Individual and Time Effects in Nonlinear Panel Data Models with Large N, T

Iván Fernández-Val — BU
Martin Weidner — USC

June 28, 2010, Preliminary Version

Abstract

Fixed effects estimators of panel models can be severely biased because of the well-known incidental parameters problem. We develop analytical and jackknife bias corrections for nonlinear models with both individual and time effects. For asymptotics where the time-dimension (T) grows with the cross-sectional dimension (N), the time effects introduce additional incidental parameter bias. As the existing bias expressions apply to models with only individual effects, we derive the appropriate corrections. The basis for the corrections are asymptotic expansions of the fixed effects score and estimator for panel models with incidental parameters in both individual and time dimensions. These expansions apply to M-estimators with concave objective functions, which cover fixed effects estimators of the most popular limited dependent variable models such as logit, probit, Tobit and Poisson models. We therefore extend the use of large-T bias adjustments to an important class of models.

1 Introduction

Fixed effects estimators of panel models can be severely biased because of the well-known incidental parameters problem (Neyman and Scott (1948), Heckman (1981), Lancaster (2000), and Greene (2004)). A recent literature, surveyed in Arellano and Hahn (2005) and including Phillips and Moon (1999), Hahn and Kuersteiner (2002), Lancaster (2002), Woutersen (2002), Hahn and Kuersteiner (2004), Hahn and Newey (2004), Carro (2007), and Fernandez-Val (2009), provides a range of solutions, so-called large-T bias corrections, to reduce the incidental parameters problem in long panels. These papers derive the analytical expression of the bias (up to a certain order of the time dimension T), which can be employed to adjust the biased fixed effects estimators. While the existing large-T methods cover a large class of models with individual effects, they do not apply to panel models with individual and time effects. Time effects are important for economic modeling because they allow the researcher to control for aggregate common shocks.

We develop analytical and jackknife bias corrections for nonlinear models with *both* individual and time effects. We consider asymptotics where T grows with the individual dimension N, as an approximation to the properties of the estimators in econometric applications where T is moderately large. Under these asymptotics, the time effects introduce additional incidental parameter bias in the time dimension. As the existing bias expressions apply to models with only individual effects, we derive the appropriate correction. This correction does not correspond to sequential application of the existing corrections to each dimension. We also give bias corrections for functions of the data, parameters and individual and time effects including average marginal effects. These marginal effects are often the quantities of interest in nonlinear models.

The basis for the bias corrections are asymptotic expansions of the fixed effects score and estimator for panel models with incidental parameters in both individual and time dimensions. Bai (2009) and Moon and Weidner (2010) derive similar expansions for quasi-likelihood estimators of linear models with interactive individual and time effects. We consider nonlinear single index models, where the single index specification contains additive individual and time effects. In that case, the nonlinearity of the model introduces nonseparability between the estimators of the model and incidental parameters, so we need to deal with an infinite dimensional non-diagonal Hessian matrix. We focus on M-estimators with concave objective functions, which cover fixed effects estimators of the most popular limited dependent variable models such as logit, probit, Tobit and Poisson models (Olsen (1978), and Pratt (1981)). Our analysis therefore extends the use of large-T bias adjustments to an important class of models.

Our corrections eliminate the leading term of the bias from the asymptotic expansions. Under asymptotic sequences where N and T grow at the same rate with the sample size, we find that this term has two components: one of order $O(T^{-1})$ coming from the estimation of the individual effects; and one of order $O(N^{-1})$ coming from the estimation of the time effects. WORK IN PROGRESS: We consider analytical methods similar to Hahn and

Kuersteiner (2004) and Hahn and Newey (2004), and suitable modifications of the leave one observation out and split panel jackknife methods of Hahn and Newey (2004) and Dhaene and Jochmans (2010). These methods apply to a wide array of sampling conditions allowing for cross sectional and time series dependencies. The corrections can be implemented over the objective function, score or estimator. Simulation evidence indicates our approach works well in finite samples and an empirical example illustrates the applicability of our estimator.

Notation: We write A' for the transpose of a matrix or vector A. We use $\mathbb{1}_n$ for the $n \times n$ identity matrix, and 1_n for the column vector of length n whose entries are all unity. For a $n \times m$ matrix A, we define the projectors $\mathcal{P}_A = A(A'A)^{-1}A'$ and $\mathcal{M}_A = \mathbb{1}_n - A(A'A)^{-1}A'$, where $(A'A)^{-1}$ denotes a generalized inverse if A is not of full column rank. For square matrices B, C, we use $B > C$ (or $B \geq C$) to indicate that $B - C$ is positive (semi) definite. We use the vector norms $\|v\| = \sqrt{v'v}$ and $\|v\|_\infty = \max_i |v_i|$, the matrix infinity norm $\|A\|_\infty = \max_i \sum_j |A_{ij}|$, and the matrix maximum norm $\|A\|_{\max} = \max_{ij} |A_{ij}|$. We write wpa1 for "with probability approaching one".

2 Model and Estimators

The data consist of $N \cdot T$ observations $Y = \{Y_{it}, i = 1, ..., N, t = 1, ..., T\}$ and $X = \{X_{it}, i = 1, ..., N, t = 1, ..., T\}$, for a dependent variable Y_{it} and a K-vector of regressors X_{it}. We are interested in some feature of the conditional distribution of Y_{it} given X_{it} such as the conditional mean or some regression function. We focus on single index models where the dependence on the regressors X_{it} can be summarized by a linear index z_{it}, which takes one of the following specifications

(I) $z_{it}(\beta, \phi) = X_{it}'\beta + \alpha_i, \qquad \phi = (\alpha_1, \ldots, \alpha_N)', \qquad G_{NT} = N.$

(II) $z_{it}(\beta, \phi) = X_{it}'\beta + \alpha_i + \gamma_t, \qquad \phi = (\alpha_1, \ldots, \alpha_N, \gamma_1, \ldots, \gamma_T)', \qquad G_{NT} = N + T.$

In both specifications, β is a K-vector of regression coefficients, α_i is a time invariant individual effect that often represents individual heterogeneity, and γ_t is a common time effect that often captures aggregate shocks. Examples of single index models include linear regression, logit, probit, Tobit and Poisson models.

The parameters of interest are the regression coefficients β, while the individual and/or time effects are treated as nuisance parameters. The total number of nuisance parameters is denoted by G_{NT}, where the subscript indicates the dependence on the sample size, and the G_{NT}-vector of nuisance parameters is denoted by ϕ. The true value of the parameters, denoted by β^0 and ϕ^0, solve the population problem for every i, t, i.e.

$$\max_{\beta,\phi} \mathbb{E}\big[L\left(Y_{it}, z_{it}(\beta, \phi)\right)\big] , \tag{2.1}$$

where $\mathbb{E}$ denotes the expectation with respect to the distribution of (Y_{it}, X_{it}) conditional on the realization of ϕ^0. $L(y, z)$ is a scalar individual-period objective function. For example, in binary choice models we can set $L(y, z) = y \log F(z) + (1 - y) \log[1 - F(z)]$, where F is a

CDF. We write z_{it}^0 for the single index evaluated the true parameter values. Other quantities of interest are averages over the data and nuisance parameters

$$\Delta \;=\; \overline{\mathbb{E}}\,\mathbb{E}\left[\Delta(z_{it}^0, \beta^0, \phi^0)\right], \tag{2.2}$$

where $\overline{\mathbb{E}}$ denotes the expectation with respect to the distribution of ϕ^0. This includes average marginal effects, which are often the quantities of interest in nonlinear models. For example, in binary choice we can set $\Delta(z_{it}, \beta, \phi) = \beta F'(z_{it})$, where F' is the derivative of F. To estimate the parameters, we consider the sample version of problem (2.1), i.e.

$$\max_{\beta,\phi}\; \mathcal{L}_{NT}(\beta,\phi)\;, \qquad\qquad \mathcal{L}_{NT}(\beta,\phi) \;=\; \sum_{i=1}^{N}\sum_{t=1}^{T} L_{it}(z_{it}(\beta,\phi))\;, \tag{2.3}$$

where $L_{it}(z) = L(Y_{it}, z)$. We also use $L_{it}^{(q)}(z) = \partial^q L(Y_{it}, z)/\partial z^q$. If no argument is specified for $L_{it}^{(q)}(z)$, then it is evaluated at $z = z_{it}^0$.

In specification (II), there is an ambiguity in ϕ, because adding a constant ρ to all α_i, while subtracting it from all γ_t does not change z_{it}, i.e. for $v = (1_N', -1_T')'$ we have $z_{it}(\beta, \phi + \rho v) = z_{it}(\beta, \phi)$. To eliminate this ambiguity, we can normalize ϕ to satisfy $v'\phi = 0$, i.e. $\sum_i \alpha_i = \sum_t \gamma_t$. For our purposes, it is more convenient not to impose the constraint $v'\phi = 0$ directly, but to impose it indirectly by modifying the objective function as follows for the second specification

$$\mathcal{L}_{NT}(\beta,\phi) \;=\; \sum_{i=1}^{N}\sum_{t=1}^{T} L_{it}\big(z_{it}(\beta,\phi)\big) \;-\; b\,(v'\phi)^2/2, \tag{2.4}$$

where $b > 0$ is an arbitrary constant. The maximizer of $\mathcal{L}_{NT}(\beta, \phi)$ automatically satisfies $v'\phi = 0$, and we choose $v'\phi^0 = 0$ for the true value.

To analyze the properties of the estimator of β, it is convenient to concentrate out first the nuisance parameters. For given β, we define the optimal $\hat{\phi}(\beta)$ and the profile objective function by

$$\hat{\phi}(\beta) = \operatorname*{argmax}_{\phi\in\mathbb{R}^{G_{NT}}} \mathcal{L}_{NT}(\beta,\,\phi)\;, \qquad\qquad \mathcal{L}_{NT}(\beta) \;=\; \max_{\phi\in\mathbb{R}^{G_{NT}}} \mathcal{L}_{NT}(\beta,\,\phi)\;. \tag{2.5}$$

The fixed effects estimators of β and ϕ are

$$\hat{\beta} \;=\; \operatorname*{argmax}_{\beta\in\mathbb{R}^K} \mathcal{L}_{NT}(\beta)\;, \qquad\qquad \hat{\phi} \;=\; \hat{\phi}(\hat{\beta})\;. \tag{2.6}$$

Estimators of averages over the data and nuisance parameters can be formed as

$$\hat{\Delta} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta(z_{it}(\hat{\beta},\hat{\phi}),\hat{\beta},\hat{\phi}).$$

Let the support of Y_{it} be $\mathcal{Y}$, e.g. $\mathcal{Y} = \{0, 1\}$ for the binary choice model. We make the following assumptions

Assumption 2.1. (Strictly Exogenous Regressors)

(i) *We consider limits of sequences where* $N, T \to \infty$ *and* $N/T \to \kappa^2$, $0 < \kappa < \infty$.

(ii) *For all* $y \in \mathcal{Y}$, *the function* $L(y, z)$ *is strictly concave and four times continuously differentiable in* $z \in \mathbb{R}$.

