal kernel for this purpose. Furthermore, I use the optimal bandwidth derived by Honoré and Kyriazidou (2000) which is a constant multiple of $T^{-1/5}$. I set this constant to values 1, 8, and 64 just as Honoré and Kyriazidou (2000) do so in their own simulations.

The Monte Carlo results in Table 2 indicate that score-based corrections are performing quite well relative to estimator-based corrections for the design I consider. The bias of the

Table 2: Finite sample performance of estimators of β_0 and ρ_0

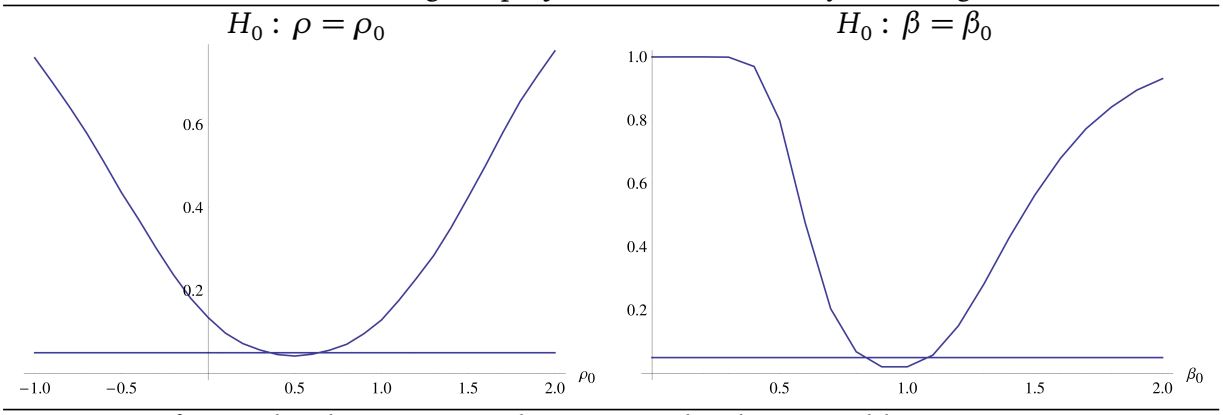
Estimator	Mean bias		Median bias		Standard deviation		Median AD		RMSE	
	ρ	β	ρ	β	ρ	β	ρ	β	ρ	β
Uncorrected MLE	-2.660	0.867	-2.588	0.792	0.902	0.457	0.558	0.272	2.341	1.921
Hahn and Kuersteiner (2011)										
Bandwidth 0	-0.552	-0.759	-1.066	-0.301	3.049	2.477	0.606	0.371	3.049	2.488
Bandwidth 1	-0.341	-0.028	-0.723	0.242	2.870	2.019	0.689	0.390	2.874	2.240
Bandwidth 2	-1.957	0.865	-1.870	0.783	0.897	0.485	0.538	0.281	1.711	1.927
Fernandez-Val (2009)										
Bandwidth 0	-1.994	0.225	-1.984	0.232	0.555	0.213	0.363	0.099	1.593	1.243
Bandwidth 1	-1.807	0.217	-1.795	0.226	0.554	0.212	0.363	0.096	1.419	1.235
Bandwidth 2	-1.948	0.211	-1.940	0.220	0.552	0.212	0.359	0.095	1.550	1.230
Honoré and Kyriazidou (2000)										
Bandwidth 1			0.268	0.550			1.771	0.885		
Bandwidth 8			-0.049	0.126			0.561	0.265		
Bandwidth 64			-0.059	0.131			0.541	0.250		
Woutersen (2003)	-0.183	-0.047	-0.181	-0.052	0.348	0.124	0.236	0.080	0.471	0.961
Carro (2007)	-0.505	-0.047	-0.506	-0.056	0.329	0.121	0.219	0.080	0.329	0.961
(1 case nonconvergent)										
Projected score	0.011	0.048	0.004	0.023	1.170	0.402	0.366	0.136	1.277	1.122
(24 cases nonconvergent)										

Note: The true values are given by $\rho_0 = 0.5$ and $\beta_0 = 1$. Results are based on 5000 replications.

root of the projected score is almost eliminated for both coefficients of the linear predictor. In contrast, the other score-based estimators are having problems eliminating the bias in the autoregressive coefficient. There seems to be a point at which a higher bandwidth will not improve finite sample performance of estimator-based corrections. In fact, the estimator-based correction by Hahn and Kuersteiner (2011) almost has the same performance as uncorrected MLE when the bandwidth is equal to 2. Furthermore, the dispersion of the corrected estimators is less than half that of the uncorrected MLE with the exception of the correction by Hahn and Kuersteiner (2011). The dispersion of the root of the projected score is more in line with that of the uncorrected MLE.

I also present two power curves in Table 3 for the projected score in the dynamic logit model. I do not present the results for the competing procedures because the estimated biases are large relative to the estimated standard deviation. The rejection probability of the test $\rho = 0.5$ is almost 5% while that of the test $\beta = 1$ is about 2%. The power curves are steeper for the coefficient of the exogenous regressor than for the autoregressive coefficient.

Table 3: Inference using the projected score for the dynamic logit model



Note: Significance level set at 5% and represented as horizontal line

It is clear from the Monte Carlo results that the projected score is an extremely attractive alternative to some of the competing bias-reduction procedures. The biggest downside is the computational time. For the designs considered, setting up of the projected score, the calculation of the root, and the standard error calculations took about 2 to 5 minutes for every replication on a laptop with 8 GB memory and an i7-processor. Even if we exploit parallel processing, the memory requirement is almost too great for all cores to be used all at once, especially when conducting Monte Carlo simulations. The reason for the high memory requirement is in the nature of the correction – a symbolic representation is created and the data are substituted into this representation. Despite these issues, the implementation is very straightforward and would not require us to program new procedures every time we make changes to the model.

4 Concluding remarks

This paper develops a bias reduction method for the estimators of common parameters of a linear or nonlinear panel data model with individual-specific fixed effects. The past decades saw a spur of research on bias reduction methods. It is easier to see what these methods have in common by considering what is called the projected score. This projected score is calculated by projecting the score vector for the common parameters onto the orthogonal complement of a space characterized by incidental parameter fluctuations.

I show that projected scores reduce the asymptotic bias of the estimators of common parameters in panel data models. Although the projected score has been introduced two decades ago, its widespread use has been hindered by computational issues. Relative to other bias reduction procedures, computation (in terms of processor time and memory) may be prohibitive but programming is less error-prone and more intuitive. I hope that this will encourage applied researchers to use the projected score. Monte Carlo simulations indicate that the bias-reducing properties of the projected score already take effect even for very small sample sizes usually encountered when panel data models are estimated for subsamples. Finally, the applied researcher need not choose a bandwidth anymore.

Future work on practical aspects include extensions to nonsmooth functions arising, say, in quantile regression. In addition, the projection idea has to be modified when one wants to extend to nonlikelihood settings and when one wants to include time effects. I intend to pursue these extensions in the future.

References

Arellano, M and S Bonhomme (2009). Robust Priors in Nonlinear Panel Data Models. *Econometrica* **77**.(2), 489–536.

Arellano, M and J Hahn (2006). *A Likelihood-Based Approximate Solution To The Incidental Parameter Problem In Dynamic Nonlinear Models With Multiple Effects*. Working Papers. CEMFI. `http://ideas.repec.org/p/cmf/wpaper/wp2006_0613.html`.

Arellano, M and J Hahn (2007). "Understanding Bias in Nonlinear Panel Models: Some Recent Developments". In: *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*. Ed. by R Blundell, W Newey, and T Persson. Vol. 3. Cambridge University Press. Chap. 12, pp.381–409.

Bartolucci, F, R Bellio, A Salvan, and N Sartori (2014). *Modified Profile Likelihood for Fixed-Effects Panel Data Models*. Tech. rep. Forthcoming in Econometric Reviews.

Bellio, R and N Sartori (2003). Extending conditional likelihood in models for stratified binary data. *Statistical Methods and Applications* **12.**(2), 121–132.

Bester, CA and C Hansen (2009). A penalty function approach to bias reduction in nonlinear panel models with fixed effects. *Journal of Business and Economic Statistics* **27**.(2), 131–148.

Bhattacharyya, A (1946). On Some Analogues of the Amount of Information and Their Use in Statistical Estimation. *Sankhyā: The Indian Journal of Statistics (1933-1960)* **8**.(1), 1–14.

Bhattacharyya, A (1947). On Some Analogues of the Amount of Information and Their Use in Statistical Estimation (Contd.) *Sankhyā: The Indian Journal of Statistics (1933-1960)* **8**.(3), 201–218.

Bhattacharyya, A (1948). On Some Analogues of the Amount of Information and Their Use in Statistical Estimation (Concluded). *Sankhyā: The Indian Journal of Statistics (1933-1960)* **8**.(4), 315–328.

Bonhomme, S (2012). Functional Differencing. *Econometrica* **80**.(4), 1337–1385.

