cemmap

centre for microdata methods and practice

# Individual and time effects in nonlinear panel models with large *N*, *T*

**Iván Fernández-Val**
**Martin Weidner**

The Institute for Fiscal Studies
Department of Economics, UCL

**cemmap** working paper CWP60/13

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

# Individual and Time Effects in Nonlinear Panel Models with Large $N$, $T$*

Iván Fernández-Val[‡] Martin Weidner[§]

November 27, 2013

**Abstract**

Fixed effects estimators of nonlinear panel data models can be severely biased because of the well-known incidental parameter problem. We develop analytical and jackknife bias corrections for nonlinear models with both individual and time effects. Under asymptotic sequences where the time-dimension ($T$) grows with the cross-sectional dimension ($N$), the time effects introduce additional incidental parameter bias. As the existing bias corrections apply to models with only individual effects, we derive the appropriate corrections for the case when both effects are present. The basis for the corrections are general asymptotic expansions of fixed effects estimators with incidental parameters in multiple dimensions. We apply the expansions to M-estimators with concave objective functions in parameters for panel models with additive individual and time effects. These estimators cover fixed effects estimators of the most popular limited dependent variable models such as logit, probit, ordered probit, Tobit and Poisson models. Our analysis therefore extends the use of large-$T$ bias adjustments to an important class of models. We also develop bias corrections for functions of the data, parameters and individual and time effects including average partial effects. In this case, the incidental parameter bias can be asymptotically of second order, but the corrections still improve finite-sample properties.

**Keywords:** Panel data, asymptotic bias correction, fixed effects.

**JEL:** C13, C23.

*We thank Richard Blundell, Ben Boury, Mingli Chen, Geert Dhaene, Jinyong Hahn, Bo Honoré, Koen Jochmans, Hiroaki Kaido, Dennis Kristensen, Guido Kuersteiner, Roger Moon, Zhongjun Qu, Sami Stouli, and Francis Vella for helpful discussions. We also benefitted from comments by seminar participants at Cambridge, Georgetown, Iowa, Maryland, Michigan State, Princeton, UCL, and USC, and by conference participants at the 16th and 19th International Conference on Panel Data, as well as the 2013 North American Summer Meetings of the Econometric Society. Financial support from the National Science Foundation and the Economic and Social Research Council through the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0002 is gratefully acknowledged.

‡Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215-1403, USA. Email: ivanf@bu.edu

§Department of Economics, University College London, Gower Street, London WC1E 6BT, UK, and and CeMMaP. Email: m.weidner@ucl.ac.uk

# 1 Introduction

Fixed effects estimators of nonlinear panel data models can be severely biased because of the well-known incidental parameter problem (Neyman and Scott (1948), Heckman (1981), Lancaster (2000), and Greene (2004)). A recent literature, surveyed in Arellano and Hahn (2007) and including Phillips and Moon (1999), Hahn and Kuersteiner (2002), Lancaster (2002), Woutersen (2002), Hahn and Newey (2004), Carro (2007), Arellano and Bonhomme (2009), Fernandez-Val (2009), Hahn and Kuersteiner (2011), Fernandez-Val and Vella (2011), and Kato, Galvao and Montes-Rojas (2012), provides a range of solutions, so-called large-$T$ bias corrections, to reduce the incidental parameter 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 modelling because they allow the researcher to control for aggregate common shocks and to parsimoniously introduce dependence across individuals.

We develop analytical and jackknife bias corrections for nonlinear models with *both* individual and time effects. To justify the corrections, we rely on asymptotic sequences where $T$ grows with the cross-sectional dimension $N$, as an approximation to the properties of the estimators in econometric applications where $T$ is moderately large relative to $N$. Examples include empirical applications that use U.S. state or country level panel data, or trade flows across countries. Under these asymptotics, the incidental parameter problem becomes a finite-sample bias problem in the time dimension and the presence of time effects introduces additional bias in the cross sectional dimension. As the existing bias corrections apply to models with only individual effects, we derive the appropriate correction. This correction corresponds to a properly adjusted sequential application of the existing corrections to each dimension.

In addition to model parameters, we provide bias corrections for average partial effects, which are often the ultimate quantities of interest in nonlinear models. These effects are functions of the data, parameters and individual and time effects in nonlinear models. The asymptotic distribution of the fixed effects estimators of these quantities depends on the sampling properties of the individual and time effects, unlike for model parameters. We find that in general the incidental parameters problem for average effects is of second order asymptotically, because the rate of convergence for these effects is generally slower than for model parameters.[1] The bias corrections, while not necessary to center the asymptotic distribution, improve the finite-sample properties of the estimators specially in dynamic models.

The basis for the bias corrections are asymptotic expansions of fixed effects estimators with incidental parameters in multiple dimensions. Bai (2009) and Moon and Weidner (2010; 2013) derive similar expansions for least squares estimators of linear models with interactive individual and time effects. We consider non-linear models with additive individual and time effects. In our case, the nonlinearity of the model introduces nonseparability between the estimators of the model parameters and incidental parameters (individual and time effects). Moreover, we need to deal with an asymptotically infinite

[1] Galvao and Kato (2013) also found slow rates of convergence for fixed effects estimators in linear models with individual effects under misspecification.

dimensional non-diagonal Hessian matrix for the incidental parameters. We focus on M-estimators with concave objective functions in all the parameters, because concavity facilitates showing consistency in our setting where the dimension of the parameter space grows with the sample size. This case covers fixed effects estimators of the most popular limited dependent variable models such as logit, probit, ordered 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, 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. We consider analytical methods similar to Hahn and Newey (2004) and Hahn and Kuersteiner (2011), and suitable modifications of the split panel jackknife of Dhaene and Jochmans (2010).[2] However, the theory of the previous papers does not cover the models that we consider, because, in addition to not allowing for time effects, they assume either identical distribution or stationarity over time for the processes of the observed variables, conditional on the unobserved effects. These assumptions are violated in our models due to the presence of the time effects. We therefore need to extend the validity of the bias corrections to heterogenous processes in multiple dimensions under weak time series dependence conditions.