(iii) *The regressors* $X_{it,k}$ *are non-colinear, even after subtracting the cross-sectional mean (for specification (I)), or subtracting both the cross-sectional mean and the time-average (for specification (II)).*[1]

(iv) $\|X_{it}\|$ *is uniformly bounded over* i, t, N, T.

(v) *There exists* $\mathcal{Z} \subseteq \mathbb{R}$ *and constants* $b_{\min}$ *and* $b_{\max}$ *such that* $z^0_{it} \in \mathcal{Z}$ *for all* i, t, N, T, *and for all* $y \in \mathcal{Y}$ *and* $z \in \mathcal{Z}$ *we have*

$$0 < b_{\min} < |L^{(2)}(y, z)| < b_{\max} , \quad |L^{(3)}(y, z)| < b_{\max} , \quad |L^{(4)}(y, z)| < b_{\max} .$$

(vi) *For all* i, t, N, T *we assume*

$$\mathbb{E}\left(L^{(1)}_{it} \,\middle|\, X\right) = 0 .$$

(vii) *We assume that* Y *conditional on* X *is independent across* i *and* t, *and that the eighth moment of* $L^{(1)}_{it}$ *conditional on* X *is uniformly bounded over* i, t, N, T.

Assumption 2.1(iii) rules out time-invariant regressors, and also cross-sectionally invariant regressors for the model with both effects. Assumption 2.1(vi) imposes that the regressors are strictly exogenous because we condition on X, i.e. the regressors for all the individuals and time periods. This assumption rules out lags of Y_{it} as regressors and any dynamic feedback from the dependent variable to the regressors. We also consider the following alternative assumptions to accommodate predetermined regressors.

Assumption 2.2. (Predetermined Regressors)

Let the assumptions 2.1 (i) *to* (v) *be satisfied. In addition*

(vi) *For all* i, t, N, T *we assume*

$$\mathbb{E}\left(L^{(1)}_{it} \,\middle|\, \mathcal{F}_{it}\right) = 0 ,$$

where $\mathcal{F}_{it} = \sigma(X^t_i,\ Y^{t-1}_i)$, *for* $X^t_i = (X_{it}, X_{i,t-1}, \ldots)$ *and* $Y^{t-1}_i = (Y_{i,t-1}, Y_{i,t-2}, \ldots)$.

(vii) *We assume that* (Y, X) *is independent across* i *and stationary and mixing across* t *for each* i; $\sup_i |\alpha_i(m)| \leq Ca^m$ *for some* a *such that* $0 < a < 1$ *and some* $C > 0$, *where* $\alpha_i(m) = \sup_t \sup_{A \in \mathcal{F}_{it}, B \in \bar{\mathcal{F}}_{i,t+m}} |P(A \cap B) - P(A)P(B)|$, *for* $\bar{\mathcal{F}}_{i,t} = \sigma(X_{it}, Y_{i,t-1}, X_{i,t+1}, Y_{i,t}, ...)$; *and that the eighth moment of* $L^{(1)}_{it}$ *conditional on* $\mathcal{F}_{it}$ *is uniformly bounded over* i, t, N, T.

[1] This means that the $K \times K$ matrix with entries $\mathrm{Tr}(X'_{k_1} \mathcal{M}_{1_N} X_{k_2})$, for specification (I), and $\mathrm{Tr}(X_{k_1} \mathcal{M}_{1_T} X'_{k_2} \mathcal{M}_{1_N})$, for specification (II), is non-degenerate for all N, T, where X_k is the $N \times T$ regressor matrix for each $k = 1, \ldots, K$.

Assumption 2.1(ii) implies strict concavity of the objective function $\mathcal{L}_{NT}(\beta,\phi)$ on $\mathbb{R}^K \times \mathbb{R}^{G_{NT}}$, which is convenient for the analysis of the properties of the fixed effects estimator. This assumption applies to the likelihood function of the most popular limited dependent variable models including linear, binary choice, censored and count data models. In particular, the log-likelihood function of the logit and probit model are strictly concave (Pratt, 1981), and the same applies to Tobit models under an appropriate parametrization (Olsen (1978)) and Poisson models.

Note also that the assumption of strict concavity restricts the objective function, not the model itself. To illustrate this, we introduce the conditional mean function F, and choose a concrete period objective function as follows

$$F\left(z_{it}\right) = \mathbb{E}\left(Y_{it} \,\middle|\, X_{it}\right) , \qquad\qquad L(y,z) = yz - \int_{\zeta^0}^{z} F(\zeta)\, d\zeta , \tag{2.7}$$

where ζ^0 can be chosen appropriately. For this objective function the second derivative of $L(y,z)$ with respect to z is given by $L^{(2)}(y,z) = -F'(z)$, i.e. if $F(z)$ is strictly increasing, then $L(y,z)$ is strictly concave in z. Thus, a strictly concave objective function can be chosen for a all single index models with strictly increasing conditional mean function, even if the log-likelihood function of the model is not strictly concave.

3 Asymptotic Expansions

In this section, we derive the asymptotic expansion of the score of the profile objective function around β^0. Here, we do not employ the panel or single index structure of the model, nor the particular structure of the objective function described in the last section. Instead, we consider a model with sample size NT, and objective function $\mathcal{L}_{NT}(\beta,\phi)$, with K parameters of interest β, and G_{NT} incidental parameters, and we only make use of the high-level Assumption 3.1 below. It is convenient to start by introducing some notation that will be extensively used in the analysis. Let

$$\begin{aligned}
\Psi(\beta,\phi) &= \frac{\partial \mathcal{L}_{NT}(\beta,\,\phi)}{\partial\phi} , \qquad \mathcal{H}(\beta,\phi) = \frac{\partial^2 \mathcal{L}_{NT}(\beta,\,\phi)}{\partial\phi\,\partial\phi'} , \quad \mathcal{L}^{(1,0)}(\beta,\phi) = \frac{\partial \mathcal{L}_{NT}(\beta,\phi)}{\partial\beta} , \\
\mathcal{L}^{(1,1)}(\beta,\phi) &= \frac{\partial^2 \mathcal{L}_{NT}(\beta,\phi)}{\partial\beta\partial\phi'} , \quad \mathcal{L}^{(2,0)}(\beta,\phi) = \frac{\partial^2 \mathcal{L}_{NT}(\beta,\phi)}{\partial\beta\partial\beta'} , \quad \mathcal{L}^{(0,3)}_g(\beta,\phi) = \frac{\partial^3 \mathcal{L}_{NT}(\beta,\phi)}{\partial\phi\partial\phi'\partial\phi_g}, \\
\mathcal{L}^{(1,2)}_g(\beta,\phi) &= \frac{\partial^3 \mathcal{L}_{NT}(\beta,\phi)}{\partial\beta\partial\phi'\partial\phi_g} .
\end{aligned} \tag{3.1}$$

We refer to the G_{NT} vector $\Psi(\beta,\phi)$ as incidental parameter score, and to the $G_{NT} \times G_{NT}$ matrix $\mathcal{H}(\beta,\phi)$ as incidental parameter Hessian. $\mathcal{L}^{(1,0)}(\beta,\phi)$ is a K-vector, $\mathcal{L}^{(1,1)}$ is a $K \times G_{NT}$ matrix, $\mathcal{L}^{(2,0)}(\beta,\phi)$ is a $K \times K$ matrix, and for each $g = 1,\ldots,G_{NT}$ we have a $G_{NT} \times G_{NT}$ matrix $\mathcal{L}^{(0,3)}_g(\beta,\phi)$ and a $K \times G_{NT}$ matrix $\mathcal{L}^{(1,2)}_g(\beta,\phi)$. Note that Ψ, $\mathcal{H}$, $\mathcal{L}^{(1,0)}$, $\mathcal{L}^{(1,1)}$, $\mathcal{L}^{(2,0)}$, $\mathcal{L}^{(0,3)}$ and $\mathcal{L}^{(1,2)}$ are functions of NT, but we suppress this dependence to simplify notation. We omit the argument of the functions when they are evaluated at the true parameter values $(\beta^0,\,\phi^0)$, e.g. $\mathcal{H} = \mathcal{H}(\beta^0,\phi^0)$. We use a bar to indicate expectation, e.g. $\overline{\mathcal{H}} = \mathbb{E}(\mathcal{H})$, and a tilde to denote that the variables are in deviation with respect to their expectation, e.g. $\widetilde{\mathcal{H}} = \mathcal{H} - \overline{\mathcal{H}}$. For $c \geq 0$, we also define the sets $\mathcal{B}(c, \beta^0) = \left\{\beta : \|\beta - \beta^0\|_\infty \leq c\right\}$, and $\mathcal{B}(c, \beta^0, \phi^0) = \left\{(\beta, \phi) : \|\beta - \beta^0\|_\infty < c, \ \|\phi - \phi^0\|_\infty < c\right\}$, which are the closed balls of radius c around the true parameters β^0 and (β^0, ϕ^0), respectively, under the infinity norm.

Assumption 3.1 (**Asymptotic Expansion**).

(i) $\frac{G_{NT}}{\sqrt{NT}} \to a$, $0 < a < \infty$, *as* $NT \to \infty$.

(ii) There exists $c > 0$ *such that* $\mathcal{L}_{NT}(\beta, \phi)$ *is four times continuously differentiable in* $\mathcal{B}(c, \beta^0, \phi^0)$, *and for all non-negative integers* p, q *with* $2 \leq p + q \leq 4$ *and* $k_1, \ldots, k_p \in \{1, \ldots, K\}$ *we have*[2]

$$\sup_{(\beta,\phi)\in\mathcal{B}(c,\beta^0,\phi^0)} \ \max_{g_1\in\{1,\ldots,G_{NT}\}} \sum_{g_2,\ldots,g_q=1}^{G_{NT}} \left| \frac{\partial^{p+q}\mathcal{L}_{NT}(\beta, \phi)}{\partial\beta_{k_1}\ldots\partial\beta_{k_p}\partial\phi_{g_1}\ldots\partial\phi_{g_q}} \right| = \mathcal{O}_P\left(\sqrt{NT}\right) .$$

(iii) There exists α, *with* $0 < \alpha < 1/12$, *such that*

$$\|\Psi\|_\infty = \mathcal{O}_P\left((NT)^{1/4+\alpha}\right) , \qquad \left\|\overline{\mathcal{H}}^{-1}\right\|_\infty = \mathcal{O}\left((NT)^{-1/2}\right) , \qquad \widetilde{\mathcal{L}}^{(2,0)} = o_P(NT) ,$$
$$\left\|\widetilde{\mathcal{H}}\right\|_\infty = \mathcal{O}_P\left((NT)^{1/4+\alpha}\right) , \quad \left\|\widetilde{\mathcal{L}}^{(1,1)}\right\|_{\max} = \mathcal{O}_P\left((NT)^{1/4+\alpha}\right) ,$$

and

$$\sum_{g=1}^{G_{NT}} \left\|\widetilde{\mathcal{L}}_g^{(0,3)}\right\|_\infty = \mathcal{O}_P\left((NT)^{1/4+\alpha}\right), \quad \max_{g\in\{1,\ldots,G_{NT}\}} \left\|\widetilde{\mathcal{L}}_g^{(1,2)}\right\|_\infty = \mathcal{O}_P\left((NT)^{1/4+\alpha}\right) .$$

Assumption 3.2 (**Uniform Consistency of** $\hat{\phi}$)**.** *For all deterministic series* $\eta_{NT} > 0$, *with* $\eta_{NT} \to 0$ *as* $NT \to \infty$, *we assume* $\sup_{\beta\in\mathcal{B}(\eta_{NT},\beta^0)} \left\|\hat{\phi}(\beta) - \phi^0\right\| = o_P(1)$.