Bun, MJG and MA Carree (2005). Bias-Corrected Estimation in Dynamic Panel Data Models. *Journal of Business & Economic Statistics* **23**.(2), 200–210.

Carro, JM (2007). Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics* **140**.(2), 503–528.

Carro, JM and A Traferri (2012). State Dependence and Heterogeneity in Health Using a Bias-corrected Fixed Effects Estimator. *Journal of Applied Econometrics*.

Chamberlain, G (1980). Analysis of Covariance with Qualitative Data. *Review of Economic Studies* **47**.(1), 225–38.

Chamberlain, G (1984). "Panel data". In: *Handbook of Econometrics*. Ed. by Z Griliches and MD Intriligator. Vol. 2. Handbook of Econometrics. Elsevier. Chap. 22, pp.1247–1318.

Chen, M (2014). *Estimation of Nonlinear Panel Models with Multiple Unobserved Effects*. Tech. rep. `http://blogs.bu.edu/mlchen/files/2014/11/JMP-Nov15th-non-url.pdf`.

Cox, DR and N Reid (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)* **49**.(1), 1–39.

De Bin, R, N Sartori, and TA Severini (2015). Integrated likelihoods in models with stratum nuisance parameters. *Electronic Journal of Statistics* **9**, 1474–1491.

Dhaene, G and K Jochmans (2015a). *Likelihood inference in an autoregression with fixed effects*. Tech. rep. Forthcoming in Econometric Theory.

Dhaene, G and K Jochmans (2015b). *Profile-score adjustments for nonlinear fixed-effect models*. Tech. rep. Katholieke Universiteit Leuven.

Fernandez-Val, I and M Weidner (2013). Individual and Time Effects in Nonlinear Panel Models with Large N, T. *ArXiv e-prints*. arXiv: `1311.7065 [stat.ME]`.

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

Firth, D (1993). Bias Reduction of Maximum Likelihood Estimates. *Biometrika* **80**.(1),

Hahn, J and G Kuersteiner (2011). Bias Reduction for Dynamic Nonlinear Panel Models with Fixed Effects. *Econometric Theory* **27**.(06), 1152–1191.

Hahn, J and W Newey (2004). Jackknife and Analytical Bias Reduction for Nonlinear Panel Models. *Econometrica* **72**.(4), 1295–1319.

Hausman, JA and ML Pinkovskiy (2013). "A Nonlinear Least Squares Approach to Estimating Fixed Effects Panel Data Models with Lagged Dependent Variables, with Applications to the Incidental Parameters Problem".

Honoré, BE and E Kyriazidou (2000). Panel Data Discrete Choice Models with Lagged Dependent Variables. *Econometrica* **68**.(4), 839–874.

Kim, MS and Y Sun (2009). "k-step bootstrap bias correction for fixed effects estimators in nonlinear panel models".

Kosmidis, I and D Firth (2010). A generic algorithm for reducing bias in parametric estimation. *Electronic Journal of Statistics* **4**, 1097–1112.

Lancaster, T (2000). The incidental parameter problem since 1948. *Journal of Econometrics* **95**.(2), 391–413.

Lancaster, T (2002). Orthogonal Parameters and Panel Data. *Review of Economic Studies* **69**.(3), 647–66.

Li, H, BG Lindsay, and RP Waterman (2003). Efficiency of projected score methods in rectangular array asymptotics. *Journal of the Royal Statistical Society B* **65**.(1), 191–208.

Moon, HR and PCB Phillips (2004). GMM Estimation of Autoregressive Roots Near Unity with Panel Data. *Econometrica* **72**.(2), 467–522.

Moral-Benito, E (2013). Likelihood-Based Estimation of Dynamic Panels With Predetermined Regressors. *Journal of Business & Economic Statistics* **31**.(4), 451–472.

Moral-Benito, E (2014). GROWTH EMPIRICS IN PANEL DATA UNDER MODEL UNCERTAINTY AND WEAK EXOGENEITY. *Journal of Applied Econometrics*.

Mundlak, Y (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica* **46**.(1), 69–85.

Neyman, J and EL Scott (1948). Consistent Estimates Based on Partially Consistent Observations. *Econometrica* **16**.(1), 1–32.

Small, CG, J Wang, and Z Yang (2000). Eliminating multiple root problems in estimation (with discussion). *Statistical Science* **15**.(4), 313–341.

Tibshirani, R and L Wasserman (1994). Some aspects of the reparametrization of statistical models. *Canadian Journal of Statistics* **22**.(1), 163–173.

Waterman, RP (1993). "Projective Score Methods". PhD thesis. Pennsylvannia State University.

Wooldridge, JM (2000). A framework for estimating dynamic, unobserved effects panel data models with possible feedback to future explanatory variables. *Economics Letters* **68**.(3), 245–250.

Woutersen, T (2003). “Robustness against incidental parameters”.
Woutersen, T (2011). “Consistent estimation and orthogonality”. In: *Missing Data Methods: Cross-sectional Methods and Applications*. Vol. 27A. Advances in Econometrics. Emerald Group Publishing Limited, pp.155–178.

A Projected score calculations

Projected score for the AR(1) linear dynamic panel data model

The model specification is as follows:

$$\begin{aligned} Y_{i,t-1} &= \{y_{i0}, y_{i1}, \ldots, y_{i,t-1}\}, \\ y_{it}|Y_{i,t-1} &\sim \text{iid}\, N(\alpha_i + \rho y_{i,t-1}, \sigma^2),\ i = 1, \ldots, n; t = 1 \ldots, T \end{aligned} \tag{15}$$

Assume y_{i0} is available and we condition on y_{i0}. Let $u_{it} = y_{it} - \alpha_i - \rho y_{i,t-1}$. The scores for the common parameters ρ and σ^2 and the incidental parameter α_i are given by:

$$\begin{aligned} U^{\rho}_{i,0} &= \frac{1}{\sigma^2}\sum_{t=1}^{T} u_{it} y_{i,t-1}, \\ U^{\sigma^2}_{i,0} &= -\frac{T}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T} u_{it}^2, \\ V^{(1)}_{i} &= \frac{1}{\sigma^2}\sum_{t=1}^{T} u_{it}. \end{aligned}$$

To calculate the second-order projected score, we need the following elements:

$$V^{(2)}_{i} = -\frac{T}{\sigma^2} + \left[\frac{1}{\sigma^2}\sum_{t=1}^{T} u_{it}\right]^2,$$

$$\begin{aligned} E[(V^{(1)}_{i})^2] &= \frac{1}{\sigma^4} E\left[\sum_{t=1}^{T} u_{it}^2 + 2\sum_{t=2}\sum_{s<t} u_{is}u_{it}\right] \\ &= \frac{1}{\sigma^4}\left[\sum_{t=1}^{T}\sigma^2 + 2\sum_{t=2}\sum_{s<t} E(u_{it}u_{is})\right] \\ &= \frac{T}{\sigma^2}. \end{aligned}$$

Note that only the second moments of (15) are used for the above calculation.

$$
\begin{aligned}
E(V_i^{(1)}V_i^{(2)}) &= -\frac{T}{\sigma^4}\sum_{t=1}^{T}E(u_{it}) + \frac{1}{\sigma^6}E\left(\sum_{t=1}^{T}u_{it}\right)^3 \\
&= \frac{1}{\sigma^6}E\left[\left(\sum_{t=1}^{T}u_{it}\right)\left(\sum_{t=1}^{T}u_{it}^2 + 2\sum_{t=2}^{T}\sum_{s<t}u_{is}u_{it}\right)\right] \\
&= \frac{1}{\sigma^6}\left[\sum_{t=1}^{T}E(u_{it}^3) + \sum_{t=1}^{T}\sum_{s\neq t}E(u_{it}^2u_{is}) + 2\sum_{r=1}^{T}\sum_{t=2}^{T}\sum_{s<t}u_{is}u_{it}u_{ir}\right] \\
&= 0.
\end{aligned}
$$

Thus, $V_i^{(1)}$ and $V_i^{(2)}$ are orthogonal. Note that we used the third moments of (15) for the preceding calculation.

$$
\begin{aligned}
E[(V_i^{(2)})^2] &= \frac{T^2}{\sigma^4} - \frac{2T}{\sigma^6}E\left(\sum_{t=1}^{T}u_{it}\right)^2 + \frac{1}{\sigma^8}E\left(\sum_{t=1}^{T}u_{it}^2 + 2\sum_{t=2}^{T}\sum_{s<t}u_{is}u_{it}\right)^2 \\
&= \frac{T^2}{\sigma^4} - \frac{2T^2}{\sigma^4} + \frac{1}{\sigma^8}E\left[\sum_{t=1}^{T}u_{it}^4 + 2\sum_{t=2}^{T}\sum_{s<t}u_{it}^2u_{is}^2 + 4\sum_{t=2}^{T}\sum_{s<t}\left(\sum_{r=2}^{T}\sum_{q<r}u_{iq}u_{ir}\right)u_{is}u_{it}\right] \\
&= -\frac{T^2}{\sigma^4} + \frac{1}{\sigma^4}[3T + T(T-1) + 2(T-1)(T)] \\
&= \frac{2T^2}{\sigma^4}.
\end{aligned}
$$