Simulation evidence indicates that our corrections improve the estimation and inference performance of the fixed effects estimators of parameters and average effects. The analytical corrections dominate the jackknife corrections in probit and Poisson models for sample sizes that are relevant for empirical practice. We illustrate the corrections with an empirical application on the relationship between competition and innovation using a panel of U.K. industries, following Aghion, Blundell, Griffith, and Howitt (2005). We find that the inverted-U pattern relationship found by Aghion et al is robust to relaxing the strict exogeneity assumption of competition with respect to the innovation process and to the inclusion of innovation dynamics. We also uncover substantial state dependence in the innovation process.

The large-$T$ panel literature on models with individual and time effects is sparse. Regarding linear regression models, there is a literature on interactive fixed effects that includes some of the papers mentioned above (e.g. Pesaran (2006), Bai (2009), Moon and Weidner (2010; 2013)). Furthermore, Hahn and Moon (2006) considered bias corrected fixed effects estimators in panel linear autoregressive models with additive individual and time effects. Regarding non-linear models, there is independent and contemporaneous work by Charbonneau (2012), which extended the conditional fixed effects estimators to logit and Poisson models with exogenous regressors and additive individual and time effects. She differences out the individual and time effects by conditioning on sufficient statistics. The conditional approach completely eliminates the asymptotic bias coming from the estimation of the incidental parameters, but it does not permit estimation of average partial effects and has not been developed for models with predetermined regressors. We instead consider estimators of model parameters and average partial effects in nonlinear models with predetermined regressors. The two approaches can therefore be considered as complementary.

In Section 2, we introduce the model and fixed effects estimators. Section 3 describes the bias

[2]A similar split panel jackknife bias correction method was outlined in Hu (2002).

corrections to deal with the incidental parameters problem and illustrates how the bias corrections work through an example. Section 4 provides the asymptotic theory. Sections 5 and 6 give Monte Carlo and empirical results. We collect the proofs of all the results and additional technical details in the Appendix.

# 2 Model and Estimators

## 2.1 Model

The data consist of $N \times T$ observations $\{(Y_{it}, X_{it}')' : 1 \leq i \leq N, 1 \leq t \leq T\}$, for a scalar outcome variable of interest $Y_{it}$ and a vector of explanatory variables $X_{it}$. We assume that the outcome for individual $i$ at time $t$ is generated by the sequential process:

$$Y_{it} \mid X_i^t, \alpha, \gamma, \beta \sim f_Y(\cdot \mid X_{it}, \alpha_i, \gamma_t, \beta), \quad (i = 1, ..., N; t = 1, ..., T),$$

where $X_i^t = (X_{i1}, \ldots, X_{it})$, $\alpha = (\alpha_1, \ldots, \alpha_N)$, $\gamma = (\gamma_1, \ldots, \gamma_T)$, $f_Y$ is a known probability function, and $\beta$ is a finite dimensional parameter vector.

The variables $\alpha_i$ and $\gamma_t$ are unobserved individual and time effects that in economic applications capture individual heterogeneity and aggregate shocks, respectively. The model is semiparametric because we do not specify the distribution of these effects nor their relationship with the explanatory variables. The conditional distribution $f_Y$ represents the parametric part of the model. The vector $X_{it}$ contains predetermined variables with respect to $Y_{it}$. Note that $X_{it}$ can include lags of $Y_{it}$ to accommodate dynamic models.

We consider two running examples throughout the analysis:

**Example 1** (Binary response model)**.** *Let $Y_{it}$ be a binary outcome and $F$ be a cumulative distribution function, e.g. the standard normal or logistic distribution. We can model the conditional distribution of $Y_{it}$ using the single-index specification*

$$f_Y(y \mid X_{it}, \alpha_i, \gamma_t, \beta) = F(X_{it}'\beta + \alpha_i + \gamma_t)^y [1 - F(X_{it}'\beta + \alpha_i + \gamma_t)]^{1-y}, \quad y \in \{0, 1\}.$$

**Example 2** (Count response model)**.** *Let $Y_{it}$ be a non-negative interger-valued outcome, and $f(\cdot; \lambda)$ be the probability mass function of a Poisson random variable with mean $\lambda > 0$. We can model the conditional distribution of $Y_{it}$ using the single index specification*

$$f_Y(y \mid X_{it}, \alpha_i, \gamma_t, \beta) = f(y; \exp[X_{it}'\beta + \alpha_i + \gamma_t]), \quad y \in \{0, 1, 2, ....\}.$$

For estimation, we adopt a fixed effects approach treating the realization of the unobserved individual and time effects as parameters to be estimated. We collect all these effects in the vector $\phi_{NT} = (\alpha_1, ..., \alpha_N, \gamma_1, ..., \gamma_T)'$. The model parameter $\beta$ usually includes regression coefficients of interest, while the unobserved effects $\phi_{NT}$ are treated as a nuisance parameter. The true values of the parameters, denoted by $\beta^0$ and $\phi_{NT}^0 = (\alpha_1^0, ..., \alpha_N^0, \gamma_1^0, ..., \gamma_T^0)'$, are the solution to the population problem

$$\begin{aligned} &\max_{(\beta, \phi_{NT}) \in \mathbb{R}^{\dim \beta + \dim \phi_{NT}}} \ \mathbb{E}_\phi[\mathcal{L}_{NT}(\beta, \phi_{NT})], \\ &\mathcal{L}_{NT}(\beta, \phi_{NT}) := (NT)^{-1/2} \left\{ \textstyle\sum_{i,t} \log f_Y(Y_{it} \mid X_{it}, \alpha_i, \gamma_t, \beta) - b(v_{NT}'\phi_{NT})^2/2 \right\}, \end{aligned} \tag{2.1}$$