Our assumptions on the single index models that were introduced in the last section are sufficient for Assumption 3.1 and 3.2 (see Lemma 4.1 below). More general models that satisfy Assumption 3.1 and 3.2 are conceivable however. In particular, models without concave objective function can still satisfy the assumptions imposed here. Strict concavity of the objective function is convenient to show consistency of $\hat{\beta}$ and uniform consistency of $\hat{\phi}$ (see Theorem 3.3 below), but if alternative consistency proofs are available, then no shape restriction for the objective functions is required.

Theorem 3.1. *Let Assumption 3.1 and 3.2 hold, and assume that there exists* $c > 0$ *such that* $\mathcal{L}_{NT}(\beta, \phi)$ *is four times continuously differentiable in* $\mathcal{B}(c, \beta^0, \phi^0)$. *Then we have*

$$\hat{\phi}(\beta) - \phi^0 = -\overline{\mathcal{H}}^{-1}\Psi + \overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\Psi - \tfrac{1}{2}\overline{\mathcal{H}}^{-1} \sum_{g=1}^{G_{NT}} \overline{\mathcal{L}}_g^{(0,3)}\overline{\mathcal{H}}^{-1}\Psi[\overline{\mathcal{H}}^{-1}\Psi]_g - \overline{\mathcal{H}}^{-1}\overline{\mathcal{L}}^{(1,1)\prime}(\beta - \beta^0) + r_{NT}(\beta),$$

[2] Note that for $q = 0$ both the maximum and the sum over g's drop out, and we only take the supremum over $\left|\frac{\partial^p\mathcal{L}_{NT}(\beta,\phi)}{\partial\beta_{k_1}\ldots\partial\beta_{k_p}}\right|$. For $q = 1$ we have no sum, but take the maximum over g_1.

and

$$\frac{1}{\sqrt{NT}} \frac{\partial \mathcal{L}_{NT}(\beta)}{\partial \beta} = S_{NT} - W_{NT} \sqrt{NT}(\beta - \beta^0) + R_{NT}(\beta),$$

where $S_{NT} = S_{NT}^{(0)} + S_{NT}^{(1)} + S_{NT}^{(2)}$ *and*

$$\begin{aligned}
W_{NT} &= -\frac{1}{NT} \left(\overline{\mathcal{L}}^{(2,0)} - \overline{\mathcal{L}}^{(1,1)} \, \overline{\mathcal{H}}^{-1} \, \overline{\mathcal{L}}^{(1,1)\prime} \right), \\
S_{NT}^{(0)} &= \frac{1}{\sqrt{NT}} \left(\mathcal{L}^{(1,0)} - \overline{\mathcal{L}}^{(1,1)} \, \overline{\mathcal{H}}^{-1} \Psi \right), \\
S_{NT}^{(1)} &= \frac{1}{\sqrt{NT}} \left(\overline{\mathcal{L}}^{(1,1)} \, \overline{\mathcal{H}}^{-1} \, \widetilde{\mathcal{H}} \, \overline{\mathcal{H}}^{-1} \, \Psi - \widetilde{\mathcal{L}}^{(1,1)} \overline{\mathcal{H}}^{-1} \Psi \right), \\
S_{NT}^{(2)} &= \frac{1}{2\sqrt{NT}} \sum_{g=1}^{G_{NT}} \left(\overline{\mathcal{L}}_g^{(1,2)} - \overline{\mathcal{L}}^{(1,1)} \, \overline{\mathcal{H}}^{-1} \overline{\mathcal{L}}_g^{(0,3)} \right) [\overline{\mathcal{H}}^{-1} \Psi]_g \overline{\mathcal{H}}^{-1} \Psi,
\end{aligned}$$

and the remainder terms $r_{NT}(\beta)$ *and* $R_{NT}(\beta)$ *satisfy for all series* $\eta_{NT} > 0$ *with* $\eta_{NT} \to 0$

$$\sup_{\beta \in \mathcal{B}(\eta_{NT}, \beta^0)} \frac{\sqrt{NT} \, \|r_{NT}(\beta)\|_\infty}{1 + \sqrt{NT} \|\beta - \beta^0\|} = o_P(1) \ , \qquad \sup_{\beta \in \mathcal{B}(\eta_{NT}, \beta^0)} \frac{R_{NT}(\beta)}{1 + \sqrt{NT} \|\beta - \beta^0\|} = o_P(1) \ .$$

Theorem 3.1 provides asymptotic expansions of the incidental parameter estimator $\hat{\phi}(\beta)$ and of the profile objective score function $\partial \mathcal{L}_{NT}(\beta)/\partial \beta$. These are joint expansions in $\beta - \beta^0$ up to linear order and in the incidental parameter score Ψ up to quadratic order. Bounds on the remainder terms of the expansions are given, which make the expansions applicable for β values within a shrinking neighborhood of β^0. Thus, if consistency of $\hat{\beta}$ is known, then the asymptotic expansion of the profile objective score can be applied to the first order condition for $\hat{\beta}$, which reads $\partial \mathcal{L}_{NT}(\hat{\beta})/\partial \beta = 0$. This gives rise to the following corollary.

Corollary 3.2. *Let the assumptions of Theorem 3.1 be satisfied. Let* $\operatorname{plim}_{NT \to \infty} W_{NT} > 0$ *and* $\|\hat{\beta} - \beta^0\| = o_P(1)$. *Then we have* $\sqrt{NT}(\hat{\beta} - \beta^0) = W_{NT}^{-1} S_{NT} + o_P(1)$.

Using this corollary we can derive the first order asymptotic theory of $\hat{\beta}$ from the limiting distributions of the approximated Hessian W_{NT} and the approximated score S_{NT}. However, before we can apply the corollary we need to show consistency of $\hat{\beta}$ and uniform consistency of $\hat{\phi}$ (Assumption 3.2). The following theorem provides these consistency results for the case of strictly concave objective function $\mathcal{L}_{NT}(\beta, \phi)$.

Theorem 3.3. *Let Assumption 3.1 hold, let* $\mathcal{L}_{NT}(\beta, \phi)$ *be four times continuously differentiable and strictly concave over* $\mathbb{R}^K \times \mathbb{R}^{G_{NT}}$, *and let* $\operatorname{plim}_{NT \to \infty} W_{NT} > 0$. *Then Assumption 3.2 is satisfied and we have* $\|\hat{\beta} - \beta^0\| = o_P(1)$. *From Corollary 3.2 we can therefore conclude that* $\sqrt{NT}(\hat{\beta} - \beta^0) = W_{NT}^{-1} S_{NT} + o_P(1)$.

Thus, for strictly concave objective function no extra consistency proof for $\hat{\phi}(\beta)$ and $\hat{\beta}$ is required to derive the first order asymptotic theory of $\hat{\beta}$.

4 Application of the Expansion to Single Index Models

We apply the general expansion results of the last section to the the single index models that were introduced in Section 2. The following lemma provides the link between the two sections.

Lemma 4.1. *Consider the single index model with either only individual specific effects (specification (I)) or individual and time effects (specification (II)). Let either Assumption 2.1 or Assumption 2.2 hold. Then Assumption 3.1 and Assumption 3.2 are satisfied, and* $\mathcal{L}_{NT}(\beta,\phi)$ *is four times continuously differentiable and strictly concave over* $\mathbb{R}^K \times \mathbb{R}^{G_{NT}}$.

The lemma guarantees that Theorem 3.3 is applicable to the single index models. To obtain the asymptotic distribution of $\hat{\beta}$ we need to establish the limit of $W_{NT}^{-1} S_{NT}$. The particular form of the approximate Hessian W_{NT} and the approximate score S_{NT} for the single index model is presented below. The probability limit of W_{NT} is assumed to exist and to be non-degenerate — which could be justified further by specifying a data generating process for the regressors, and for the fixed and time effects. Establishing the asymptotics of the approximate score S_{NT} is more involved. In the following, it is convenient to separately consider the model with only individual effects and the model with both individual and time effects.

4.1 Only Individual Specific Effects

Here, we consider the single index model with only individual specific effects. The theory for this case was already developed previously, see e.g. Arellano and Hahn (2007) for a review of the literature. We still work out this case explicitly with our methodology, in order to have a benchmark for comparison when considering the more general case with also time effects below.

In specification (I) we have $\phi = \alpha$. The objective function $\mathcal{L}(\beta,\alpha) = \mathcal{L}(\beta,\phi)$ was defined in equation (2.3). Its partial derivatives read

$$\frac{\partial^{p+q}\mathcal{L}_{NT}(\beta,\,\alpha)}{\partial\beta_{k_1}\ldots\partial\beta_{k_p}\partial\alpha_{i_1}\ldots\partial\alpha_{i_q}} = \begin{cases} \sum_{t=1}^{T} X_{it,k_1}X_{it,k_2}\ldots X_{it,k_p}\, L_{it}^{(p+q)} & \text{if } i_1 = i_2 = \ldots = i_q, \\ 0 & \text{otherwise.} \end{cases} \tag{4.1}$$

For $i \neq j$ a partial derivative with respect to both α_i and α_j is zero, since the data for each unit i,t only depends on exactly one incidental parameter. Applying equation (4.1) we find that, evaluated at the true parameters, the elements of the incidental parameter score read $\Psi_i = \sum_{t=1}^{T} L_{it}^{(1)}$, and the incidental parameter Hessian is now a diagonal $N \times N$ matrix with diagonal elements $\mathcal{H}_{ii} = \sum_{t=1}^{T} L_{it}^{(2)}$. It is also convenient to define the K-vector $\mathcal{X}_i$ for $i = 1,\ldots,N$ by[3]

$$\mathcal{X}_i = \frac{\sum_{t=1}^{T} \mathbb{E}\left(X_{it}\, L_{it}^{(2)}\right)}{\sum_{t=1}^{T} \mathbb{E}\left(L_{it}^{(2)}\right)}. \tag{4.2}$$

[3]For the $K \times N$ matrix $\mathcal{X} = (\mathcal{X}_1,\ldots,\mathcal{X}_N)$ we have $\mathcal{X} = \overline{\mathcal{L}}^{(1,1)}\,\overline{\mathcal{H}}^{-1}$.