We have used fourth moments of (15) for the preceding calculation.

$$
\begin{aligned}
E(U_{i,0}^{\rho}V_i) &= \frac{1}{\sigma^4}E\left[\left(\sum_{t=1}^{T}u_{it}y_{i,t-1}\right)\left(\sum_{t=1}^{T}u_{it}\right)\right] \\
&= \frac{1}{\sigma^4}\left[\sum_{t=1}^{T}E(u_{it}^2)y_{i,t-1} + \sum_{t=2}^{T}\sum_{s<t}E(u_{it}u_{is})y_{i,s-1} + \sum_{t=2}^{T}\sum_{s<t}E(u_{it}u_{is})y_{i,t-1}\right] \\
&= \frac{1}{\sigma^2}\sum_{t=1}^{T}y_{i,t-1},
\end{aligned}
$$

$$
\begin{aligned}
E(U_{i,0}^{\sigma^2}V_i) &= -\frac{T}{2\sigma^4}\sum_{t=1}^{T}E(u_{it}) + \frac{1}{2\sigma^6}E\left[\left(\sum_{t=1}^{T}u_{it}^2\right)\left(\sum_{t=1}^{T}u_{it}\right)\right] \\
&= \frac{1}{2\sigma^6}E\left[\sum_{s=1}^{T}\sum_{t=1}^{T}u_{it}^2u_{is}\right] \\
&= \frac{1}{2\sigma^6}\left[\sum_{t=1}^{T}E(u_{it}^3) + \sum_{t=2}^{T}\sum_{s<t}E(u_{it}^2u_{is}) + \sum_{t=2}^{T}\sum_{s<t}E(u_{is}^2u_{it})\right] \\
&= 0,
\end{aligned}
$$

$$
\begin{aligned}
E(U_{i,0}^{\sigma^2}V_i^{(2)}) &= \frac{T^2}{2\sigma^4}-\frac{T}{2\sigma^6}E\left[\sum_{t=1}^{T}u_{it}\right]^2+\frac{1}{2\sigma^8}E\left[\left(\sum_{t=1}^{T}u_{it}^2\right)\left(\sum_{t=1}^{T}u_{it}^2+2\sum_{t=2}^{T}\sum_{s<t}u_{it}u_{is}\right)\right] \\
&\quad -\frac{T}{2\sigma^6}E\left[\sum_{t=1}^{T}u_{it}\right]^2 \\
&= \frac{T^2}{2\sigma^4}-\frac{T^2}{2\sigma^4}+\frac{1}{2\sigma^8}T(3\sigma^4+(T-1)\sigma^4))-\frac{T^2}{2\sigma^4} \\
&= \frac{T}{\sigma^4},
\end{aligned}
$$

$$
\begin{aligned}
E(U_{i,0}^{\rho}V_i^{(2)}) &= -\frac{T}{\sigma^4}E\left[\sum_{t=1}^{T}u_{it}y_{i,t-1}\right]+\frac{1}{\sigma^6}E\left[\left(\sum_{t=1}^{T}u_{it}y_{i,t-1}\right)\left(\sum_{t=1}^{T}u_{it}^2+2\sum_{t=2}^{T}\sum_{s<t}u_{it}u_{is}\right)\right] \\
&= \frac{1}{\sigma^6}E\left[\sum_{t=1}^{T}u_{it}^3y_{i,t-1}+\sum_{t=1}^{T}\sum_{s\neq t}u_{is}^2u_{it}y_{i,t-1}+2\sum_{t=2}^{T}\sum_{s=t-1}u_{is}u_{it}^2y_{is}+2\sum_{t=2}^{T}\sum_{s\neq t-1}u_{is}u_{it}^2y_{is}\right] \\
&= \frac{2}{\sigma^6}E\left[\sum_{t=2}^{T}\sum_{s=t-1}u_{is}u_{it}^2y_{is}\right] \\
&= \frac{2(T-1)}{\sigma^2}.
\end{aligned}
$$

Thus, the second-order projected score is given by the system of equations below:

$$
\begin{aligned}
U_{i,2} &= \begin{bmatrix}U_{i,0}^{\rho}\\ U_{i,0}^{\sigma^2}\end{bmatrix}-\begin{bmatrix}\frac{1}{\sigma^2}\sum_{t=1}^{T}y_{i,t-1} & \frac{2(T-1)}{\sigma^2}\\ 0 & \frac{T}{\sigma^4}\end{bmatrix}\begin{bmatrix}\frac{\sigma^2}{T} & 0\\ 0 & \frac{\sigma^4}{2T^2}\end{bmatrix}\begin{bmatrix}\frac{1}{\sigma^2}\sum_{t=1}^{T}u_{it}\\ -\frac{T}{\sigma^2}+\left[\frac{1}{\sigma^2}\sum_{t=1}^{T}u_{it}\right]^2\end{bmatrix} \\
&= \begin{bmatrix}U_{i,0}^{\rho}\\ U_{i,0}^{\sigma^2}\end{bmatrix}-\begin{bmatrix}\frac{1}{T}\sum_{t=1}^{T}y_{i,t-1} & \frac{\sigma^2(T-1)}{T}\\ 0 & \frac{1}{2T}\end{bmatrix}\begin{bmatrix}\frac{1}{\sigma^2}\sum_{t=1}^{T}u_{it}\\ -\frac{T}{\sigma^2}+\left[\frac{1}{\sigma^2}\sum_{t=1}^{T}u_{it}\right]^2\end{bmatrix}
\end{aligned}
$$

Projected score for the static binary choice model with an exogenous regressor

Suppose $y_{it}|x_{i1},x_{i2}\sim Ber(p_{it})$ with

$$p_{it}=E(y_{it}|x_{i1},\ldots,x_{iT})=F(\alpha_i+x_{it}^{\mathrm{T}}\beta)\equiv F_{it}$$

for $i=1,\ldots,N$ and $t=1,2$. The uncentered moments of this conditional distribution are all equal to F_{it}. The second and third centered moments are given by

$$
\begin{aligned}
Var(y_{it}|x_{i1},\ldots,x_{iT}) &= F_{it}(1-F_{it})\\
E((y_{it}-F_{it})^3|x_{i1},\ldots,x_{iT}) &= F_{it}(1-F_{it})(1-2F_{it})
\end{aligned}
$$

The log-likelihood contribution for the ith cross-section is

$$\sum_{t=1}^{2}[y_{it}\log F_{it}+(1-y_{it})\log(1-F_{it})]$$

Define

$$h_{it}=\frac{f_{it}}{F_{it}(1-F_{it})}$$

where $f_{it}\equiv f(\alpha_i+x_{it}^{\mathrm{T}}\beta)$. For the logit case, we must have $h_{it}=1$. The scores for the structural parameter and the incidental parameter are

$$\begin{aligned}U_{i,0} &= \sum_{t=1}^{2}(y_{it}-F_{it})h_{it}x_{it}\\ V_i^{(1)} &= \sum_{t=1}^{2}(y_{it}-F_{it})h_{it}\end{aligned}$$

The components of the second-order projected score are calculated below. First, $E(V_i^2)$ is given by

$$\begin{aligned}(V_i^{(1)})^2 &= \sum_{t=1}^{2}(y_{it}-F_{it})^2h_{it}^2+2(y_{i1}-F_{i1})(y_{i2}-F_{i2})h_{i1}h_{i2}\\ \Rightarrow E[(V_i^{(1)})^2] &= \sum_{t=1}^{2}F_{it}(1-F_{it})h_{it}^2\end{aligned}$$

Note that we used second moment properties and the independence. Next, $V_i^{(2)}$ is given by:

$$\begin{aligned}V_i^{(2)} &= \partial_{\alpha_i}V_i^{(1)}+(V_i^{(1)})^2\\ &= \sum_{t=1}^{2}\left[(y_{it}-F_{it})\partial_{\alpha_i}h_{it}-h_{it}f_{it}\right]+(V_i^{(1)})^2\end{aligned}$$

To orthogonalize $V_i^{(2)}$, we compute

$$V_i^{(2),*}=V_i^{(2)}-\frac{E(V_i^{(1)}V_i^{(2)})}{E[(V_i^{(1)})^2]}V_i^{(1)}$$