for every $N, T$, where $\mathbb{E}_\phi$ denotes the expectation with respect to the distribution of the data conditional on the unobserved effects and initial conditions, $b > 0$ is an arbitrary constant, $v_{NT} = (1_N', -1_T')'$, and $1_N$ and $1_T$ denote vectors of ones with dimensions $N$ and $T$. We will assume that the solution to the population problem exists and is unique. This will be justified, for example, by a concavity assumption on the objective function that we impose in Section 4. The second term of $\mathcal{L}_{NT}$ is a penalty that imposes a normalization needed to identify $\phi_{NT}$ in models with scalar individual and time effects that enter additively into the log-likelihood function as $\alpha_i + \gamma_t$.[3] In this case, adding a constant to all $\alpha_i$, while subtracting it from all $\gamma_t$, does not change $\alpha_i + \gamma_t$. To eliminate this ambiguity, we normalize $\phi^0_{NT}$ to satisfy $v_{NT}'\phi^0_{NT} = 0$, i.e. $\sum_i \alpha_i^0 = \sum_t \gamma_t^0$. The penalty produces a maximizer of $\mathcal{L}_{NT}$ that is automatically normalized. We could equivalently impose the constraint $v_{NT}'\phi_{NT} = 0$ in the program, but for technical reasons we prefer to work with an unconstrained optimization problem.[4] The pre-factor $(NT)^{-1/2}$ in $\mathcal{L}_{NT}(\beta, \phi_{NT})$ is just a convenient rescaling when discussing the structure of the Hessian of the incidental parameters below.

Other quantities of interest involve averages over the data and unobserved effects

$$\delta^0_{NT} = \mathbb{E}[\Delta_{NT}(\beta^0, \phi^0_{NT})], \quad \Delta_{NT}(\beta, \phi_{NT}) = (NT)^{-1} \sum_{i,t} \Delta(X_{it}, \beta, \alpha_i, \gamma_t), \tag{2.2}$$

where $\mathbb{E}$ denotes the expectation with respect to the joint distribution of the data and the unobserved effects, provided that the expectation exists. They are indexed by $N$ and $T$ because the marginal distribution of $\{(X_{it}, \alpha_i, \gamma_t) : 1 \leq i \leq N, 1 \leq t \leq T\}$ can be heterogeneous across $i$ and/or $t$; see Section 4.2. These averages include average partial effects (APEs), which are often the ultimate quantities of interest in nonlinear models. Some examples of partial effects, motivated by the numerical examples of Sections 5 and 6, are the following:

**Example 1** (Binary response model). *If $X_{it,k}$, the $k$th element of $X_{it}$, is binary, its partial effect on the conditional probability of $Y_{it}$ is*

$$\Delta(X_{it}, \beta, \alpha_i, \gamma_t) = F(\beta_k + X_{it,-k}'\beta_{-k} + \alpha_i + \gamma_t) - F(X_{it,-k}'\beta_{-k} + \alpha_i + \gamma_t), \tag{2.3}$$

*where $\beta_k$ is the $k$th element of $\beta$, and $X_{it,-k}$ and $\beta_{-k}$ include all elements of $X_{it}$ and $\beta$ except for the $k$th element. If $X_{it,k}$ is continuous and $F$ is differentiable, the partial effect of $X_{it,k}$ on the conditional probability of $Y_{it}$ is*

$$\Delta(X_{it}, \alpha_i, \gamma_t) = \beta_k \partial F(X_{it}'\beta + \alpha_i + \gamma_t), \tag{2.4}$$

*where $\partial F$ is the derivative of $F$.*

[3] In Appendix B we derive asymptotic expansions that apply to more general models. In order to use these expansions to obtain the asymptotic distribution of the panel fixed effects estimators, we need to derive the properties of the expected Hessian of the incidental parameters, a matrix with increasing dimension, and to show the consistency of the estimator of the incidental parameter vector. The additive specification $\alpha_i + \gamma_t$ is useful to characterize the Hessian and we impose strict concavity of the objective function to show the consistency.

[4] There are alternative normalizations for $\phi_{NT}$ such as $\alpha_1 = 0$. The normalization has no effect on the model parameter and average partial effects. Our choice is very convenient for certain intermediate results that involve the incidental parameters $\phi_{NT}$, their score vector and their Hessian matrix.

**Example 2** (Count response model). *If $X_{it}$ includes $Z_{it}$ and some known transformation $H(Z_{it})$ with coefficients $\beta_k$ and $\beta_j$, the partial effect of $Z_{it}$ on the conditional expectation of $Y_{it}$ is*

$$\Delta(X_{it}, \beta, \alpha_i, \gamma_t) = [\beta_k + \beta_j \partial H(Z_{it})] \exp(X_{it}'\beta + \alpha_i + \gamma_t). \tag{2.5}$$

## 2.2 Fixed effects estimators

We estimate the parameters by solving the sample analog of problem (2.1), i.e.

$$\max_{(\beta,\phi_{NT})\in\mathbb{R}^{\dim\beta+\dim\phi_{NT}}} \mathcal{L}_{NT}(\beta, \phi_{NT}).$$

To analyze the properties of the estimator of $\beta$ it is convenient to first concentrate out the nuisance parameter $\phi_{NT}$. For given $\beta$, we define the optimal $\widehat{\phi}_{NT}(\beta)$ as

$$\widehat{\phi}_{NT}(\beta) = \operatorname*{argmax}_{\phi_{NT}\in\mathbb{R}^{\dim\phi_{NT}}} \mathcal{L}_{NT}(\beta,\, \phi_{NT})\ . \tag{2.6}$$

The fixed effects estimators of $\beta^0$ and $\phi_{NT}^0$ are

$$\widehat{\beta}_{NT} \;=\; \operatorname*{argmax}_{\beta\in\mathbb{R}^{\dim\beta}} \mathcal{L}_{NT}(\beta, \widehat{\phi}_{NT}(\beta))\ , \qquad \widehat{\phi}_{NT} \;=\; \widehat{\phi}_{NT}(\widehat{\beta}). \tag{2.7}$$

As in the population case, we will impose conditions guaranteeing that the solutions to the previous programs exist and are unique a.s.