The $\mathcal{X}_i$ are linear projections of the regressor X_{it} under a metric defined by $L_{it}^{(2)}$ and the expectation operator, namely for every $k = 1, \ldots, K$ we have

$$\mathcal{X}_{i,k} = \underset{\alpha_i \in \mathbb{R}}{\operatorname{argmin}} \sum_{t=1}^{T} \mathbb{E}\left[L_{it}^{(2)} \, (X_{it,k} - \alpha_i)^2 \right]. \tag{4.3}$$

In the linear model with additive individual effects we have $L_{it}^{(2)} = -1$ and thus $\mathcal{X}_i = \sum_t \mathbb{E} X_{it}/T$, so that $\mathbb{E}(X_{it}) - \mathcal{X}_i$ is the expected value of the familiar fixed effects transformation of the regressors.

We can now express the approximated Hessian W_{NT} and the terms in the approximated score $S_{NT} = S_{NT}^{(0)} + S_{NT}^{(1)} + S_{NT}^{(2)}$, defined in Theorem 3.1, as follows

$$\begin{aligned}
W_{NT} &= -\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}\left[L_{it}^{(2)} \left(X_{it} - \mathcal{X}_i\right) \left(X_{it} - \mathcal{X}_i\right)' \right], \\
S_{NT}^{(0)} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} L_{it}^{(1)} \left(X_{it} - \mathcal{X}_i\right), \\
S_{NT}^{(1)} &= -\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \mathcal{H}_{ii}^{-1} \, \Psi_i \, \sum_{t=1}^{T} L_{it}^{(2)} \left(X_{it} - \mathcal{X}_i\right), \\
S_{NT}^{(2)} &= \frac{1}{2\sqrt{NT}} \sum_{i=1}^{N} \mathcal{H}_{ii}^{-2} \, \Psi_i^2 \, \sum_{t=1}^{T} \mathbb{E}\left[L_{it}^{(3)} \, \left(X_{it} - \mathcal{X}_i\right) \right].
\end{aligned} \tag{4.4}$$

Next, we derive the asymptotics of these terms and then apply Corollary 3.2 to obtain the asymptotic distribution of $\hat{\beta}$. The result of this analysis is summarized in the following theorem.

Theorem 4.2. *Assume that either Assumption 2.1 [strictly exogenous regressors] or Assumption 2.2 [predetermined regressors] holds and that the following limits exist*

$$W = \lim_{N,T\to\infty} W_{NT} \, ,$$

$$V = \lim_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}\left[\left(L_{it}^{(1)}\right)^2 \, \left(X_{it} - \mathcal{X}_i\right) \left(X_{it} - \mathcal{X}_i\right)' \right] \, ,$$

$$B_1 = \begin{cases} -\displaystyle\lim_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[L_{it}^{(1)} L_{it}^{(2)} \left(X_{it} - \mathcal{X}_i\right) \right]}{\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left(L_{it}^{(2)} \right)}, & \textit{for strictly exogenous regr.,} \\[2ex] -\displaystyle\lim_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{\frac{1}{T}\sum_{t=1}^{T} \sum_{\tau=t}^{T} \mathbb{E}\left[L_{it}^{(1)} L_{i\tau}^{(2)} \left(X_{i\tau} - \mathcal{X}_i\right) \right]}{\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left(L_{it}^{(2)} \right)}, & \textit{for predetermined regr.,} \end{cases}$$

$$B_2 = \frac{1}{2} \lim_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{\left\{ \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[\left(L_{it}^{(1)}\right)^2 \right] \right\} \left\{ \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[L_{it}^{(3)} \, \left(X_{it} - \mathcal{X}_i\right) \right] \right\}}{\left(\frac{1}{T}\sum_{t=1}^{T} L_{it}^{(2)} \right)^2} \, ,$$

and let $W > 0$. We then have

$$S_{NT}^{(0)} \underset{d}{\longrightarrow} \mathcal{N}\left(0, \; V\,\right) \, , \qquad S_{NT}^{(1)} \underset{p}{\longrightarrow} \kappa \, B_1 \, , \qquad S_{NT}^{(2)} \underset{p}{\longrightarrow} \kappa \, B_2 \, ,$$

and therefore

$$\sqrt{NT}\left(\hat{\beta}-\beta^0\right) \underset{d}{\longrightarrow} \mathcal{N}\left[\,\kappa\, W^{-1}\,(B_1+B_2)\,,\; W^{-1}\,V\,W^{-1}\right]\;.$$

If the objective function is the true likelihood function, then the information equality holds, which reads $\mathbb{E}\left[\left(L^{(1)}_{it}\right)^2 \middle| X\right] = -\mathbb{E}\left(L^{(2)}_{it}\middle| X\right)$ for strictly exogenous regressors and $\mathbb{E}\left[\left(L^{(1)}_{it}\right)^2 \middle| \mathcal{F}_{it}\right] = -\mathbb{E}\left(L^{(2)}_{it}\middle| \mathcal{F}_{it}\right)$ for predetermined regressors. In that case we have $V = W$, and the asymptotic variance of $\hat{\beta}$ is simply W^{-1}.

4.2 Individual and Time Specific Effects

We now consider specification (II), where we have both individual and time effects such that the incidental parameter vector ϕ is $N+T$ dimensional. We use the modified objective function (2.4), where we subtracted off the term $b(v'\phi)^2/2$, with $(N+T)$-vector $v = (1'_N, -1'_T)'$ and some contant $b > 0$. As explained above, this modification is just a convenient way to resolve the degeneracy of the incidental parameter space, without directly imposing a constraint on the incidental parameters. The incidental parameter score and the incidental parameter Hessian evaluated at the true parameters are

$$\Psi = \begin{pmatrix} \left[\sum_{t=1}^T L^{(1)}_{it}\right]_{i=1,\ldots,N} \\ \left[\sum_{i=1}^N L^{(1)}_{it}\right]_{t=1,\ldots,T} \end{pmatrix}\,,$$

$$\mathcal{H} = \begin{pmatrix} \operatorname{diag}\left\{\left[\sum_{t=1}^T L^{(2)}_{it}\right]_{i=1,\ldots,N}\right\} & \left[L^{(2)}_{it}\right]_{\substack{i=1,\ldots,N \\ t=1,\ldots,T}} \\ \left[L^{(2)}_{it}\right]_{\substack{t=1,\ldots,T \\ i=1,\ldots,N}} & \operatorname{diag}\left\{\left[\sum_{i=1}^N L^{(2)}_{it}\right]_{t=1,\ldots,T}\right\} \end{pmatrix} - b\,vv'\,. \tag{4.5}$$

Note that without the additional term $b\,vv'$ the incidental parameter Hessian $\mathcal{H}$ and its expectation $\overline{\mathcal{H}}$ would not be invertible[4]. We introduce the following notation for the different blocks of the inverse of $\overline{\mathcal{H}}$

$$\overline{\mathcal{H}}^{-1} = \begin{pmatrix} \overline{\mathcal{H}}^{-1}_{\alpha\alpha} & \overline{\mathcal{H}}^{-1}_{\alpha\gamma} \\ \overline{\mathcal{H}}^{-1}_{\gamma\alpha} & \overline{\mathcal{H}}^{-1}_{\gamma\gamma} \end{pmatrix}\,, \tag{4.6}$$

where $\overline{\mathcal{H}}^{-1}_{\alpha\alpha}$, $\overline{\mathcal{H}}^{-1}_{\alpha\gamma}$, $\overline{\mathcal{H}}^{-1}_{\gamma\alpha}$, and $\overline{\mathcal{H}}^{-1}_{\gamma\gamma}$ are $N\times N$, $N\times T$, $T\times N$ and $T\times T$ matrices, respectively. The analogue of the projected regressor $\mathcal{X}_i$ in the case of only individual specific effects is now given by

$$\mathcal{X}_{it} = \sum_{j=1}^N\sum_{\tau=1}^T \left(\overline{\mathcal{H}}^{-1}_{\alpha\alpha,ij} + \overline{\mathcal{H}}^{-1}_{\gamma\alpha,tj} + \overline{\mathcal{H}}^{-1}_{\alpha\gamma,i\tau} + \overline{\mathcal{H}}^{-1}_{\gamma\gamma,t\tau}\right)\,\mathbb{E}\left(L^{(2)}_{j\tau}\,X_{j\tau}\right)\,. \tag{4.7}$$

Again, this projected regressor matrix describes a linear projection of X_{it}, which is induced by the incidental parameters under a metric given by $L^{(2)}_{it}$ and the expectation operator,

[4]We would then have $\mathcal{H}v = 0$ and $\overline{\mathcal{H}}v = 0$, i.e. both $\mathcal{H}$ and $\overline{\mathcal{H}}$ would have a non-trivial Kernel.

namely for all $k = 1, \ldots, K$ we have[5]

$$\mathcal{X}_{it,k} = \alpha^{(X)}_{i,k} + \gamma^{(X)}_{t,k}, \quad \text{with} \quad \left(\alpha^{(X)}_{i,k}, \gamma^{(X)}_{t,k}\right) = \underset{\alpha,\gamma}{\operatorname{argmin}} \sum_{i,t} \mathbb{E}\left[L^{(2)}_{it} \left(X_{it,k} - \alpha_i - \gamma_t\right)^2 \right]. \quad (4.8)$$

For the linear model with additive individual and time effects we have $L^{(2)}_{it} = -1$ and therefore $\mathcal{X}_{it} = \mathbb{E}\left(\sum_t X_{it}/T + \sum_i X_{it}/N - \sum_i \sum_t X_{it}/NT\right)$, so that $\mathbb{E}(X_{it}) - \mathcal{X}_{it}$ is the expected value of the familiar fixed effects transformation for this model. Similar to the projection $\mathcal{X}_{it}$ of X_{it} it is also convenient to introduce the projection of $L^{(1)}_{it}$ as

$$\Lambda_{it} = \sum_{j=1}^{N} \sum_{\tau=1}^{T} \left(\overline{\mathcal{H}}^{-1}_{\alpha\alpha,ij} + \overline{\mathcal{H}}^{-1}_{\gamma\alpha,tj} + \overline{\mathcal{H}}^{-1}_{\alpha\gamma,i\tau} + \overline{\mathcal{H}}^{-1}_{\gamma\gamma,t\tau} \right) L^{(1)}_{j\tau}. \quad (4.9)$$

Using these definitions we can express the approximated Hessian W_{NT} and the terms of the approximated score $S_{NT} = S^{(0)}_{NT} + S^{(1)}_{NT} + S^{(2)}_{NT}$ as follows

$$\begin{aligned}
W_{NT} &= -\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}\left[L^{(2)}_{it} \left(X_{it} - \mathcal{X}_{it}\right) \left(X_{it} - \mathcal{X}_{it}\right)' \right], \\
S^{(0)}_{NT} &= \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} L^{(1)}_{it} \left(X_{it} - \mathcal{X}_{it}\right), \\
S^{(1)}_{NT} &= -\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \Lambda_{it} L^{(2)}_{it} \left(X_{it} - \mathcal{X}_{it}\right), \\
S^{(2)}_{NT} &= \frac{1}{2\sqrt{NT}} \sum_{i=1}^{N} \sum_{j=1}^{T} \Lambda^2_{it} \mathbb{E}\left[L^{(3)}_{it} \left(X_{it} - \mathcal{X}_{it}\right) \right]. \quad (4.10)
\end{aligned}$$

Compared to the expressions in equation (4.4) for only individual specific effects, we find that the structure of W_{NT} and S_{NT} is very similar here, only $\mathcal{X}_i$ is replaced by $\mathcal{X}_{it}$, and $\mathcal{H}^{-1}_{ii}\Psi_i$ is replaced by Λ_{it}. The evaluation of $S^{(0)}_{NT}$, which contributes the asymptotic variance of $\hat{\beta}$, is therefore essentially unchanged relative to the case of only individual specific effects. However, to evaluate the terms $S^{(1)}_{NT}$ and $S^{(2)}_{NT}$, which contribute the asymptotic bias of $\hat{\beta}$, we need further knowledge about Λ_{it}, which in turn requires further insights into the structure of the inverse expected incidental parameter Hessian $\overline{\mathcal{H}}^{-1}$.