Note that

$$\begin{aligned}V_i^{(1)}V_i^{(2)} &= (y_{i1}-F_{i1})^3h_{i1}^3+3(y_{i1}-F_{i1})^2h_{i1}^2(y_{i2}-F_{i2})h_{i2}\\ &\quad+3(y_{i2}-F_{i2})h_{i2}(y_{i1}-F_{i1})^2h_{i1}^2+(y_{i2}-F_{i2})^3h_{i2}^3\\ &\quad+(y_{i1}-F_{i1})^2(\partial_{\alpha_i}h_{i1})h_{i1}+(y_{i2}-F_{i2})^2(\partial_{\alpha_i}h_{i2})h_{i2}\\ &\quad+(y_{i1}-F_{i1})(\partial_{\alpha_i}h_{i1})(y_{i2}-F_{i2})h_{i2}+(y_{i2}-F_{i2})(\partial_{\alpha_i}h_{i2})(y_{i1}-F_{i1})h_{i1}\end{aligned}$$

$$\begin{aligned}\Rightarrow E(V_i^{(1)}V_i^{(2)}) &= F_{i1}(1-F_{i1})(1-2F_{i1})h_{i1}^3 + F_{i2}(1-F_{i2})(1-2F_{i2})h_{i2}^3 \\ &+ F_{i1}(1-F_{i1})(\partial_{\alpha_i}h_{i1})h_{i1} + F_{i2}(1-F_{i2})(\partial_{\alpha_i}h_{i2})h_{i2}\end{aligned}$$

The last two terms above are equal to zero in the logit case since $\partial_{\alpha_i}h_{it} = 0$. Now, $E(V_i^{(2),*})^2 = E(V_i^{(2),*}V_i^{(2)}) = E(\partial_{\alpha_i}^2 V_i^{(2),*})$. The resulting expression will be a very complicated expression and has to be computed symbolically. Next, we compute $E(U_{i,0}V_i^{(1)})$:

$$U_{i,0}V_i^{(1)} = \sum_{t=1}^{2}\left[(y_{it}-F_{it})^2 h_{it}^2 x_{it}\right] + (y_{i1}-F_{i1})(y_{i2}-F_{i2})h_{i2} + (y_{i2}-F_{i2})(y_{i1}-F_{i1})h_{i1}$$

As a result, we have

$$E(U_{i,0}V_i^{(1)}) = \sum_{t=1}^{2} F_{it}(1-F_{it})h_{it}^2 x_{it}$$

Finally, $E(U_{i,0}V_i^{(2),*})$ has to be computed symbolically.

B Proofs of some results

Proof of Lemma 1

To show that $U_{i,2}$ is an unbiased estimating equation, we have to show that $\mathbb{E}\left[U_{i,0}\right] = 0$, $\mathbb{E}[V_i] = 0$, and $\mathbb{E}\left[vec\left[V_i^{(2)}\right]\right] = 0$. The first two statements follow from the zero-mean property of the scores. Since $\mathbb{E}\left[vec\left[V_i^{(2)}\right]\right] = vec\left(\mathbb{E}\left[V_i^{(2)}\right]\right)$, we have to show that $\mathbb{E}\left[V_i^{(2)}\right] = 0$. Differentiating $\mathbb{E}[V_i] = 0$ with respect to α_i gives the desired result. Thus, we have shown that $U_{i,2}$ is an unbiased estimating equation. To show second-order $\mathbb{E}$-ancillarity, we can show that (12) satisfies the moment conditions in (3) for $k = 1, 2$. This follows by construction. ■

Proof of Proposition 1

To simplify the exposition, I return to the case where incidental parameter is scalar. To show (13), consider a second-order Taylor series expansion of the plug-in second-order projected score for the ith individual about the true value α_{i0}, i.e.

$$\begin{aligned}\widehat{U_{i,2}}(\theta_0) &= U_{i,2} + \partial_{\alpha_i}U_{i,2}(\widehat{\alpha_i}(\theta_0)-\alpha_{i0}) + \frac{1}{2}\partial_{\alpha_i}^2 U_{i,2}(\widehat{\alpha_i}(\theta_0)-\alpha_{i0})^2 \\ &+ O_p\left(T^{-1/2}\right). \qquad (16)\end{aligned}$$

Under regularity conditions for maximum likelihood estimation, the three terms in (16) are $O_p\left(T^{1/2}\right)$, $O_p\left(T^{1/2}\right)$, and $O_p(1)$. The final term is a zero mean $O_p\left(T^{-1/2}\right)$ term. Note that the first-order conditions used to obtain a plug-in estimator for α_i can be expanded in the

following way:

$$\widehat{V_i}(\theta_0) = V_i + \partial_{\alpha_i} V_i (\widehat{\alpha_i}(\theta_0) - \alpha_{i0}) + O_p(1).$$

Since the right hand size is equal to zero, we can write

$$\widehat{\alpha_i}(\theta_0) - \alpha_{i0} = -\frac{V_i}{\mathbb{E}\left(\partial_{\alpha_i} V_i\right)} + O_p\left(T^{-1}\right). \tag{17}$$

Furthermore, the square of $\widehat{\alpha_i}(\theta_0) - \alpha_{i0}$ can be written as

$$\begin{aligned}
(\widehat{\alpha_i}(\theta_0) - \alpha_{i0})^2 &= \frac{V_i^2}{\left[\mathbb{E}\left(\partial_{\alpha_i} V_i\right)\right]^2} - 2\frac{V_i}{\mathbb{E}\left(\partial_{\alpha_i} V_i\right)} O_p\left(T^{-1}\right) + O_p\left(T^{-2}\right) \\
&= \frac{V_i^2}{\left[\mathbb{E}\left(\partial_{\alpha_i} V_i\right)\right]^2} - 2\frac{\mathbb{E}(V_i)}{\mathbb{E}\left(\partial_{\alpha_i} V_i\right)} O_p\left(T^{-1}\right) - 2\frac{O_p\left(T^{1/2}\right)}{\mathbb{E}\left(\partial_{\alpha_i} V_i\right)} O_p\left(T^{-1}\right) + O_p\left(T^{-2}\right) \\
&= \frac{\mathbb{E}\left(V_i^2\right)}{\left[\mathbb{E}\left(\partial_{\alpha_i} V_i\right)\right]^2} + \frac{O_p\left(T^{1/2}\right)}{\left[\mathbb{E}\left(\partial_{\alpha_i} V_i\right)\right]^2} + O_p\left(T^{-3/2}\right) \\
&= \frac{\mathbb{E}\left(V_i^2\right)}{\left[\mathbb{E}\left(\partial_{\alpha_i} V_i\right)\right]^2} + O_p\left(T^{-3/2}\right).
\end{aligned} \tag{18}$$

Note that $\mathbb{E}(V_i) = 0$ because the α_i-score is an unbiased estimating equation. Central limit theorems for V_i and V_i^2 allow us to obtain (18). After substituting (17) into $\partial_{\alpha_i} U_{i,2}(\widehat{\alpha_i}(\theta_0) - \alpha_{i0})$, we have

$$\begin{aligned}
\partial_{\alpha_i} U_{i,2}(\widehat{\alpha_i}(\theta_0) - \alpha_{i0}) &= -\frac{V_i \partial_{\alpha_i} U_{i,2}}{\mathbb{E}\left(\partial_{\alpha_i} V_i\right)} + \partial_{\alpha_i} U_{i,2} O_p\left(T^{-1}\right) \\
&= -\frac{V_i \partial_{\alpha_i} U_{i,2}}{\mathbb{E}\left(\partial_{\alpha_i} V_i\right)} + \mathbb{E}\left(\partial_{\alpha_i} U_{i,2}\right) O_p\left(T^{-1}\right) + O_p\left(T^{-1/2}\right) \\
&= -\frac{V_i \partial_{\alpha_i} U_{i,2}}{\mathbb{E}\left(\partial_{\alpha_i} V_i\right)} + O_p\left(T^{-1/2}\right)
\end{aligned} \tag{19}$$

A central limit theorem for $\partial_{\alpha_i} U_{i,2}$ and second-order local $\mathbb{E}$-ancillarity allow us to produce the previous derivation. The expression in (19) involves the product of $\partial_{\alpha_i} U_{i,2}$ and V_i and a zero mean $O_p\left(T^{-1/2}\right)$ term. As a result, the expectation of the term $\partial_{\alpha_i} U_{i,2}(\widehat{\alpha_i}(\theta_0) - \alpha_{i0})$ is $O\left(T^{-1}\right)$.

Next, we substitute (18) into $\partial_{\alpha_i}^2 U_{i,2}(\widehat{\alpha_i}(\theta_0) - \alpha_{i0})^2$. As a result, we obtain

$$\begin{aligned}
\partial_{\alpha_i}^2 U_{i,2}(\widehat{\alpha_i}(\theta_0) - \alpha_{i0})^2 &= \frac{\partial_{\alpha_i}^2 U_{i,2} \mathbb{E}\left(V_i^2\right)}{\left[\mathbb{E}\left(\partial_{\alpha_i} V_i\right)\right]^2} + \partial_{\alpha_i}^2 U_{i,2} O_p\left(T^{-3/2}\right) \\
&= \frac{\partial_{\alpha_i}^2 U_{i,2} \mathbb{E}\left(V_i^2\right)}{\left[\mathbb{E}\left(\partial_{\alpha_i} V_i\right)\right]^2} + \mathbb{E}\left(\partial_{\alpha_i}^2 U_{i,2}\right) O_p\left(T^{-3/2}\right) + O_p\left(T^{1/2}\right) O_p\left(T^{-3/2}\right)
\end{aligned}$$

$$= \frac{\partial^2_{\alpha_i} U_{i,2} \mathbb{E}\left(V_i^2\right)}{\left[\mathbb{E}\left(\partial_{\alpha_i} V_i\right)\right]^2} + O_p\left(T^{-1}\right) \tag{20}$$

The expression in (20) involves $\partial^2_{\alpha_i} U_{i,2}$, which has zero expectation because of second-order local $\mathbb{E}$-ancillarity, and an $O_p\left(T^{-1}\right)$ term. As a result, the expectation of the term $\partial^2_{\alpha_i} U_{i,2}(\widehat{\alpha_i}(\theta_0) - \alpha_{i0})$ is $O\left(T^{-1}\right)$.