Estimators of APEs can be formed by plugging-in the estimators of the model parameters in the sample version of (2.2), i.e.

$$\widehat{\delta}_{NT} = \Delta_{NT}(\widehat{\beta}, \widehat{\phi}_{NT}). \tag{2.8}$$

# 3 Incidental parameter problem and bias corrections

In this section we give a heuristic discussion of the main results, leaving the technical details to Section 4. We illustrate the analysis with numerical calculations based on a variation of the classical Neyman and Scott (1948) variance example.

## 3.1 Incidental parameter problem

Fixed effects estimators in nonlinear or dynamic models suffer from the incidental parameter problem (Neyman and Scott, 1948). The individual and time effects are incidental parameters that cause the estimators of the model parameters to be inconsistent under asymptotic sequences where either $N$ or $T$ are fixed. To describe the problem let

$$\overline{\beta}_{NT} := \operatorname*{argmax}_{\beta\in\mathbb{R}^{\dim\beta}} \mathbb{E}_\phi\left[\mathcal{L}_{NT}(\beta, \widehat{\phi}_{NT}(\beta))\right]. \tag{3.1}$$

In general, $\operatorname{plim}_{N\to\infty} \overline{\beta}_{NT} \neq \beta^0$ and $\operatorname{plim}_{T\to\infty} \overline{\beta}_{NT} \neq \beta^0$ because of the estimation error in $\widehat{\phi}_{NT}(\beta)$ when one of the dimensions is fixed. If $\widehat{\phi}_{NT}(\beta)$ is replaced by $\phi_{NT}(\beta) = \operatorname{argmax}_{\phi_{NT}\in\mathbb{R}^{\dim\phi_{NT}}} \mathbb{E}_\phi[\mathcal{L}_{NT}(\beta,\, \phi_{NT})]$, then $\overline{\beta}_{NT} = \beta^0$. We consider analytical and jackknife corrections for the bias $\overline{\beta}_{NT} - \beta^0$.

### 3.2 Bias Corrections

Some expansions can be used to explain our corrections. Under suitable sampling conditions, the bias is small for large enough $N$ and $T$, i.e., $\text{plim}_{N,T\to\infty} \overline{\beta}_{NT} = \beta^0$. For smooth likelihoods and under appropriate regularity conditions, as $N, T \to \infty$,

$$\overline{\beta}_{NT} = \beta^0 + \overline{B}_\infty/T + \overline{D}_\infty/N + o_P(T^{-1} \vee N^{-1}), \tag{3.2}$$

for some $\overline{B}_\infty$ and $\overline{D}_\infty$ that we characterize in Section 4. Unlike in nonlinear models without incidental parameters, the order of the bias is higher than the inverse of the sample size $(NT)^{-1}$ due to the slow rate of convergence of $\widehat{\phi}_{NT}$. Note also that by the properties of the maximum likelihood estimator

$$\sqrt{NT}(\widehat{\beta}_{NT} - \overline{\beta}_{NT}) \to_d \mathcal{N}(0, \overline{V}_\infty).$$

Under asymptotic sequences where $N/T \to \kappa^2$ as $N, T \to \infty$, the fixed effects estimator is asymptotically biased because

$$\begin{aligned}\sqrt{NT}(\widehat{\beta}_{NT} - \beta^0) &= \sqrt{NT}(\widehat{\beta}_{NT} - \overline{\beta}_{NT}) + \sqrt{NT}(\overline{B}_\infty/T + \overline{D}_\infty/N + o_P(T^{-1} \vee N^{-1})) \\ &\to_d \mathcal{N}(\kappa\overline{B}_\infty + \kappa^{-1}\overline{D}_\infty, \overline{V}_\infty). \end{aligned} \tag{3.3}$$

This is the large-$N$ large-$T$ version of the incidental parameters problem that invalidates any inference based on the asymptotic distribution. Relative to fixed effects estimators with only individual effects, the presence of time effects introduces additional asymptotic bias through $\overline{D}_\infty$.

The analytical bias correction consists of removing estimates of the leading terms of the bias from the fixed effect estimator of $\beta^0$. Let $\widehat{B}_{NT}$ and $\widehat{D}_{NT}$ be estimators of $\overline{B}_\infty$ and $\overline{D}_\infty$, respectively. The bias corrected estimator can be formed as

$$\widetilde{\beta}^A_{NT} = \widehat{\beta}_{NT} - \widehat{B}_{NT}/T - \widehat{D}_{NT}/N.$$

If $N/T \to \kappa^2$, $\widehat{B}_{NT} \to_P \overline{B}_\infty$, and $\widehat{D}_{NT} \to_P \overline{D}_\infty$, then

$$\sqrt{NT}(\widetilde{\beta}^A_{NT} - \beta^0) \to_d \mathcal{N}(0, \overline{V}_\infty).$$

The analytical correction therefore centers the asymptotic distribution at the true value of the parameter, without increasing asymptotic variance.