Lemma 4.3. *Let assumption 2.1 or 2.2 be satisfied. We then have*

$$\overline{\mathcal{H}}^{-1} = \begin{pmatrix} \operatorname{diag}\left\{ \left[\sum_{t=1}^{T} \mathbb{E}\left(L^{(2)}_{it}\right) \right]_{i=1,\ldots,N} \right\} & 0_{N\times T} \\ 0_{T\times N} & \operatorname{diag}\left\{ \left[\sum_{i=1}^{N} \mathbb{E}\left(L^{(2)}_{it}\right) \right]_{t=1,\ldots,T} \right\} \end{pmatrix}^{-1} + \mathcal{R},$$

where the $(N+T)\times(N+T)$ matrix $\mathcal{R}$ satisfies $\|\mathcal{R}\|_{\max} = \mathcal{O}\left(N^{-2}\right)$.

[5] Solving the optimization problem gives

$$\begin{pmatrix} \alpha^{(X)}_k \\ \gamma^{(X)}_k \end{pmatrix} = \overline{\mathcal{H}}^{-1} \begin{pmatrix} \left[\sum_t \mathbb{E}\left(L^{(2)}_{it} X_{it,k}\right) \right]_{i=1,\ldots,N} \\ \left[\sum_i \mathbb{E}\left(L^{(2)}_{it} X_{it,k}\right) \right]_{t=1,\ldots,T} \end{pmatrix},$$

from which equation (4.7) for $\mathcal{X}_{it,k} = \alpha^{(X)}_{i,k} + \gamma^{(X)}_{t,k}$ is obtained directly.

The lemma states that $\overline{\mathcal{H}}^{-1}$ can be approximated by a diagonal matrix, whose entries are of order N^{-1}, while the entries of the remainder $\mathcal{R}$ are uniformly bounded of order N^{-2}. Applying this result to the definition of Λ_{it} we find that

$$\Lambda_{it} = \frac{\sum_{\tau=1}^{T} L_{i\tau}^{(1)}}{\sum_{\tau=1}^{T} \mathbb{E}\left(L_{i\tau}^{(2)}\right)} + \frac{\sum_{j=1}^{N} L_{jt}^{(1)}}{\sum_{j=1}^{N} \mathbb{E}\left(L_{jt}^{(2)}\right)} + \text{terms that stem from } \mathcal{R}.$$

One can show that the terms that stem from the remainder $\mathcal{R}$ only contribute asymptotically vanishing terms to $S_{NT}^{(1)}$ and $S_{NT}^{(2)}$, i.e. only the two leading terms in Λ_{it} need to be considered further. The first term in Λ_{it} is equal to $\mathcal{H}_{ii}^{-1}\Psi_i$ from the previous subsection, i.e. this term gives contributions to the asymptotic bias of $\hat{\beta}$ that are equivalent to those in the model with only individual specific effects (apart from replacing $\mathcal{X}_i$ by $\mathcal{X}_{it}$). The second term in Λ_{it} gives additional contributions to the asymptotic bias of $\hat{\beta}$ that are not present in the model with only individual specific effects.

Theorem 4.4. *Assume that either Assumption 2.1 [strictly exogenous regressors] or Assumption 2.2 [predetermined regressors] holds and that the following limits exist*

$$W = \lim_{N,T\to\infty} W_{NT} ,$$

$$V = \lim_{N,T\to\infty} \frac{1}{NT} \sum_{i=1}^{N}\sum_{t=1}^{T} \mathbb{E}\left[\left(L_{it}^{(1)}\right)^2 (X_{it} - \mathcal{X}_{it})(X_{it} - \mathcal{X}_{it})'\right] ,$$

$$B_1 = \begin{cases} -\lim_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[L_{it}^{(1)} L_{it}^{(2)} (X_{it} - \mathcal{X}_{it})\right]}{\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left(L_{it}^{(2)}\right)}, & \textit{for strictly exogenous regr.,} \\ -\lim_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{\frac{1}{T}\sum_{t=1}^{T}\sum_{\tau=t}^{T} \mathbb{E}\left[L_{it}^{(1)} L_{i\tau}^{(2)} (X_{i\tau} - \mathcal{X}_{it})\right]}{\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left(L_{it}^{(2)}\right)}, & \textit{for predetermined regr.,} \end{cases}$$

$$B_2 = \frac{1}{2} \lim_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{\left\{\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[\left(L_{it}^{(1)}\right)^2\right]\right\}\left\{\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[L_{it}^{(3)} (X_{it} - \mathcal{X}_{it})\right]\right\}}{\left(\frac{1}{T}\sum_{t=1}^{T} L_{it}^{(2)}\right)^2} ,$$

$$D_1 = -\lim_{N,T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \frac{\frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\left[L_{it}^{(1)} L_{it}^{(2)} (X_{it} - \mathcal{X}_{it})\right]}{\frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\left(L_{it}^{(2)}\right)},$$

$$D_2 = \frac{1}{2} \lim_{N,T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \frac{\left\{\frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\left[\left(L_{it}^{(1)}\right)^2\right]\right\}\left\{\frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\left[L_{it}^{(3)} (X_{it} - \mathcal{X}_{it})\right]\right\}}{\left(\frac{1}{N}\sum_{i=1}^{N} L_{it}^{(2)}\right)^2} ,$$

and let $W > 0$. *We then have*

$$S_{NT}^{(0)} \underset{d}{\longrightarrow} \mathcal{N}(0,\ V) , \qquad S_{NT}^{(1)} \underset{p}{\longrightarrow} \kappa B_1 + \kappa^{-1} D_1 , \qquad S_{NT}^{(2)} \underset{p}{\longrightarrow} \kappa B_2 + \kappa^{-1} D_2 ,$$

and therefore

$$\sqrt{NT}\left(\hat{\beta} - \beta^0\right) \underset{d}{\longrightarrow} \mathcal{N}\left[W^{-1}\left(\kappa B_1 + \kappa B_2 + \kappa^{-1} D_1 + \kappa^{-1} D_2\right),\ W^{-1}\, V\, W^{-1}\right] .$$

The structure of the bias terms B_1 and B_2 is the same as for the case of only fixed effects, but the projected regressor $\mathcal{X}_{it}$ is different. The structure of D_1 and D_2 is symmetric to that of B_1 and B_2, with the role of time and cross-sectional dimensions interchanged.

5 Bias Corrections

The results of the previous sections show that the asymptotic distribution of the fixed effects estimator has a bias of the same order as the asymptotic variance under sequences where T grows at the same rate as N. This is the large-T version of the incidental parameters problem that invalidates any inference based on the asymptotic distribution. In this section we consider analytical and jackknife bias corrections. We focus on the single index model with individual and time effects, because the existing bias corrections do not cover this model.

The analytical bias correction consists of removing an estimate of the expression of the bias given in Theorem 4.4 from the fixed effect estimator of β. Hence, it only requires estimating the Hessian matrix W and the bias of the score $\mathcal{B} = \kappa B_1 + \kappa B_2 + \kappa^{-1} D_1 + \kappa^{-1} D_2$. The components of $\mathcal{B}$ and W can be estimated using sample analogs evaluated at the fixed effects estimates of β and ϕ. The bias-corrected estimator is then

$$\tilde{\beta}_A = \hat{\beta} - \hat{W}^{-1} \left(\hat{B}_1/T + \hat{B}_2/T + \hat{D}_1/N + \hat{D}_2/N\right). \tag{5.1}$$

Details and consistency of $\hat{W}$, $\hat{B}_{1/2}$, $\hat{D}_{1/2}$ is work in progress.

For the Jackknife bias correction, note that the probability limit of the fixed effects estimator of β has the expansion

$$\operatorname*{plim}_{N,T\to\infty} \hat{\beta} = \beta_{NT} = \beta^0 + B/T + D/N + o(T^{-1}), \tag{5.2}$$

where $B = W^{-1}(B_1 + B_2)$ and $D = W^{-1}(D_1 + D_2)$. We consider two versions of the Jackknife: one is based on applying a version of the leave one observation out panel jackknife of Hahn and Newey (2004) to both the individual and time dimension, and the other is based on applying a suitable combination of the leave one observation out to the individual dimension and the split panel of Dhaene and Jochmans (2010) to the time dimension. The first method applies to panels where there is no cross sectional and time series dependencies, whereas the second methods allows for time series dependencies.

To describe the first jackknife correction, let $\bar{\beta}_{N-1,T}$ be the average of the N jackknife estimators that leave out one individual, $\bar{\beta}_{N,T-1}$ be the average of the T jackknife estimators that leave out one time period, and $\bar{\beta}_{N-1,T-1}$ be the average of the $N \cdot T$ jackknife estimators that leave out one individual and one time period. The bias corrected estimator is

$$\tilde{\beta}_{J1} = NT\hat{\beta} - (N-1)T\bar{\beta}_{N-1,T} - N(T-1)\bar{\beta}_{N,T-1} + (N-1)(T-1)\bar{\beta}_{N-1,T-1}.$$

To give some intuition about how the corrections works, note that

$$
\begin{aligned}
\text{plim}_{NT\to\infty}\tilde{\beta}_{J1} &= [NT-(N-1)T-N(T-1)+(N-1)(T-1)]\beta^0 \\
&\quad + [N-(N-1)-N+(N-1)]B + [T-T-(T-1)+T-1]D \\
&\quad + [NT-(N-1)T-N(T-1)+(N-1)(T-1)]o(T^{-1}) = \beta^0 + o(T^{-1}).
\end{aligned}
$$

where we use the expansion (5.2) under suitable assumptions in the remainder term.

To describe the second jackknife correction, let $\bar{\beta}_{N/2,T/2}$ be the average of the split jackknife estimators that leave out half of the cross-sectional units and the first or second halve of the time periods. The bias corrected estimator is

$$
\tilde{\beta}_{J2} = 2\hat{\beta} - \bar{\beta}_{N/2,T/2}.
$$

To give some intuition about how the corrections works, note that

$$
\text{plim}_{NT\to\infty}\tilde{\beta}_{J2} = (2-1)\beta^0 + [2/T-1/(T/2)]B + [2/N-1/(N/2)]D + o(T^{-1}) = \beta^0 + o(T^{-1}).
$$

where we use the expansion (5.2) under suitable assumptions on the remainder term.