Proof of Proposition 2

Assume that the system of equations implied by the plug-in second-order projected score has a solution in some neighborhood of the true value θ_0. We denote this solution by $\widehat{\theta^c}$ and it satisfies $\sum_{i=1}^N \widehat{U_{i,2}}\left(\widehat{\theta^c}\right) = 0$. Consider the following first-order Taylor series expansion of the plug-in second-order projected score around θ_0, i.e.

$$\sum_{i=1}^n \widehat{U_{i,2}}\left(\widehat{\theta^c}\right) = \sum_{i=1}^n \widehat{U_{i,2}}(\theta_0) + \sum_{i=1}^n \frac{d}{d\theta}\widehat{U_{i,2}}\left(\bar{\theta}\right)\left(\widehat{\theta^c} - \theta_0\right). \tag{21}$$

Note that the left hand side of (21) is equal to zero because $\widehat{\theta^c}$ is the root of the plug-in second-order projected score. Rewrite (21) as

$$\sqrt{nT}\left(\widehat{\theta^c} - \theta_0\right) = \left(\frac{1}{nT}\sum_{i=1}^n \frac{d}{d\theta}\widehat{U_{i,2}}\left(\bar{\theta}\right)\right)^{-1} \frac{1}{\sqrt{nT}}\sum_{i=1}^n \widehat{U_{i,2}}(\theta_0). \tag{22}$$

Let $n, T \to \infty$ and $n/T \to c \in (0, \infty)$. Note that

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^n \left[\widehat{U_{i,2}}(\theta_0) - U_{i,2}\right] = \frac{1}{\sqrt{nT}}\sum_{i=1}^n \mathbb{E}\left[\widehat{U_{i,2}}(\theta_0) - U_{i,2}\right] + O_p(1) = O_p\left(\sqrt{\frac{n}{T^3}}\right) + O_p(1)$$

The first equality comes from replacing the empirical mean with an expectation and leaving behind a zero-mean $O_p(1)$ term. The second equality comes from the order calculation in Proposition 1. Provided that $n/T^3 \to 0$, $\frac{1}{\sqrt{nT}}\sum_i \widehat{U_{i,2}}(\theta_0)$ can be approximated by $\frac{1}{\sqrt{nT}}\sum_i U_{i,2}$ and the latter quantity is asymptotically normal. A central limit theorem applies to $\frac{1}{\sqrt{nT}}\sum_i U_{i,2}$ (similar to the score in likelihood settings), i.e.

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^n U_{i,2} \xrightarrow{d} N\left(0, \lim_{n,T\to\infty} \frac{1}{nT}\sum_{i=1}^n \mathbb{E}\left[U_{i,2}U_{i,2}^{\mathrm{T}}\right]\right).$$

Next, note that

$$\left.\frac{d}{d\theta}\widehat{U_{i,2}}(\theta)\right|_{\theta=\bar{\theta}} = \left.\left[\partial_\theta \widehat{U_{i,2}}(\theta) + \left(\partial_{\alpha_i}\widehat{U_{i,2}}(\theta)\right)\left(\partial_\theta \widehat{\alpha_i}(\theta)\right)\right]\right|_{\theta=\bar{\theta}} \tag{23}$$

by the chain rule. Replacing $\widehat{U_{i,2}}(\theta)$ with its Taylor series expansion

$$\widehat{U_{i,2}}(\theta) = U_{i,2}(\theta, \alpha_{i0}) + \partial_{\alpha_i} U_{i,2}(\theta, \alpha_{i0})(\widehat{\alpha_i}(\theta) - \alpha_{i0}) + O_p(1) \tag{24}$$

and calculating the derivatives in (23) yields

$$\begin{aligned}
\partial_\theta \widehat{U_{i,2}}(\theta) &= \partial_\theta U_{i,2}(\theta, \alpha_{i0}) + \partial^2_{\theta\alpha_i} U_{i,2}(\theta, \alpha_{i0})(\widehat{\alpha_i}(\theta) - \alpha_{i0}) \\
&\quad + \partial_{\alpha_i} U_{i,2}(\theta, \alpha_{i0})(\partial_\theta \widehat{\alpha_i}(\theta)) + O_p(1), \qquad (25) \\
\partial_{\alpha_i} \widehat{U_{i,2}}(\theta) &= \partial_{\alpha_i} U_{i,2}(\theta, \alpha_{i0}) + \partial^2_{\alpha_i} U_{i,2}(\theta, \alpha_{i0})(\widehat{\alpha_i}(\theta) - \alpha_{i0}) \\
&\quad + \partial_{\alpha_i} U_{i,2}(\theta, \alpha_{i0})\left(\partial_{\alpha_i} \widehat{\alpha_i}(\theta)\right) + O_p(1). \qquad (26)
\end{aligned}$$

Taking probability limits, we have the following components:

$$\begin{aligned}
\operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \operatorname*{plim}_{T\to\infty} \frac{1}{T} \partial_\theta U_{i,2}(\theta, \alpha_{i0}) \Big|_{\theta=\bar{\theta}} &= \lim_{n,T\to\infty} \frac{1}{nT} \sum_{i=1}^{n} \mathbb{E}\left[\partial_\theta U_{i,2}\right], \\
\operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \operatorname*{plim}_{T\to\infty} \frac{1}{T} \partial^2_{\theta\alpha_i} U_{i,2}(\theta, \alpha_{i0})(\widehat{\alpha_i}(\theta) - \alpha_{i0}) \Big|_{\theta=\bar{\theta}} &= 0, \\
\operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \operatorname*{plim}_{T\to\infty} \frac{1}{T} \partial_{\alpha_i} U_{i,2}(\theta, \alpha_{i0})\left(\partial_{\alpha_i} \widehat{\alpha_i}(\theta)\right) \Big|_{\theta=\bar{\theta}} &= 0, \\
\operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \operatorname*{plim}_{T\to\infty} \frac{1}{T} \partial_{\alpha_i} U_{i,2}(\theta, \alpha_{i0}) \Big|_{\theta=\bar{\theta}} &= 0, \\
\operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \operatorname*{plim}_{T\to\infty} \frac{1}{T} \partial^2_{\alpha_i} U_{i,2}(\theta, \alpha_{i0})(\widehat{\alpha_i}(\theta) - \alpha_{i0}) \Big|_{\theta=\bar{\theta}} &= 0, \\
\operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \operatorname*{plim}_{T\to\infty} \frac{1}{T} \partial_{\alpha_i} U_{i,2}(\theta, \alpha_{i0})\left(\partial_{\alpha_i} \widehat{\alpha_i}(\theta)\right) \Big|_{\theta=\bar{\theta}} &= 0.
\end{aligned}$$

Note that as $T \to \infty$, we have both $\bar{\theta} \xrightarrow{p} \theta_0$ and $\widehat{\alpha_i}\left(\bar{\theta}\right) \xrightarrow{p} \alpha_{i0}$. The second and fifth equalities follow $\widehat{\alpha_i}\left(\bar{\theta}\right) \xrightarrow{p} \alpha_{i0}$ as $T \to \infty$. The third, fourth, and sixth equalities would follow from the law of large numbers and second-order ancillarity. The $O_p(1)$ terms in (24), (25), and (26) all converge to zero because $\widehat{\alpha_i}\left(\bar{\theta}\right) \xrightarrow{p} \alpha_{i0}$ as $T \to \infty$. We can then conclude that

$$\operatorname*{plim}_{n,T\to\infty} \frac{1}{nT} \sum_{i=1}^{n} \frac{d}{d\theta} \widehat{U_{i,2}}(\theta) \Big|_{\theta=\bar{\theta}} = \lim_{n,T\to\infty} \frac{1}{nT} \sum_{i=1}^{n} \mathbb{E}\left[\partial_\theta U_{i,2}\right].$$

Notice that $U_{i,2}$ behaves like $U_{i,0}$ asymptotically because the correction in (12) has expectation zero at the true value. As long as the information identity holds, we have $\mathbb{E}\left[U_{i,2} U_{i,2}^{\mathrm{T}}\right] = \mathbb{E}\left[U_{i,0} U_{i,0}^{\mathrm{T}}\right] = \mathbb{E}\left[\partial_\theta U_{i,0}\right]$. Otherwise, we have the usual sandwich-type asymptotic covariance matrix.