We consider a jackknife bias correction method that does not require explicit estimation of the bias, but is computationally more intensive. This method is based on the split panel jackknife (SPJ) of Dhaene and Jochmans (2010) applied to the two dimensions of the panel.[5] To describe it, let $\widetilde{\beta}_{N,T/2}$ be the average of the 2 split jackknife estimators that leave out the first and second halves of the time periods, and let $\widetilde{\beta}_{N/2,T}$ be the average of the 2 split jackknife estimators that leave out half of the individuals. In choosing the cross sectional division of the panel, we might want to take into account

[5] Alternative jackknife corrections based on the leave-one-observation-out panel jackknife (PJ) of Hahn and Newey (2004) and combinations of PJ and SPJ are also possible. We do not consider corrections based on PJ because they are theoretically justified by second-order expansions of $\overline{\beta}_{NT}$ that are beyond the scope of this paper.

individual clustering structures to preserve and account for cross sectional dependencies.[6] The bias corrected estimator is

$$\widetilde{\beta}^J_{NT} = 3\widehat{\beta}_{NT} - \widetilde{\beta}_{N,T/2} - \widetilde{\beta}_{N/2,T}. \tag{3.4}$$

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

$$\widetilde{\beta}^J_{NT} - \beta_0 = (\widehat{\beta}_{NT} - \beta_0) - (\widetilde{\beta}_{N,T/2} - \widehat{\beta}_{NT}) - (\widetilde{\beta}_{N/2,T} - \widehat{\beta}_{NT}),$$

where $\widetilde{\beta}_{N,T/2} - \widehat{\beta}_{NT} = \overline{B}_\infty/T + o_P(T^{-1} \vee N^{-1})$ and $\widetilde{\beta}_{N/2,T} - \widehat{\beta}_{NT} = \overline{D}_\infty/N + o_P(T^{-1} \vee N^{-1})$. The time series split removes the bias term $\overline{B}_\infty$ and the cross sectional split removes the bias term $\overline{D}_\infty$.

### 3.3 Illustrative Example

To illustrate how the bias corrections work in finite samples, we consider a simple model where the solution to the population program (3.1) has closed form. This model corresponds to the classical Neyman and Scott (1948) variance example with individual and time effects, $Y_{it} \mid \alpha, \gamma, \beta \sim \mathcal{N}(\alpha_i + \gamma_t, \beta)$. It is well-know that in this case

$$\widehat{\beta}_{NT} = (NT)^{-1} \sum_{i,t} \left(Y_{it} - \bar{Y}_{i.} - \bar{Y}_{.t} + \bar{Y}_{..}\right)^2,$$

where $\bar{Y}_{i.} = T^{-1}\sum_t Y_{it}$, $\bar{Y}_{.t} = N^{-1}\sum_i Y_{it}$, and $\bar{Y}_{..} = (NT)^{-1}\sum_{i,t} Y_{it}$. Moreover,

$$\overline{\beta}_{NT} = \mathbb{E}_\phi[\widehat{\beta}_{NT}] = \beta^0 \frac{(N-1)(T-1)}{NT} = \beta^0\left(1 - \frac{1}{T} - \frac{1}{N} + \frac{1}{NT}\right),$$

so that $\overline{B}_\infty = -\beta^0$ and $\overline{D}_\infty = -\beta^0$.

To form the analytical bias correction we can set $\widehat{B}_{NT} = -\widehat{\beta}_{NT}$ and $\widehat{D}_{NT} = -\widehat{\beta}_{NT}$. This yields $\widetilde{\beta}^A_{NT} = \widehat{\beta}_{NT}(1 + 1/T + 1/N)$ with

$$\overline{\beta}^A_{NT} = \mathbb{E}_\phi[\widetilde{\beta}^A_{NT}] = \beta^0\left(1 - \frac{1}{T^2} - \frac{1}{N^2} - \frac{1}{NT} + \frac{1}{NT^2} + \frac{1}{N^2T}\right).$$

This correction reduces the order of the bias from $(T^{-1} \vee N^{-1})$ to $(T^{-2} \vee N^{-2})$, and introduces additional higher order terms. The analytical correction increases finite-sample variance because the factor $(1 + 1/T + 1/N) > 1$. We compare the biases and standard deviations of the fixed effects estimator and the corrected estimator in a numerical example below.

For the Jackknife correction, straightforward calculations give

$$\overline{\beta}^J_{NT} = \mathbb{E}_\phi[\widetilde{\beta}^J_{NT}] = 3\overline{\beta}_{NT} - \overline{\beta}_{N,T/2} - \overline{\beta}_{N/2,T} = \beta^0\left(1 - \frac{1}{NT}\right).$$

The correction therefore reduces the order of the bias from $(T^{-1} \vee N^{-1})$ to $(TN)^{-1}$.[7]

---

[6] When $T$ is odd we define $\widetilde{\beta}_{N,T/2}$ as the average of the 2 split jackknife estimators that use overlapping subpanels with $t \leq (T+1)/2$ and $t \geq (T+1)/2$. We define $\widetilde{\beta}_{N/2,T}$ similarly when $N$ is odd.

[7] In this example it is possible to develop higher-order jackknife corrections that completely eliminate the bias because we know the entire expansion of $\overline{\beta}_{NT}$. For example, $\mathbb{E}_\phi[4\widehat{\beta}_{NT} - 2\widetilde{\beta}_{N,T/2} - 2\widetilde{\beta}_{N/2,T} + \widetilde{\beta}_{N/2,T/2}] = \beta^0$, where $\widetilde{\beta}_{N/2,T/2}$ is the average of the four split jackknife estimators that leave out half of the individuals and the first or the second halves of the time periods. See Dhaene and Jochmans (2010) for a discussion on higher-order bias corrections of panel fixed effects estimators.

Table 1 presents numerical results for the bias and standard deviations of the fixed effects and bias corrected estimators in finite samples. We consider panels with $N, T \in \{10, 25, 50\}$, and only report the results for $T \leq N$ since all the expressions are symmetric in $N$ and $T$. All the numbers in the table are in percentage of the true parameter value, so we do not need to specify the value of $\beta^0$. We find that the analytical and jackknife corrections offer substantial improvements over the fixed effects estimator in terms of bias. The first and fourth row of the table show that the bias of the fixed effects estimator is of the same order of magnitude as the standard deviation, where $\overline{V}_{NT} = \text{Var}[\widehat{\beta}_{NT}] = 2(N-1)(T-1)(\beta^0)^2/(NT)^2$ under independence of $Y_{it}$ over $i$ and $t$ conditional on the unobserved effects. The last row shows the increase in standard deviation due to analytical bias correction is small compared to the bias reduction, where $\overline{V}^A_{NT} = \text{Var}[\widetilde{\beta}^A_{NT}] = (1 + 1/N + 1/T)^2 \overline{V}_{NT}$.