A Proofs for Section 3 (Expansion)

Lemma A.1. *Let assumption 3.1 be satisfied. Then for all ϕ_{NT} with $\|\phi_{NT}-\phi^0\|_\infty = o_P(1)$ we have*

$$
\begin{aligned}
\sum_{g=1}^{G_{NT}} \left\| \mathcal{L}_g^{(0,3)}(\beta^0,\phi_{NT}) - \overline{\mathcal{L}}_g^{(0,3)} \right\|_\infty &= \mathcal{O}_P\left((NT)^{1/4+\alpha}\right) + \mathcal{O}_P\left((NT)^{1/2}\|\phi_{NT}-\phi^0\|_\infty\right) , \\
\max_{g=1\ldots G_{NT}} \left\| \mathcal{L}_g^{(1,2)}(\beta^0,\phi_{NT}) - \overline{\mathcal{L}}_g^{(1,2)} \right\|_\infty &= \mathcal{O}_P\left((NT)^{1/4+\alpha}\right) + \mathcal{O}_P\left((NT)^{1/2}\|\phi_{NT}-\phi^0\|_\infty\right) .
\end{aligned}
$$

Furthermore, for all series $\eta_{NT}>0$ with $\eta_{NT}\to 0$

$$
\begin{aligned}
\sup_{(\beta,\phi)\in\mathcal{B}(\eta_{NT},\beta^0,\phi^0)} \left\| \mathcal{L}^{(1,1)}(\beta,\phi) - \overline{\mathcal{L}}^{(1,1)} \right\|_{\max} &= o_P\left(\sqrt{NT}\right) , \\
\sup_{(\beta,\phi)\in\mathcal{B}(\eta_{NT},\beta^0,\phi^0)} \left\| \mathcal{L}^{(2,0)}(\beta,\phi) - \overline{\mathcal{L}}^{(2,0)} \right\|_{\max} &= o_P\left(NT\right) .
\end{aligned}
$$

Proof of Lemma A.1. For simplicity write ϕ instead of ϕ_{NT}. Let $\mathcal{L}_{g,h}^{(0,4)}(\beta,\phi) = \partial\mathcal{L}_g^{(0,3)}(\beta,\phi)/\partial\phi_h$. By Taylor expanding $\partial\mathcal{L}_g^{(0,3)}(\beta,\phi)$ we obtain

$$
\begin{aligned}
\sum_g \left\| \mathcal{L}_g^{(0,3)}(\beta^0,\phi) - \overline{\mathcal{L}}_g^{(0,3)} \right\|_\infty &= \sum_g \left\| \mathcal{L}_g^{(0,3)} + \sum_h \mathcal{L}_{g,h}^{(0,4)}(\beta^0,\bar{\phi})(\phi_h-\phi_h^0) - \overline{\mathcal{L}}_g^{(0,3)} \right\|_\infty \\
&\le \sum_g \left\| \widetilde{\mathcal{L}}_g^{(0,3)} \right\|_\infty + \sum_{g,h} \left\| \mathcal{L}_{g,h}^{(0,4)}(\beta^0,\bar{\phi})(\phi_h-\phi_h^0) \right\|_\infty \\
&\le \sum_g \left\| \widetilde{\mathcal{L}}_g^{(0,3)} \right\|_\infty + \sum_{g,h} \left\| \mathcal{L}_{g,h}^{(0,4)}(\beta^0,\bar{\phi}) \right\|_\infty \|\phi-\phi^0\|_\infty \\
&= \mathcal{O}_P\left((NT)^{1/4+\alpha}\right) + \mathcal{O}_P\left((NT)^{1/2}\|\phi-\phi^0\|_\infty\right) , \qquad \text{(A.1)}
\end{aligned}
$$

where $\bar{\phi}$ lies between ϕ and ϕ^0, and in the last step we applied the assumptions. The proof of the other statements in the lemma is analogous. ■

Proof of Theorem 3.1, Part 1: Expansion of $\hat{\phi}(\beta)$. Let $\beta \in \mathcal{B}(\eta_{NT}, \beta^0)$. We suppress the NT-dependence of β, i.e. we just write β instead of β_{NT}. A Taylor expansion of the first order condition (FOC) for $\hat{\phi}(\beta)$ around ϕ^0 and β^0 gives

$$\begin{aligned}
0 &= \Psi(\beta, \hat{\phi}(\beta)) = \Psi(\beta^0, \hat{\phi}(\beta)) + \mathcal{L}^{(1,1)\prime}(\bar{\beta}, \hat{\phi}(\beta))(\beta - \beta^0) \\
&= \Psi(\beta^0, \phi^0) + \mathcal{H}(\beta^0, \phi^0)(\hat{\phi}(\beta) - \phi^0) + \sum_{g=1}^{G_{NT}} \mathcal{L}_g^{(0,3)}(\beta^0, \bar{\phi})(\hat{\phi}_g(\beta) - \phi_g^0)(\hat{\phi}(\beta) - \phi^0)/2 \\
&\quad + \mathcal{L}^{(1,1)\prime}(\bar{\beta}, \hat{\phi}(\beta))(\beta - \beta^0) \\
&= \Psi + \overline{\mathcal{H}}\, (\hat{\phi}(\beta) - \phi^0) + \widetilde{\mathcal{H}}\, (\hat{\phi}(\beta) - \phi^0) + \sum_{g=1}^{G_{NT}} \overline{\mathcal{L}}_g^{(0,3)} (\hat{\phi}_g(\beta) - \phi_g^0)(\hat{\phi}(\beta) - \phi^0)/2 \\
&\quad + \overline{\mathcal{L}}^{(1,1)\prime}(\beta - \beta^0) + r_{NT,1}(\beta)\,,
\end{aligned} \tag{A.2}$$

where $\bar{\beta}$ is between β and β^0, and $\bar{\phi}$ is between $\hat{\phi}(\beta)$ and ϕ^0, and the remainder is

$$\begin{aligned}
r_{NT,1}(\beta) = \sum_{g=1}^{G_{NT}} &\left(\mathcal{L}_g^{(0,3)}(\beta^0, \bar{\phi}) - \overline{\mathcal{L}}_g^{(0,3)}\right)(\hat{\phi}_g(\beta) - \phi_g^0)(\hat{\phi}(\beta) - \phi^0)/2 \\
&+ \left(\mathcal{L}^{(1,1)}(\bar{\beta}, \hat{\phi}(\beta)) - \overline{\mathcal{L}}^{(1,1)}\right)'(\beta - \beta^0).
\end{aligned} \tag{A.3}$$

Using Lemma A.1 we find

$$\begin{aligned}
\|r_{NT,1}(\beta)\|_\infty = \mathcal{O}_P &\left((NT)^{1/4+\alpha}\|\hat{\phi}(\beta) - \phi^0\|_\infty^2\right) + \mathcal{O}_P\left((NT)^{1/2}\|\hat{\phi}(\beta) - \phi^0\|_\infty^3\right) \\
&+ o_P\left(\sqrt{NT}\|\beta - \beta^0\|_\infty\right)\,.
\end{aligned} \tag{A.4}$$

Under our assumptions we can conclude from this expansion of the FOC that

$$\begin{aligned}
\hat{\phi}(\beta) - \phi^0 &= -\overline{\mathcal{H}}^{-1}\Psi + r_{NT,2}(\beta), \\
\|r_{NT,2}(\beta)\|_\infty &= o_P\left((NT)^{-1/4+\alpha}\|\hat{\phi}(\beta) - \phi^0\|_\infty\right) + \mathcal{O}_P\left(\|\hat{\phi}(\beta) - \phi^0\|_\infty^2\right) + \mathcal{O}_P\left(\|\beta - \beta^0\|_\infty\right).
\end{aligned} \tag{A.5}$$

This also implies $\|\hat{\phi}(\beta) - \phi^0\|_\infty = \mathcal{O}_P\left((NT)^{-1/4+\alpha}\right) + \mathcal{O}_P\left(\|\beta - \beta^0\|_\infty\right)$, and therefore $\|r_{NT,2}(\beta)\|_\infty = \mathcal{O}_P\left((NT)^{-1/2+2\alpha}\right) + \mathcal{O}_P\left(\|\beta - \beta^0\|_\infty\right)$. Plugging $\hat{\phi}(\beta) - \phi^0 = \overline{\mathcal{H}}^{-1}\Psi + r_{NT,2}(\beta)$ back into the expansion, and solving for $\hat{\phi}(\beta)$, we obtain

$$\hat{\phi}(\beta) - \phi^0 = -\overline{\mathcal{H}}^{-1}\Psi + \overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\Psi - \overline{\mathcal{H}}^{-1}\sum_{g=1}^{G_{NT}} \overline{\mathcal{L}}_g^{(0,3)}\overline{\mathcal{H}}^{-1}\Psi[\overline{\mathcal{H}}^{-1}\Psi]_g/2 - \overline{\mathcal{H}}^{-1}\overline{\mathcal{L}}^{(1,1)\prime}(\beta - \beta^0) + r_{NT}(\beta), \tag{A.6}$$

where

$$\begin{aligned}
r_{NT}(\beta) = -\overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\, r_{NT,2}(\beta) &- \overline{\mathcal{H}}^{-1}\sum_{g=1}^{G_{NT}} \overline{\mathcal{L}}_g^{(0,3)}(\beta^0, \phi^0) r_{NT,2,g}(\beta) r_{NT,2}(\beta)/2 \\
&+ \overline{\mathcal{H}}^{-1}\sum_{g=1}^{G_{NT}} \overline{\mathcal{L}}_g^{(0,3)}(\beta^0, \phi^0) r_{NT,2,g}(\beta)\overline{\mathcal{H}}^{-1}\Psi + \mathcal{O}_P\left((NT)^{-1/4+\alpha}\|\hat{\phi}(\beta) - \phi^0\|_\infty^2\right) \\
&+ \mathcal{O}_P\left((NT)^{1/2}\|\hat{\phi}(\beta) - \phi^0\|_\infty^3\right) + o_P\left(\|\beta - \beta^0\|_\infty\right)\,,
\end{aligned} \tag{A.7}$$

and therefore we conclude

$$\|r_{NT}(\beta)\|_\infty = \mathcal{O}_P\left((NT)^{-3/4+3\alpha}\right) + o_P(\|\beta-\beta^0\|) = o_P\left((NT)^{-1/2}\right) + o_P(\|\beta-\beta^0\|)\ . \tag{A.8}$$

■

Proof of Theorem 3.1, Part 2: Expansion of profile score. Let $\beta \in \mathcal{B}(\eta_{NT}, \beta^0)$. Let $\hat\phi^0 = \hat\phi(\beta^0)$. From the first part of the theorem we can deduce

$$\begin{aligned}\hat\phi(\beta) - \hat\phi^0 &= -\overline{\mathcal{H}}^{-1}\overline{\mathcal{L}}^{(1,1)\prime}(\beta-\beta^0) + r_{NT,3}(\beta)\ ,\\ \|r_{NT,3}(\beta)\|_\infty &= o_P\left((NT)^{-1/2}\right) + o_P(\|\beta-\beta^0\|)\ .\end{aligned} \tag{A.9}$$