Alternative proof of (14)

In this appendix, we prove the main results in the spirit of the papers by Hahn and Newey (2004) and Hahn and Kuersteiner (2011). We also note some departures from their proof. Let F_i and $\widehat{F_i}$ denote the CDF and its empirical counterpart for the ith individual. Define $F_i(\epsilon) = F_i + \epsilon\sqrt{T}(\widehat{F_i} - F_i)$ and $\Delta_{iT} = \sqrt{T}(\widehat{F_i} - F_i)$, where $\epsilon \in [0, T^{-1/2}]$. We have $\mathbf{F}(\epsilon) = \mathbf{F} + \epsilon\sqrt{T}(\widehat{\mathbf{F}} - \mathbf{F})$ in vector form.

Let $\alpha_i(\boldsymbol{\theta}, F_i(\epsilon))$ and $\boldsymbol{\theta}(\mathbf{F}(\epsilon))$ be the solutions to the estimating equations below:

$$\int V_i(\boldsymbol{\theta}, \alpha_i(\boldsymbol{\theta}, F_i(\epsilon)); \mathbf{z}_i)\, dF_i(\epsilon) = 0 \tag{27}$$

$$\sum_{i=1}^{n} \int U_{i,2}(\boldsymbol{\theta}(\mathbf{F}(\epsilon)), \alpha_i(\boldsymbol{\theta}(\mathbf{F}(\epsilon)), F_i(\epsilon)); \mathbf{z})\, dF_i(\epsilon) = 0 \tag{28}$$

The plug-in used for the α_i's in the second-order projected score can be written as $\widehat{\alpha_i}(\theta) = \alpha_i(\boldsymbol{\theta}, F_i(T^{-1/2}))$. The root for the plug-in version of the second-order projected score can be written as $\widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}(\mathbf{F}(T^{-1/2}))$. On the other hand, the true values can be written as $\boldsymbol{\theta}_0 = \boldsymbol{\theta}(\mathbf{F}(0)) = \boldsymbol{\theta}(\mathbf{F})$ and $\alpha_{i0} = \alpha_i(\boldsymbol{\theta}, F_i)$.

Expand the functional $\boldsymbol{\theta}(\widehat{\mathbf{F}})$ about the true value $\boldsymbol{\theta}(\mathbf{F})$ up to the third order, i.e.

$$\boldsymbol{\theta}(\widehat{\mathbf{F}}) - \boldsymbol{\theta}(\mathbf{F}) = \frac{1}{\sqrt{T}}\boldsymbol{\theta}^{\epsilon}(0) + \frac{1}{2}\left(\frac{1}{\sqrt{T}}\right)^2 \boldsymbol{\theta}^{\epsilon\epsilon}(0) + \frac{1}{6}\left(\frac{1}{\sqrt{T}}\right)^3 \boldsymbol{\theta}^{\epsilon\epsilon\epsilon}(\tilde{\epsilon}) \tag{29}$$

where

$$\boldsymbol{\theta}^{\epsilon}(0) = \partial_{\epsilon}\boldsymbol{\theta}(\mathbf{F}(\epsilon))|_{\epsilon=0},\ \boldsymbol{\theta}^{\epsilon\epsilon}(0) = \partial_{\epsilon}^2\boldsymbol{\theta}(\mathbf{F}(\epsilon))|_{\epsilon=0},\ \boldsymbol{\theta}^{\epsilon\epsilon\epsilon}(0) = \partial_{\epsilon}^3\boldsymbol{\theta}(\mathbf{F}(\epsilon))|_{\epsilon=\tilde{\epsilon}\in[0,T^{-1/2}]} \tag{30}$$

Define the object

$$h_i(\epsilon) = U_{i,2}(\boldsymbol{\theta}(\mathbf{F}(\epsilon)), \alpha_i(\boldsymbol{\theta}(\mathbf{F}(\epsilon)), F_i(\epsilon))) \tag{31}$$

where the dependence on the data is suppressed. Hahn and Newey (2004) and Hahn and Kuersteiner (2011) use $U_{i,1}$ instead of $U_{i,2}$. It follows that (28) can be rewritten as

$$\frac{1}{n}\sum_{i=1}^{n} \int h_i(\epsilon)\, dF_i(\epsilon) = 0 \tag{32}$$

We show that when $n, T \to \infty$ such that $n/T \to c \in (0, \infty)$,

$$\sqrt{nT}(\boldsymbol{\theta}(\widehat{\mathbf{F}}) - \boldsymbol{\theta}(\mathbf{F})) \xrightarrow{d} N\left(\mathbf{0}, \left(\lim_{n,T\to\infty} \frac{1}{nT}\sum_{i=1}^{n} \mathbf{I}_i\right)^{-1}\right) \tag{33}$$

in the following manner:

1. Differentiate (32) with respect to ϵ twice. The resulting expressions can be decom-

posed into two terms: a term that requires integration with respect to $F_i(\epsilon)$ and a term that characterizes the "tail" or the remainder. We have

$$\frac{1}{n}\sum_{i=1}^{n}\int \frac{dh_i(\epsilon)}{d\epsilon}\, dF_i(\epsilon) + \frac{1}{n}\sum_{i=1}^{n}\int h_i(\epsilon)\, d\Delta_{iT} = 0 \tag{34}$$

$$\frac{1}{n}\sum_{i=1}^{n}\int \frac{d^2h_i(\epsilon)}{d\epsilon^2}\, dF_i(\epsilon) + \frac{2}{n}\sum_{i=1}^{n}\int \frac{dh_i(\epsilon)}{d\epsilon}\, d\Delta_{iT} = 0 \tag{35}$$

2. Compute the total derivatives in the previous equations noting the dependence of $\boldsymbol{\theta}(\mathbf{F}(\epsilon))$ and $\alpha_i(\boldsymbol{\theta}(\mathbf{F}(\epsilon)), F_i(\epsilon))$ on ϵ.

$$\begin{aligned}
\frac{dh_i(\epsilon)}{d\epsilon} &= \partial_{\boldsymbol{\theta}} h_i(\epsilon)\partial_\epsilon \boldsymbol{\theta} + \partial_{\alpha_i} h_i(\epsilon)(\partial_{\boldsymbol{\theta}}\alpha_i)^{\mathrm{T}}\partial_\epsilon\boldsymbol{\theta} + \partial_{\alpha_i} h_i(\epsilon)\partial_\epsilon\alpha_i \\
\frac{d^2h_i(\epsilon)}{d\epsilon^2} &= \underbrace{\partial_\epsilon\boldsymbol{\theta}\left[\partial^2_\theta h_i(\epsilon)\partial_\epsilon\boldsymbol{\theta} + \partial^2_{\theta,\alpha_i} h_i(\epsilon)\partial_{\boldsymbol{\theta}}\alpha_i\partial_\epsilon\boldsymbol{\theta} + \partial^2_{\theta,\alpha_i} h_i(\epsilon)\partial_\epsilon\alpha_i\right] + \partial_{\boldsymbol{\theta}} h_i(\epsilon)\partial^2_\epsilon\boldsymbol{\theta}}_{\partial_\epsilon(\partial_\theta h_i(\epsilon)\partial_\epsilon\boldsymbol{\theta})} \\
&\quad + \partial_{\boldsymbol{\theta}}\alpha_i\partial_\epsilon\boldsymbol{\theta}\underbrace{\left[\partial^2_{\theta,\alpha_i} h_i(\epsilon)\partial_\epsilon\boldsymbol{\theta} + \partial^2_{\alpha_i} h_i(\epsilon)\partial_{\boldsymbol{\theta}}\alpha_i\partial_\epsilon\boldsymbol{\theta} + \partial^2_{\alpha_i} h_i(\epsilon)\partial_\epsilon\alpha_i\right]}_{\partial_\epsilon(\partial_{\alpha_i} h_i(\epsilon))} \\
&\quad + \partial_{\alpha_i} h_i(\epsilon)\underbrace{\left[(\partial_\epsilon\boldsymbol{\theta})^2\partial^2_{\boldsymbol{\theta}}\alpha_i + \partial^2_{\boldsymbol{\theta},\epsilon}\alpha_i\partial_\epsilon\boldsymbol{\theta} + \partial_{\boldsymbol{\theta}}\alpha_i\partial^2_\epsilon\boldsymbol{\theta}\right]}_{\partial_\epsilon(\partial_{\boldsymbol{\theta}}\alpha_i\partial_\epsilon\boldsymbol{\theta})} \\
&\quad + \partial_\epsilon\alpha_i\left[\partial^2_{\theta,\alpha_i} h_i(\epsilon)\partial_\epsilon\boldsymbol{\theta} + \partial^2_{\alpha_i} h_i(\epsilon)\partial_{\boldsymbol{\theta}}\alpha_i\partial_\epsilon\boldsymbol{\theta} + \partial^2_{\alpha_i} h_i(\epsilon)\partial_\epsilon\alpha_i\right] \\
&\quad + \partial_{\alpha_i} h_i(\epsilon)\underbrace{\left[(\partial^2_{\boldsymbol{\theta},\epsilon}\alpha_i)^{\mathrm{T}}\partial_\epsilon\boldsymbol{\theta} + \partial^2_\epsilon\alpha_i\right]}_{\partial_\epsilon(\partial_\epsilon\alpha_i)}
\end{aligned}$$