Table 1: Biases and Standard Deviations for $Y_{it} \mid \alpha, \gamma, \beta \sim \mathcal{N}(\alpha_i + \gamma_t, \beta)$

| | N = 10 | N=25 | | N=50 | | |
|---|---|---|---|---|---|---|
| | T = 10 | T=10 | T=25 | T=10 | T=25 | T=50 |
| $(\overline{\beta}_{NT} - \beta^0)/\beta^0$ | -.19 | -.14 | -.08 | -.12 | -.06 | -.04 |
| $(\overline{\beta}^A_{NT} - \beta^0)/\beta^0$ | -.03 | -.02 | .00 | -.01 | -.01 | .00 |
| $(\overline{\beta}^J_{NT} - \beta^0)/\beta^0$ | -.01 | .00 | .00 | .00 | .00 | .00 |
| $\sqrt{\overline{V}_{NT}}/\beta^0$ | .13 | .08 | .05 | .06 | .04 | .03 |
| $\sqrt{\overline{V}^A_{NT}}/\beta^0$ | .14 | .09 | .06 | .06 | .04 | .03 |

Table 2 illustrates the effect of the bias on the inference based on the asymptotic distribution. It shows the coverage probabilities of 95% asymptotic confidence intervals for $\beta^0$ constructed in the usual way as

$$\text{CI}_{.95}(\widehat{\beta}) = \widehat{\beta} \pm 1.96 \widehat{V}_{NT}^{1/2} = \widehat{\beta}(1 \pm 1.96\sqrt{2/(NT)})$$

where $\widehat{\beta} = \{\widehat{\beta}_{NT}, \widetilde{\beta}^A_{NT}\}$ and $\widehat{V}_{NT} = 2\widehat{\beta}^2/(NT)$ is an estimator of the asymptotic variance $\overline{V}_\infty/(NT) = 2(\beta^0)^2/(NT)$. To find the exact probabilities, we use that $NT\widehat{\beta}_{NT}/\beta^0 \sim \chi^2_{(N-1)(T-1)}$ and $\widetilde{\beta}^A_{NT} = (1 + 1/N + 1/T)\widehat{\beta}_{NT}$. These probabilities do not depend on the value of $\beta^0$ because the limits of the intervals are proportional to $\widehat{\beta}$. As a benchmark of comparison, we also consider confidence intervals constructed from the unbiased estimator $\widetilde{\beta}_{NT} = NT\widehat{\beta}_{NT}/[(N-1)(T-1)]$. Here we find that the confidence intervals based on the fixed effect estimator display severe undercoverage for all the sample sizes. The confidence intervals based on the corrected estimators have high coverage probabilities, which approach the nominal level as the sample size grows. Moreover, the bias corrected estimator produces confidence intervals with very similar coverage probabilities to the ones from the unbiased estimator.

## 4 Asymptotic Theory for Bias Corrections

In nonlinear panel data models the population problem (3.1) generally does not have closed form solution, so we need to rely on asymptotic arguments to characterize the terms in the expansion of the bias (3.2)

Table 2: Coverage probabilities for $Y_{it} \mid \alpha, \gamma, \beta \sim \mathcal{N}(\alpha_i + \gamma_t, \beta)$

| | N = 10 | N=25 | | N=50 | | |
|---|---|---|---|---|---|---|
| | T = 10 | T=10 | T=25 | T=10 | T=25 | T=50 |
| $\text{CI}_{.95}(\widehat{\beta}_{NT})$ | .56 | .55 | .65 | .44 | .63 | .68 |
| $\text{CI}_{.95}(\widetilde{\beta}^A_{NT})$ | .89 | .92 | .93 | .92 | .94 | .94 |
| $\text{CI}_{.95}(\widetilde{\beta}_{NT})$ | .91 | .93 | .94 | .93 | .94 | .94 |

Nominal coverage probability is .95.

and to justify the validity of the corrections.

## 4.1 Asymptotic distribution of model parameters

We consider panel models with scalar individual and time effects that enter the likelihood function additively through $\pi_{it} = \alpha_i + \gamma_t$. In these models the dimension of the incidental parameters is $\dim \phi_{NT} = N + T$. The leading cases are single index models, where the dependence of the likelihood function on the parameters is through an index $X'_{it}\beta + \alpha_i + \gamma_t$. These models cover the probit and Poisson specifications of Examples 1 and 2. Moreover, the additive structure only applies to the unobserved effects, so we can allow for scale parameters to cover the Tobit and negative binomial models. We focus on these additive models for computational tractability and because we can establish the consistency of the fixed effects estimators under a concavity assumption in the log-likelihood function with respect to all the parameters.

The parametric part of our panel models takes the form

$$\log f_Y(Y_{it} \mid X_{it}, \alpha_i, \gamma_t, \beta) = \ell_{it}(\beta, \pi_{it}). \tag{4.1}$$

We denote the derivatives of the log-likelihood function $\ell_{it}$ by $\partial_\beta \ell_{it}(\beta, \pi) := \partial \ell_{it}(\beta, \pi)/\partial\beta$, $\partial_{\beta\beta'}\ell_{it}(\beta, \pi) := \partial^2 \ell_{it}(\beta, \pi)/(\partial\beta\partial\beta')$, $\partial_{\pi^q}\ell_{it}(\beta, \pi) := \partial^q \ell_{it}(\beta, \pi)/\partial\pi^q$, $q = 1, 2, 3$, *etc.* We drop the arguments $\beta$ and $\pi$ when the derivatives are evaluated at the true parameters $\beta^0$ and $\pi^0_{it} := \alpha^0_i + \gamma^0_t$, e.g. $\partial_{\pi^q}\ell_{it} := \partial_{\pi^q}\ell_{it}(\beta^0, \pi^0_{it})$. We also drop the dependence on $NT$ from all the sequences of functions and parameters, e.g. we use $\mathcal{L}$ for $\mathcal{L}_{NT}$ and $\phi$ for $\phi_{NT}$.