For $\mathcal{L}_{NT}(\beta) = \mathcal{L}_{NT}(\beta, \hat\phi(\beta))$ we obtain

$$\begin{aligned}\frac{\partial \mathcal{L}_{NT}(\beta)}{\partial\beta} &= \mathcal{L}^{(1,0)}(\beta,\hat\phi(\beta)) + \frac{\partial\hat\phi(\beta)}{\partial\beta'}\Psi(\beta,\hat\phi(\beta)) = \mathcal{L}^{(1,0)}(\beta,\hat\phi(\beta))\\ &= \mathcal{L}^{(1,0)}(\beta^0,\hat\phi(\beta)) + \mathcal{L}^{(2,0)}(\bar\beta,\hat\phi(\beta))(\beta-\beta^0)\\ &= \mathcal{L}^{(1,0)}(\beta^0,\hat\phi^0) + \mathcal{L}^{(1,1)}(\beta^0,\bar\phi)(\hat\phi(\beta)-\hat\phi^0) + \mathcal{L}^{(2,0)}(\bar\beta,\hat\phi(\beta))(\beta-\beta^0)\\ &= \mathcal{L}^{(1,0)}(\beta^0,\hat\phi^0) + \left(\overline{\mathcal{L}}^{(2,0)} - \overline{\mathcal{L}}^{(1,1)}\,\overline{\mathcal{H}}^{-1}\overline{\mathcal{L}}^{(1,1)\prime}\right)(\beta-\beta^0) + o_P(\sqrt{NT}) + o_P(NT\|\beta-\beta^0\|_\infty)\ ,\end{aligned} \tag{A.10}$$

where $\bar\beta$ lies between β and β^0, and $\bar\phi$ lies between $\hat\phi(\beta^0)$ and $\hat\phi^0$. Here, in the first line we used $\Psi(\beta,\hat\phi(\beta)) = 0$, which is the FOC for $\hat\phi$. In the second line we expanded $\mathcal{L}^{(1,0)}(\beta,\hat\phi(\beta))$ in its first argument. In the third line we expanded around $\hat\phi^0$. Finally, we applied equation (A.9) and Lemma A.1 to arrive at the last line. Expanding $\mathcal{L}^{(1,0)}(\beta^0,\hat\phi^0)$ around ϕ^0 gives

$$\begin{aligned}\mathcal{L}^{(1,0)}(\beta^0,\hat\phi^0) &= \mathcal{L}^{(1,0)} + \mathcal{L}^{(1,1)}(\hat\phi^0-\phi^0) + \sum_{g=1}^{G_{NT}} \mathcal{L}_g^{(1,2)}(\beta^0,\tilde\phi)(\hat\phi_g^0-\phi_g^0)(\hat\phi^0-\phi^0)/2\\ &= \mathcal{L}^{(1,0)} + \overline{\mathcal{L}}^{(1,1)}(\hat\phi^0-\phi^0) + \widetilde{\mathcal{L}}^{(1,1)}(\hat\phi^0-\phi^0) + \sum_{g=1}^{G_{NT}} \overline{\mathcal{L}}_g^{(1,2)}(\hat\phi_g^0-\phi_g^0)(\hat\phi^0-\phi^0)/2 + o_P(\sqrt{NT}),\end{aligned} \tag{A.11}$$

where $\tilde\phi$ is between $\bar\phi^0$ and ϕ^0, and in the last line we used Lemma A.1 and $\|\hat\phi^0-\phi^0\|_\infty = \mathcal{O}_P\left((NT)^{-1/4+\alpha}\right)$. Using the expansion for $\hat\phi^0$, we obtain

$$\begin{aligned}\mathcal{L}^{(1,0)}(\beta^0,\hat\phi^0) = \mathcal{L}^{(1,0)} &- \overline{\mathcal{L}}^{(1,1)}\,\overline{\mathcal{H}}^{-1}\Psi + \overline{\mathcal{L}}^{(1,1)}\,\overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\Psi - \overline{\mathcal{L}}^{(1,1)}\,\overline{\mathcal{H}}^{-1}\sum_{g=1}^{G_{NT}}\overline{\mathcal{L}}_g^{(0,3)}\overline{\mathcal{H}}^{-1}\Psi[\overline{\mathcal{H}}^{-1}\Psi]_g/2\\ &- \widetilde{\mathcal{L}}^{(1,1)}\overline{\mathcal{H}}^{-1}\Psi + \sum_{g=1}^{G_{NT}}\overline{\mathcal{L}}_g^{(1,2)}[\overline{\mathcal{H}}^{-1}\Psi]_g\overline{\mathcal{H}}^{-1}\Psi/2 + R_{NT,2},\end{aligned} \tag{A.12}$$

where

$$\begin{aligned}R_{NT,2} &= \overline{\mathcal{L}}^{(1,1)}r_{NT}^0 + \widetilde{\mathcal{L}}^{(1,1)}r_{NT,2}^0 - \sum_{g=1}^{G_{NT}}\overline{\mathcal{L}}_g^{(1,2)}[\overline{\mathcal{H}}^{-1}\Psi]_g r_{NT,2}^0 + \sum_{g=1}^{G_{NT}}\overline{\mathcal{L}}_g^{(1,2)}r_{NT,2,g}^0 r_{NT,2}^0/2 + o_P(\sqrt{NT})\\ &= o_P(\sqrt{NT})\ .\end{aligned}$$

Here we used that $r^0_{NT,2} = r_{NT,2}(\beta^0)$ and $r^0_{NT} = r_{NT}(\beta^0)$ satisfy $\|r^0_{NT,2}\|_\infty = \mathcal{O}_P((NT)^{-1/2+2\alpha})$ and $\|r^0_{NT}\|_\infty = o_P((NT)^{-1/2})$, and that $\|\widetilde{\mathcal{L}}^{(1,1)}\|_\infty = \mathcal{O}_P((NT)^{3/4+\alpha})$. We have thus shown that

$$\frac{\partial \mathcal{L}_{NT}(\beta)}{\partial \beta} = \mathcal{L}^{(1,0)} - \overline{\mathcal{L}}^{(1,1)}\,\overline{\mathcal{H}}^{-1}\Psi + \overline{\mathcal{L}}^{(1,1)}\,\overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\Psi - \overline{\mathcal{L}}^{(1,1)}\,\overline{\mathcal{H}}^{-1}\sum_{g=1}^{G_{NT}}\overline{\mathcal{L}}_g^{(0,3)}\overline{\mathcal{H}}^{-1}\Psi[\overline{\mathcal{H}}^{-1}\Psi]_g/2 - \widetilde{\mathcal{L}}^{(1,1)}\overline{\mathcal{H}}^{-1}\Psi$$
$$+ \sum_{g=1}^{G_{NT}}\overline{\mathcal{L}}_g^{(1,2)}[\overline{\mathcal{H}}^{-1}\Psi]_g\overline{\mathcal{H}}^{-1}\Psi/2 + \left(\overline{\mathcal{L}}^{(2,0)} - \overline{\mathcal{L}}^{(1,1)}\,\overline{\mathcal{H}}^{-1}\overline{\mathcal{L}}^{(1,1)\prime}\right)(\beta-\beta^0) + o_P(\sqrt{NT}) + o_P(NT\|\beta-\beta^0\|), \tag{A.13}$$

which is the statement of the theorem. ■

Lemma A.2. *Under assumption 3.1 there exists $\tilde{c} > 0$ such that*

$$\sup_{(\beta,\phi)\in\mathcal{B}(\tilde{c},\beta^0,\phi^0)} \|\mathcal{H}^{-1}(\beta,\phi)\|_\infty = \mathcal{O}_P((NT)^{-1/2}).$$

Proof. Analogous to the proof of Lemma A.1 one can show that for $A(\beta,\phi) = \mathcal{H}(\beta,\phi) - \overline{\mathcal{H}}$ and all $\tilde{c} \le c$ we have

$$\sup_{(\beta,\phi)\in\mathcal{B}(\tilde{c},\beta^0,\phi^0)} \|A(\beta,\phi)\|_\infty = o_P\left(\sqrt{NT}\right) + \tilde{c}\,\mathcal{O}_P\left(\sqrt{NT}\right)\,, \tag{A.14}$$

i.e. there exists a constant $C_1 > 0$ such that for all $\tilde{c} \le c$

$$\sup_{(\beta,\phi)\in\mathcal{B}(\tilde{c},\beta^0,\phi^0)} \|A(\beta,\phi)\|_\infty \le \tilde{c}\,C_1\,\sqrt{NT}\,, \qquad \text{wpa1}. \tag{A.15}$$

Since $\|\overline{\mathcal{H}}^{-1}\|_\infty = \mathcal{O}_P((NT)^{-1/2})$ there also exists $C_2 > 0$ such that $\|\overline{\mathcal{H}}^{-1}\|_\infty \le C_2(NT)^{-1/2}$, wpa1. We choose $\tilde{c} = \min[c, 1/(2C_1C_2)]$. From the definition of the inverse it follows that

$$\begin{aligned}\mathcal{H}^{-1}(\beta,\phi) - \overline{\mathcal{H}}^{-1} &= -\overline{\mathcal{H}}^{-1}A(\beta,\phi)\overline{\mathcal{H}}^{-1}\left(\mathbb{1} + A(\beta,\phi)\overline{\mathcal{H}}^{-1}\right)^{-1} \\ &= -\overline{\mathcal{H}}^{-1}A(\beta,\phi)\overline{\mathcal{H}}^{-1}\sum_{q=0}^{\infty}(-A(\beta,\phi)\overline{\mathcal{H}}^{-1})^q \end{aligned} \tag{A.16}$$

and therefore for $(\beta,\phi) \in \mathcal{B}(\tilde{c},\beta^0,\phi^0)$ we have wpa1

$$\begin{aligned}\left\|\mathcal{H}^{-1}(\beta,\phi) - \overline{\mathcal{H}}^{-1}\right\|_\infty &\le \left\|\overline{\mathcal{H}}^{-1}\right\|_\infty^2 \|A(\beta,\phi)\|_\infty \sum_{q=0}^{\infty}\|A(\beta,\phi)\|_\infty^q \left\|\overline{\mathcal{H}}^{-1}\right\|_\infty^q \\ &= \left\|\overline{\mathcal{H}}^{-1}\right\|_\infty^2 \|A(\beta,\phi)\|_\infty \left(1 - \|A(\beta,\phi)\|_\infty \left\|\overline{\mathcal{H}}^{-1}\right\|_\infty\right)^{-1} \\ &\le \left\|\overline{\mathcal{H}}^{-1}\right\|_\infty^2 \|A(\beta,\phi)\|_\infty (1 - \tilde{c}C_1C_2)^{-1} \\ &\le 2\left\|\overline{\mathcal{H}}^{-1}\right\|_\infty^2 \|A(\beta,\phi)\|_\infty = \mathcal{O}_P\left((NT)^{-1/2}\right)\,. \end{aligned} \tag{A.17}$$

The statement in the lemma then follows from $\left\|\mathcal{H}^{-1}(\beta,\phi)\right\|_\infty \le \left\|\overline{\mathcal{H}}^{-1}\right\|_\infty + \left\|\mathcal{H}^{-1}(\beta,\phi) - \overline{\mathcal{H}}^{-1}\right\|_\infty$. ■

Proof of Theorem 3.3. By assumption, the objective function is strictly concave. Therefore, for given β, there is a one-to-one correspondence between $\phi \in \mathbb{R}^{G_{NT}}$ and the incidental parameter score $\Psi(\beta,\phi) \in \mathbb{R}^{G_{NT}}$. Let $\mathcal{D}_\beta = \Psi(\beta, \mathbb{R}^{G_{NT}})$ be the domain of the incidental parameter score. Let $\Phi(\beta,\psi)$ be the inverse function to $\Psi(\beta,\phi)$, i.e. $\Phi(\beta,\Psi(\beta,\phi)) = \phi$ for all β and ϕ.