3. Next, we have to derive $\boldsymbol{\theta}^\epsilon(0)$ and $\boldsymbol{\theta}^{\epsilon\epsilon}(0)$. This means that we have to evaluate the expressions in (b) at $\epsilon = 0$. Use the definitions of $\boldsymbol{\theta}_0$, α_{i0} and (31) to rewrite the resulting expressions. As a consequence, we have

$$\begin{aligned}
\left.\frac{dh_i(\epsilon)}{d\epsilon}\right|_{\epsilon=0} &= \left[\partial_{\boldsymbol{\theta}} U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0}) + \partial_{\alpha_i} U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0})(\partial_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}_0,F_i))^{\mathrm{T}}\right]\boldsymbol{\theta}^\epsilon(0) \\
&\quad + \partial_{\alpha_i} U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0})\partial_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}_0,F_i)\boldsymbol{\theta}^\epsilon(0) \\
&\quad + \partial_{\alpha_i} U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0})\partial_\epsilon\alpha_i(\boldsymbol{\theta}_0,F_i)
\end{aligned} \tag{36}$$

$$\begin{aligned}
\left.\frac{d^2h_i(\epsilon)}{d\epsilon^2}\right|_{\epsilon=0} &= \boldsymbol{\theta}^\epsilon(0)\left[\partial^2_{\boldsymbol{\theta}} U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0})\boldsymbol{\theta}^\epsilon(0) + \partial^2_{\boldsymbol{\theta},\alpha_i} U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0})\partial_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}_0,F_i)\boldsymbol{\theta}^\epsilon(0)\right] \\
&\quad + \boxed{\boldsymbol{\theta}^\epsilon(0)\partial^2_{\boldsymbol{\theta},\alpha_i} U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0})\partial_\epsilon\alpha_i(\boldsymbol{\theta}_0,F_i)} + \partial_{\boldsymbol{\theta}} U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0})\boldsymbol{\theta}^{\epsilon\epsilon}(0) \\
&\quad + \partial_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}_0,F_i)\boldsymbol{\theta}^\epsilon(0)\left[\partial^2_{\boldsymbol{\theta},\alpha_i} U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0})\boldsymbol{\theta}^\epsilon(0)\right] \\
&\quad + \boxed{\partial_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}_0,F_i)\boldsymbol{\theta}^\epsilon(0)\left[\partial^2_{\alpha_i} U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0})\partial_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}_0,F_i)\boldsymbol{\theta}^\epsilon(0)\right]}
\end{aligned}$$

$$
\begin{aligned}
&+ \boxed{\partial_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}_0, F_i)\boldsymbol{\theta}^{\epsilon}(0)\Big[\partial^2_{\alpha_i} U_{i,2}(\boldsymbol{\theta}_0, \alpha_{i0})\partial_{\epsilon}\alpha_i(\boldsymbol{\theta}_0, F_i)\Big]} \\
&+ \boxed{\partial_{\alpha_i} U_{i,2}(\boldsymbol{\theta}_0, \alpha_{i0})(\boldsymbol{\theta}^{\epsilon}(0))^{\mathrm{T}}\partial^2_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}_0, F_i)\boldsymbol{\theta}^{\epsilon}(0)} \\
&+ \boxed{\partial_{\alpha_i} U_{i,2}(\boldsymbol{\theta}_0, \alpha_{i0})\partial^2_{\boldsymbol{\theta},\epsilon}\alpha_i(\boldsymbol{\theta}_0, F_i)\boldsymbol{\theta}^{\epsilon}(0)} \\
&+ \boxed{\partial_{\alpha_i} U_{i,2}(\boldsymbol{\theta}_0, \alpha_{i0})[\partial_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}_0, F_i)\boldsymbol{\theta}^{\epsilon\epsilon}(0)]} \\
&+ \boxed{\partial_{\epsilon}\alpha_i(\boldsymbol{\theta}_0, F_i)\Big[\partial^2_{\boldsymbol{\theta},\alpha_i} U_{i,2}(\boldsymbol{\theta}_0, \alpha_{i0})\boldsymbol{\theta}^{\epsilon}(0)\Big]} \\
&+ \boxed{\partial_{\epsilon}\alpha_i(\boldsymbol{\theta}_0, F_i)\Big[\partial^2_{\alpha_i} U_{i,2}(\boldsymbol{\theta}_0, \alpha_{i0})\partial_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}_0, F_i)\boldsymbol{\theta}^{\epsilon}(0)\Big]} \\
&+ \boxed{\boxed{\partial_{\epsilon}\alpha_i(\boldsymbol{\theta}_0, F_i)\Big[\partial^2_{\alpha_i} U_{i,2}(\boldsymbol{\theta}_0, \alpha_{i0})\partial_{\epsilon}\alpha_i(\boldsymbol{\theta}_0, F_i)\Big]}} \\
&+ \boxed{\partial_{\alpha_i} U_{i,2}(\boldsymbol{\theta}_0, \alpha_{i0})\partial^2_{\boldsymbol{\theta},\epsilon}\alpha_i(\boldsymbol{\theta}_0, F_i)\boldsymbol{\theta}^{\epsilon}(0)} \\
&+ \boxed{\boxed{\partial_{\alpha_i} U_{i,2}(\boldsymbol{\theta}_0, \alpha_{i0})\partial^2_{\epsilon}\alpha_i(\boldsymbol{\theta}_0, F_i)}} \qquad (37)
\end{aligned}
$$

4. Substitute the above expressions into (34) and (35). The first sum in (34) and (35) when evaluated at $\epsilon = 0$ becomes the expectation with respect to the true values while the second sum becomes a "tail" term characterizing the difference between the realized distribution $\widehat{F_i}$ and the true one F_i. Since $\boldsymbol{\theta}^{\epsilon}(0)$ and $\boldsymbol{\theta}^{\epsilon\epsilon}(0)$ do not depend on the data, they can be treated as constants with respect to the expectation.

5. We need to derive the expressions for the first and second derivatives of $\alpha_i(\boldsymbol{\theta}, F_i)$ with respect to $\boldsymbol{\theta}$ and ϵ. Differentiate (27) with respect to $\boldsymbol{\theta}$ and ϵ. Solve the resulting system of two equations in two unknowns for $\partial_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}, F_i(\epsilon))$ and $\partial_{\epsilon}\alpha_i(\boldsymbol{\theta}, F_i(\epsilon))$. Next, get the second derivatives of (27) with respect to $\boldsymbol{\theta}$ and ϵ. Solve the resulting system of three equations in three unknowns for $\partial^2_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}, F_i(\epsilon))$, $\partial^2_{\boldsymbol{\theta},\epsilon}\alpha_i(\boldsymbol{\theta}, F_i(\epsilon))$, and $\partial^2_{\epsilon}\alpha_i(\boldsymbol{\theta}, F_i(\epsilon))$.[7] In effect, we are applying the Implicit Function Theorem and evaluating at $\epsilon = 0$ and $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. The resulting first derivatives would be

$$
\partial_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}_0, F_i) = -\frac{\mathbb{E}(\partial_{\boldsymbol{\theta}} V_i^{(1)})}{\mathbb{E}(\partial_{\alpha_i} V_i^{(1)})} = O_p(1) \qquad (38)
$$

$$
\partial_{\epsilon}\alpha_i(\boldsymbol{\theta}_0, F_i) = -\frac{T^{1/2}\frac{1}{T}\left(V_i^{(1)} - \mathbb{E}(V_i^{(1)})\right)}{\mathbb{E}(\partial_{\alpha_i} V_i^{(1)})} = O_p(T^{-1}) \qquad (39)
$$

[7] The systems of equations can be found in the appendix of Hahn and Kuersteiner (2011). Refer to pages 1178 and 1181. Solving the system of equations is not as hard as it sounds because the coefficient matrix is diagonal.