We make the following assumptions:

**Assumption 4.1** (Panel models)**.** *Let $\nu > 0$ and $\mu > 4(8+\nu)/\nu$. Let $\varepsilon > 0$ and let $\mathcal{B}^0_\varepsilon$ be a subset of $\mathbb{R}^{\dim \beta + 1}$ that contains an $\varepsilon$-neighbourhood of $(\beta^0, \pi^0_{it})$ for all $i, t, N, T$.*

*(i) Asymptotics: we consider limits of sequences where $N/T \to \kappa^2$, $0 < \kappa < \infty$, as $N, T \to \infty$.*

*(ii) Sampling: conditional on $\phi$, $\{(Y_i^T, X_i^T) : 1 \leq i \leq N\}$ is independent across $i$ and, for each $i$, $\{(Y_{it}, X_{it}) : 1 \leq t \leq T\}$ is $\alpha$-mixing with mixing coefficients satisfying $\sup_i a_i(m) = \mathcal{O}(m^{-\mu})$ as $m \to \infty$, where*

$$a_i(m) := \sup_t \sup_{A \in \mathcal{A}^i_t, B \in \mathcal{B}^i_{t+m}} |P(A \cap B) - P(A)P(B)|,$$

*and for* $Z_{it} = (Y_{it}, X_{it})$, $\mathcal{A}_t^i$ *is the sigma field generated by* $(Z_{it}, Z_{i,t-1}, \ldots)$, *and* $\mathcal{B}_t^i$ *is the sigma field generated by* $(Z_{it}, Z_{i,t+1}, \ldots)$.

*(iii) Model: for* $X_i^t = \{X_{is} : s = 1, ..., t\}$, *we assume that for all* $i, t, N, T$,

$$Y_{it} \mid X_i^t, \phi, \beta \sim \exp[\ell_{it}(\beta, \alpha_i + \gamma_t)].$$

*The realizations of the parameters and unobserved effects that generate the observed data are denoted by* $\beta^0$ *and* $\phi^0$.

*(iv) Smoothness and moments: We assume that* $(\beta, \pi) \mapsto \ell_{it}(\beta, \pi)$ *is four times continuously differentiable over* $\mathcal{B}_\varepsilon^0$ *a.s. The partial derivatives of* $\ell_{it}(\beta, \pi)$ *with respect to the elements of* $(\beta, \pi)$ *up to fourth order are bounded in absolute value uniformly over* $(\beta, \pi) \in \mathcal{B}_\varepsilon^0$ *by a function* $M(Z_{it}) > 0$ *a.s., and* $\max_{i,t} \mathbb{E}_\phi[M(Z_{it})^{8+\nu}]$ *is a.s. uniformly bounded over* $N, T$.

*(v) Concavity: For all* $N, T$, $(\beta, \phi) \mapsto \mathcal{L}(\beta, \phi) = (NT)^{-1/2}\{\sum_{i,t} \ell_{it}(\beta, \alpha_i + \gamma_t) - b(v'\phi)^2/2\}$ *is strictly concave over* $\mathbb{R}^{\dim \beta + N + T}$ *a.s. Furthermore, there exist constants* $b_{\min}$ *and* $b_{\max}$ *such that for all* $(\beta, \pi) \in \mathcal{B}_\varepsilon^0$, $0 < b_{\min} \leq -\mathbb{E}_\phi[\partial_{\pi^2}\ell_{it}(\beta, \pi)] \leq b_{\max}$ *a.s. uniformly over* $i, t, N, T$.

Assumption 4.1(*i*) defines the large-$T$ asymptotic framework and is the same as in Hahn and Kuersteiner (2011). Assumption 4.1(*ii*) does not impose identical distribution nor stationarity over the time series dimension, conditional on the unobserved effects, unlike most of the large-$T$ panel literature, e.g., Hahn and Newey (2004) and Hahn and Kuersteiner (2011). These assumptions are violated by the presence of the time effects, because they are treated as parameters. The mixing condition is used to bound covariances and moments in the application of laws of large numbers and central limit theorems – it could replaced by other conditions that guarantee the applicability of these results.

Assumption 4.1(*iii*) is the parametric part of the panel model. We rely on this assumption to guarantee that $\partial_\beta \ell_{it}$ and $\partial_\pi \ell_{it}$ have martingale difference properties. Moreover, we use certain Bartlett identities implied by this assumption to simplify some expressions, but those simplifications are not crucial for our results. Assumption 4.1(*iv*) imposes smoothness and moment conditions in the log-likelihood function and its derivatives. These conditions guarantee that the higher-order stochastic expansions of the fixed effect estimator that we use to characterize the asymptotic bias are well-defined, and that the remainder terms of these expansions are bounded. The most commonly used nonlinear models in applied economics such as logit, probit, ordered probit, Poisson, and Tobit models have smooth log-likelihoods functions that satisfy the concavity condition of Assumption 4.1(*v*), provided that all the elements of $X_{it}$ have cross sectional and time series variation.

To describe the asymptotic distribution of the fixed effects estimator $\widehat{\beta}$, it is convenient to introduce some additional notation. Let $\overline{\mathcal{H}}$ be the $(N+T) \times (N+T)$ expected Hessian matrix of the log-likelihood with respect to the nuisance parameters evaluated at the true parameters, i.e.