We first want to show that for all $\beta \in \mathbb{R}^K$ and $\psi_0, \psi_1 \in \mathbb{R}^{G_{NT}}$ which satisfy $((1-r)\psi_0 + r\psi_1) \in \mathcal{D}_\beta$, for $r \in [0,1]$, we have

$$\|\phi_1 - \phi_0\|_\infty \le \left(\sup_{r\in[0,1]} \left\| \mathcal{H}^{-1}\left(\beta, \phi_r\right) \right\|_\infty \right) \|\psi_1 - \psi_0\|_\infty \,, \tag{A.18}$$

where $\phi_r = \Phi(\beta, (1-r)\psi_0 + r\psi_1)$. Let $\psi(r) = (1-r)\psi_0 + r\psi_1$. We then have

$$\frac{d\phi(r)}{dr} = \mathcal{H}^{-1}(\beta,\phi(r))\, \frac{d\psi(r)}{dr} = \mathcal{H}^{-1}(\beta,\phi(r))\, (\psi_1 - \psi_0) \,, \tag{A.19}$$

and therefore

$$\begin{aligned} \|\phi(1) - \phi(0)\|_\infty &= \left\| \int_0^1 \frac{d\phi(r)}{dr} dr \right\|_\infty \le \int_0^1 \left\| \frac{d\phi(r)}{dr} \right\|_\infty dr \\ &\le \int_0^1 \left\| \mathcal{H}^{-1}(\beta,\phi(r)) \right\|_\infty dr \; \|\psi_1 - \psi_0\|_\infty \\ &\le \left(\sup_{r\in[0,1]} \left\| \mathcal{H}^{-1}(\beta,\phi(r)) \right\|_\infty \right) \|\psi_1 - \psi_0\|_\infty \,, \end{aligned} \tag{A.20}$$

which is the inequality that we wanted to show. Analogously, one can show that for all $\beta_0, \beta_1 \in \mathbb{R}^K$ we have

$$\|\psi_1 - \psi_0\|_\infty \le \left(\sup_{r\in[0,1]} \left\| \frac{\partial \mathcal{L}\left(\beta_r, \phi^0\right)}{\partial\phi\partial\beta'} \right\|_\infty \right) \|\beta_1 - \beta_0\|_\infty \,, \tag{A.21}$$

where $\beta_r = (1-r)\beta_0 + r\beta_1$, and $\psi_0 = \Psi(\beta_0, \phi^0)$, $\psi_1 = \Psi(\beta_1, \phi^0)$.

We now want to prove statement (i) of the theorem. Let $\eta_{NT} > 0$ be such that $\eta_{NT} \to 0$, and let $\beta = \beta_{NT}$ be such that $\beta \in \mathcal{B}(\eta_{NT}, \beta^0)$. Using inequality (A.21) we find

$$\begin{aligned} \left\|\Psi(\beta,\phi^0)\right\|_\infty &\le \left\|\Psi(\beta^0,\phi^0)\right\|_\infty + \left\|\Psi(\beta^0,\phi^0) - \Psi(\beta,\phi^0)\right\|_\infty \\ &\le \mathcal{O}_P((NT)^{1/4+\alpha}) + \left(\sup_{\beta\in\mathcal{B}(\eta_{NT},\beta^0)} \left\| \frac{\partial\mathcal{L}\left(\tilde\beta, \phi^0\right)}{\partial\phi\partial\beta'} \right\|_\infty \right) \|\beta - \beta^0\|_\infty \\ &= \mathcal{O}_P\left((NT)^{1/4+\alpha}\right) + \mathcal{O}_P(\sqrt{NT})\eta_{NT} \\ &= o_P\left(\sqrt{NT}\right) \,. \end{aligned} \tag{A.22}$$

Note that $\hat\phi(\beta) = \Phi(\beta, 0)$ and $\Psi(\beta, \hat\phi(\beta)) = 0$. Using the result in equation (A.22), inequality (A.18) and lemma A.2 we want to show that $\|\hat\phi(\beta) - \phi^0\| \le \left[\sup_\phi \left\|\mathcal{H}^{-1}(\beta,\phi)\right\|_\infty\right] \left\|0 - \Psi(\beta,\phi^0)\right\|_\infty = \mathcal{O}_P((NT)^{-1/2}) o_P\left(\sqrt{NT}\right) = o_P(1)$. Here, one has to be careful regarding the set over which the supremum is taken, to avoid a circular conclusion. We therefore present the following indirect argument.

Let $\tilde{\eta}_{NT} > 0$ be series such that $\tilde{\eta}_{NT} \to 0$ and

$$\tilde{\eta}_{NT} > \left(\sup_{\phi \in \mathcal{B}(c,\phi^0)} \left\| \mathcal{H}^{-1}(\beta,\phi) \right\|_\infty \right) \left\| \Psi(\beta,\phi^0) \right\|_\infty , \tag{A.23}$$

Such a $\tilde{\eta}_{NT} > 0$ has to exist, because the rhs of (A.23) is $o_P(1)$. Now, for $r \in [0,1]$ define $\psi_r = (1-r)\Psi(\beta,\phi^0)$, and let

$$r_{\max,NT} = \sup \left\{ r \in [0,1] \,\middle|\, \forall \tilde{r} \leq r \; \exists \phi \in \mathbb{R}^{G_{NT}} : \; \psi_{\tilde{r}} = \Psi(\beta,\phi) \text{ and } \|\phi - \phi^0\|_\infty \leq \tilde{\eta}_{NT} \right\} . \tag{A.24}$$

The subset of $[0,1]$ over which we take the supremum is non-empty since $r = 0$ satisfies the condition with $\phi = \phi^0$. Since the function ψ_r is continuous in r, and $\psi(\beta,\phi)$ is continuous in ϕ, it must be the case that either $r_{\max,NT} = 1$, or $\|\Phi(\beta,\psi_{r_{\max,NT}}) - \phi^0\|_\infty = \tilde{\eta}_{NT}$. We show that the second possibility can be ruled out asymptotically, namely by inequality (A.18), and using (A.22) and lemma A.2, we would have

$$\begin{aligned}
\tilde{\eta}_{NT} &= \|\Phi(\beta,\psi_{r_{\max,NT}}) - \phi^0\|_\infty \\
&\leq \left(\sup_{\phi \in \mathcal{B}(\tilde{\eta}_{NT},\phi^0)} \left\| \mathcal{H}^{-1}(\beta,\phi) \right\|_\infty \right) \left\| \psi_{r_{\max,NT}} - \Psi(\beta,\phi^0) \right\|_\infty \\
&= \left(\sup_{\phi \in \mathcal{B}(\tilde{\eta}_{NT},\phi^0)} \left\| \mathcal{H}^{-1}(\beta,\phi) \right\|_\infty \right) r_{\max,NT} \left\| \Psi(\beta,\phi^0) \right\|_\infty \\
&\leq \left(\sup_{\phi \in \mathcal{B}(\tilde{\eta}_{NT},\phi^0)} \left\| \mathcal{H}^{-1}(\beta,\phi) \right\|_\infty \right) \left\| \Psi(\beta,\phi^0) \right\|_\infty ,
\end{aligned} \tag{A.25}$$

which is in contradiction with (A.23). We can thus conclude that $r_{\max,NT} = 1$ wpa1. Thus $\psi_{r_{\max,NT}} = 0$ wpa1, and thus $\Phi(\beta,0) = \hat{\phi}(\beta)$ exists wpa1, and satisfies $\|\hat{\phi}(\beta) - \phi^0\|_\infty \leq \tilde{\eta}_{NT} = o_P(1)$, wpa1, which proves part (i) of the theorem.

The proof of part (ii) of the theorem is straightforward. Since the objective function $\mathcal{L}(\beta,\phi)$ is strictly convex the same must be true for the profile objective function $\mathcal{L}(\beta)$. Therefore, the FOC $\partial\mathcal{L}(\beta)/\partial\beta = 0$ uniquely determines the optimum $\hat{\beta}$. Applying the expansion of $\partial\mathcal{L}(\beta)/\partial\beta$ in theorem 3.1 then gives the required results. ■

References

Arellano, M. and Hahn, J. (2005). Understanding bias in nonlinear panel models: Some recent developments. *Working Papers*, pages 1–38.

Arellano, M. and Hahn, J. (2007). Understanding bias in nonlinear panel models: Some recent developments. *Econometric Society Monographs*, 43:381.

Bai, J. (2009). Panel data models with interactive fixed effects. *Econometrica*.

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

Dhaene, G., Leuven, K., and Jochmans, K. (2010). Split-panel jackknife estimation of fixed-effect models.

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

Greene, W. (2004). The behavior of the fixed effects estimator in nonlinear models. *The Econometrics Journal*, 7(1):98–119.

Hahn, J. and Kuersteiner, G. (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica*, 70(4):1639–1657.

Hahn, J. and Kuersteiner, G. (2004). Bias reduction for dynamic nonlinear panel models with fixed effects. *Unpublished manuscript.*

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72(4):1295–1319.

Heckman, J. (1981). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. *Structural analysis of discrete data with econometric applications*, pages 179–195.

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

Lancaster, T. (2002). Orthogonal parameters and panel data. *The Review of Economic Studies*, 69(3):647–666.

Moon, H. and Weidner, M. (2010). Linear Regression for Panel with Unknown Number of Factors as Interactive Fixed Effects. Manuscript.

Neyman, J. and Scott, E. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1–32.

Olsen, R. (1978). Note on the uniqueness of the maximum likelihood estimator for the tobit model. *Econometrica: Journal of the Econometric Society*, pages 1211–1215.

Phillips, P. C. B. and Moon, H. (1999). Linear regression limit theory for nonstationary panel data. *Econometrica*, 67(5):1057–1111.

Pratt, J. (1981). Concavity of the log likelihood. *Journal of the American Statistical Association*, pages 103–106.

Woutersen, T. (2002). Robustness against incidental parameters. *Unpublished manuscript.*