The resulting second derivatives would be

$$\begin{aligned}
\partial^2_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}_0,F_i) &= -\frac{1}{\mathbb{E}(\partial_{\alpha_i}V_i^{(1)})}\left[\mathbb{E}(\partial^2_{\boldsymbol{\theta}}V_i^{(1)})+\partial_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}_0,F_i)\mathbb{E}(\partial^2_{\boldsymbol{\theta},\alpha_i}V_i^{(1)})^{\mathrm{T}}\right]\\
&\quad -\frac{1}{\mathbb{E}(\partial_{\alpha_i}V_i^{(1)})}\left[\mathbb{E}(\partial^2_{\boldsymbol{\theta},\alpha_i}V_i^{(1)})(\partial_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}_0,F_i))^{\mathrm{T}}\right]\\
&\quad -\frac{1}{\mathbb{E}(\partial_{\alpha_i}V_i^{(1)})}\left[\mathbb{E}(\partial^2_{\alpha_i}V_i^{(1)})\partial_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}_0,F_i)\partial_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}_0,F_i)^{\mathrm{T}}\right]\\
&= O_p(1) \qquad (40)\\
\partial^2_{\boldsymbol{\theta},\epsilon}\alpha_i(\boldsymbol{\theta}_0,F_i) &= -\frac{1}{\mathbb{E}(\partial_{\alpha_i}V_i^{(1)})}\left[\mathbb{E}(\partial^2_{\boldsymbol{\theta},\alpha_i}V_i^{(1)})\partial_{\epsilon}\alpha_i(\boldsymbol{\theta}_0,F_i)+T^{1/2}\frac{1}{T}\left(\partial_{\boldsymbol{\theta}}V_i^{(1)}-\mathbb{E}(\partial_{\boldsymbol{\theta}}V_i^{(1)})\right)\right]\\
&\quad -\frac{1}{\mathbb{E}(\partial_{\alpha_i}V_i^{(1)})}\left[\mathbb{E}(\partial^2_{\alpha_i}V_i^{(1)})\partial_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}_0,F_i)\partial_{\epsilon}\alpha_i(\boldsymbol{\theta}_0,F_i)\right]\\
&\quad -\frac{1}{\mathbb{E}(\partial_{\alpha_i}V_i^{(1)})}\left[T^{1/2}\frac{1}{T}\left(\partial_{\alpha_i}V_i^{(1)}-\mathbb{E}(\partial_{\alpha_i}V_i^{(1)})\right)\partial_{\boldsymbol{\theta}}\alpha_i(\boldsymbol{\theta}_0,F_i)\right]\\
&= O_p(T^{-1}) \qquad (41)\\
\partial^2_{\epsilon}\alpha_i(\boldsymbol{\theta}_0,F_i) &= -\frac{1}{\mathbb{E}(\partial_{\alpha_i}V_i^{(1)})}\left[\mathbb{E}(\partial^2_{\alpha_i}V_i^{(1)})(\partial_{\epsilon}\alpha_i(\boldsymbol{\theta}_0,F_i))^2\right]\\
&\quad -\frac{1}{\mathbb{E}(\partial_{\alpha_i}V_i^{(1)})}\left[2T^{1/2}\frac{1}{T}\left(\partial_{\alpha_i}V_i^{(1)}-\mathbb{E}(\partial_{\alpha_i}V_i^{(1)})\right)\partial_{\epsilon}\alpha_i(\boldsymbol{\theta}_0,F_i)\right]\\
&= O_p(T^{-2}) \qquad (42)
\end{aligned}$$

Central limit theorems are applied to $\partial_{\alpha_i}V_i^{(1)}$ and $\partial_{\boldsymbol{\theta}}V_i^{(1)}$, so that the resulting order of magnitude calculations can be obtained.

6. We are now in a position to simplify $\boldsymbol{\theta}^{\epsilon}(0)$ and $\boldsymbol{\theta}^{\epsilon\epsilon}(0)$.

   (a) First, we find an expression for $\boldsymbol{\theta}^{\epsilon}(0)$. Calculate the expectation of every term in (36) at the true values. Note that $\boldsymbol{\theta}^{\epsilon}(0)$ do not depend on the data. Further note that (38) is already a constant while (39) depends on the data through $V_i^{(1)}$.[8] Second-order E-ancillarity implies that $\mathbb{E}(\partial_{\alpha_i}U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0}))=0$ and $\mathbb{E}(V_i^{(1)}\partial_{\alpha_i}U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0}))=0$, as in (7). As a result, the first sum in (34) is given by

$$\frac{1}{n}\sum_{i=1}^{n}\int\frac{dh_i(0)}{d\epsilon}\,dF_i=\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(\partial_{\boldsymbol{\theta}}U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0}))\right)\boldsymbol{\theta}^{\epsilon}(0) \qquad (43)$$

[8] A curious aspect of the proof in Hahn and Newey (2004) and Hahn and Kuersteiner (2011) is that they treat (39), (41), and (42) as constants yet they still depend on the data. We solve the system of equations mentioned in Step 5 and make the order of magnitude calculations explicit to take into account the latter fact.

The remaining term in (34) is given by

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\int h_i(\epsilon)\, d\Delta_{iT} &= \frac{1}{n}\sum_{i=1}^{n}\int U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0})\, d\Delta_{iT} \\ &= \frac{\sqrt{T}}{n}\sum_{i=1}^{n}\frac{1}{T}\left(U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0}) - \mathbb{E}(U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0}))\right) \quad (44)\end{aligned}$$

Define $\mathbf{I}_i$ as follows, provided integration and differentiation can be interchanged:

$$\mathbf{I}_i = \mathbb{E}[(U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0}))(U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0}))^{\mathrm{T}}] = \mathbb{E}(\partial_{\boldsymbol{\theta}} U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0})) \quad (45)$$

Thus, we have the following expression for $\boldsymbol{\theta}^{\epsilon}(0)$, whose asymptotic distribution we seek:

$$\boldsymbol{\theta}^{\epsilon}(0) = \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{I}_i\right)^{-1}\frac{\sqrt{T}}{n}\sum_{i=1}^{n}\frac{1}{T}\left(U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0}) - \mathbb{E}(U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0}))\right) \quad (46)$$

Assume that a central limit theorem holds for $U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0})$, i.e.

$$\sqrt{nT}\left(\frac{1}{nT}\sum_{i=1}^{n}U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0})\right) \xrightarrow{d} N\left(0, \lim_{n,T\to\infty}\frac{1}{nT}\sum_{i=1}^{n}\mathbf{I}_i\right) \quad (47)$$

As a consequence, we have

$$\begin{aligned}\sqrt{nT}\frac{1}{\sqrt{T}}\boldsymbol{\theta}^{\epsilon}(0) &= \left(\frac{1}{nT}\sum_{i=1}^{n}\mathbf{I}_i\right)^{-1}\left(\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0})\right) \\ &\xrightarrow{d} N\left(\mathbf{0}, \left(\lim_{n,T\to\infty}\frac{1}{nT}\sum_{i=1}^{n}\mathbf{I}_i\right)^{-1}\right) \quad (48)\end{aligned}$$

Therefore,

$$\boldsymbol{\theta}^{\epsilon}(0) = O_p(n^{-1/2}) \quad (49)$$

(b) Next we find an expression for $\boldsymbol{\theta}^{\epsilon\epsilon}(0)$. Calculate the expectation of every term in (37) at the true values while noting the orders of magnitude in (38), (39), (40), (41), (42), and (49). The boxed, double-boxed, oval-boxed and unboxed terms in (37) are $O_p(n^{-1/2})$, $O_p(T^{-1})$, $\mathbf{0}$, and $O_p(Tn^{-1})$ respectively. The first sum in (35) can now be written as

$$\frac{1}{n}\sum_{i=1}^{n}\int\frac{d^2h_i(\epsilon)}{d\epsilon^2}\, dF_i(\epsilon) = \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{I}_i\right)\boldsymbol{\theta}^{\epsilon\epsilon}(0) + O_p\left(\frac{T}{n}\right) + O_p\left(\frac{1}{T}\right) + O_p\left(\frac{1}{\sqrt{n}}\right) \quad (50)$$

After applying central limit theorems for $\partial_{\boldsymbol{\theta}} U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0})$ and $\partial_{\alpha_i} U_{i,2}(\boldsymbol{\theta}_0,\alpha_{i0})$ and

noting the order of magnitude calculations in (38), (39), and (49), the "tail" term in (35) can now be written as

$$\frac{2}{n}\sum_{i=1}^{n}\int \frac{dh_i(\epsilon)}{d\epsilon}\, d\Delta_{iT} = O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{T}\right) \tag{51}$$

As a consequence, we have

$$\begin{aligned}\sqrt{nT}\left(\frac{1}{\sqrt{T}}\right)^2 \boldsymbol{\theta}^{\epsilon\epsilon}(0) \;=\; & \left(\frac{1}{nT}\sum_{i=1}^{n}\mathbf{I}_i\right)^{-1}\left[O_p\left(\frac{1}{\sqrt{nT}}\right) + O_p\left(\frac{1}{\sqrt{T^3}}\right) + O_p\left(\sqrt{\frac{n}{T^5}}\right)\right] \\ & + \left(\frac{1}{nT}\sum_{i=1}^{n}\mathbf{I}_i\right)^{-1}\left[O_p\left(\frac{1}{\sqrt{nT^2}}\right) + O_p\left(\frac{1}{T^2}\right)\right]\end{aligned}$$

Under the conditions that $n, T \to \infty$ and $n/T \to c \in (0, \infty)$, the distribution of $\boldsymbol{\theta}^{\epsilon\epsilon}(0)$ becomes degenerate at $\mathbf{0}$.

(c) The last term in the Taylor series expansion (29) can be shown to be $o_p(1)$. This step mimics the derivation in Hahn and Kuersteiner (2011).