$$\overline{\mathcal{H}} = \mathbb{E}_\phi[-\partial_{\phi\phi'}\mathcal{L}] = \begin{pmatrix} \overline{\mathcal{H}}^*_{(\alpha\alpha)} & \overline{\mathcal{H}}^*_{(\alpha\gamma)} \\ [\overline{\mathcal{H}}^*_{(\alpha\gamma)}]' & \overline{\mathcal{H}}^*_{(\gamma\gamma)} \end{pmatrix} + \frac{b}{\sqrt{NT}} vv', \tag{4.2}$$

where $\overline{\mathcal{H}}^*_{(\alpha\alpha)} = \operatorname{diag}(\sum_t \mathbb{E}_\phi[-\partial_{\pi^2}\ell_{it}])/\sqrt{NT}$, $\overline{\mathcal{H}}^*_{(\alpha\gamma)it} = \mathbb{E}_\phi[-\partial_{\pi^2}\ell_{it}]/\sqrt{NT}$, and $\overline{\mathcal{H}}^*_{(\gamma\gamma)} = \operatorname{diag}(\sum_i \mathbb{E}_\phi[-\partial_{\pi^2}\ell_{it}])/\sqrt{NT}$. Furthermore, let $\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}$, $\overline{\mathcal{H}}^{-1}_{(\alpha\gamma)}$, $\overline{\mathcal{H}}^{-1}_{(\gamma\alpha)}$, and $\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}$ denote the $N \times N$, $N \times T$, $T \times N$ and $T \times T$ blocks of the inverse $\overline{\mathcal{H}}^{-1}$ of $\overline{\mathcal{H}}$. It is convenient to define the dim $\beta$-vector $\Xi_{it}$ and the operator $D_{\beta\pi^q}$ by

$$\Xi_{it} := -\frac{1}{\sqrt{NT}} \sum_{j=1}^{N} \sum_{\tau=1}^{T} \left( \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)ij} + \overline{\mathcal{H}}^{-1}_{(\gamma\alpha)tj} + \overline{\mathcal{H}}^{-1}_{(\alpha\gamma)i\tau} + \overline{\mathcal{H}}^{-1}_{(\gamma\gamma)t\tau} \right) \mathbb{E}_\phi \left( \partial_{\beta\pi} \ell_{j\tau} \right),$$

$$D_{\beta\pi^q} \ell_{it} := \partial_{\beta\pi^q} \ell_{it} - \partial_{\pi^{q+1}} \ell_{it} \Xi_{it}, \tag{4.3}$$

with $q = 0, 1, 2$. The $k$-th component of $\Xi_{it}$ corresponds to the population least squares projection of $\mathbb{E}_\phi(\partial_{\beta_k\pi}\ell_{it})/\mathbb{E}_\phi(\partial_{\pi^2}\ell_{it})$ on the space spanned by the incidental parameters under a metric given by $\mathbb{E}_\phi(-\partial_{\pi^2}\ell_{it})$, i.e.

$$\Xi_{it,k} = \alpha^*_{i,k} + \gamma^*_{t,k}, \qquad (\alpha^*_k, \gamma^*_k) = \underset{\alpha_{i,k}, \gamma_{t,k}}{\operatorname{argmin}} \sum_{i,t} \mathbb{E}_\phi(-\partial_{\pi^2}\ell_{it}) \left( \frac{\mathbb{E}_\phi(\partial_{\beta_k\pi}\ell_{it})}{\mathbb{E}_\phi(\partial_{\pi^2}\ell_{it})} - \alpha_{i,k} - \gamma_{t,k} \right)^2.$$

The operator $D_{\beta\pi^q}$ generalizes the individual and time differencing transformations from linear models to nonlinear models. To see this, consider the normal linear model $Y_{it} \mid X^t_i, \alpha_i, \gamma_t \sim \mathcal{N}(X'_{it}\beta + \alpha_i + \gamma_t, 1)$. Then, $\Xi_{it} = T^{-1}\sum_{t=1}^T \mathbb{E}_\phi[X_{it}] + N^{-1}\sum_{i=1}^N \mathbb{E}_\phi[X_{it}] - (NT)^{-1}\sum_{i=1}^N\sum_{t=1}^T \mathbb{E}_\phi[X_{it}]$, $D_\beta\ell_{it} = -\tilde{X}_{it}\varepsilon_{it}$, $D_{\beta\pi}\ell_{it} = -\tilde{X}_{it}$, and $D_{\beta\pi^2}\ell_{it} = 0$, where $\varepsilon_{it} = Y_{it} - X'_{it}\beta - \alpha_i - \gamma_t$ and $\tilde{X}_{it} = X_{it} - \Xi_{it}$ is the individual and time demeaned explanatory variable.

Let $\overline{\mathbb{E}} := \operatorname{plim}_{N,T\to\infty}$. The following theorem establishes the asymptotic distribution of the fixed effects estimator $\widehat{\beta}$.

**Theorem 4.1** (Asymptotic distribution of $\widehat{\beta}$)**.** *Suppose that Assumption 4.1 holds, that the following limits exist*

$$\overline{B}_\infty = -\overline{\mathbb{E}}\left[ \frac{1}{N}\sum_{i=1}^N \frac{\sum_{t=1}^T\sum_{\tau=t}^T \mathbb{E}_\phi\left(\partial_\pi\ell_{it}D_{\beta\pi}\ell_{i\tau}\right) + \frac{1}{2}\sum_{t=1}^T \mathbb{E}_\phi(D_{\beta\pi^2}\ell_{it})}{\sum_{t=1}^T \mathbb{E}_\phi\left(\partial_{\pi^2}\ell_{it}\right)} \right],$$

$$\overline{D}_\infty = -\overline{\mathbb{E}}\left[ \frac{1}{T}\sum_{t=1}^T \frac{\sum_{i=1}^N \mathbb{E}_\phi\left(\partial_\pi\ell_{it}D_{\beta\pi}\ell_{it} + \frac{1}{2}D_{\beta\pi^2}\ell_{it}\right)}{\sum_{i=1}^N \mathbb{E}_\phi\left(\partial_{\pi^2}\ell_{it}\right)} \right],$$

$$\overline{W}_\infty = -\overline{\mathbb{E}}\left[ \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \mathbb{E}_\phi\left(\partial_{\beta\beta'}\ell_{it} - \partial_{\pi^2}\ell_{it}\Xi_{it}\Xi'_{it}\right) \right],$$

*and that* $\overline{W}_\infty > 0$*. Then,*

$$\sqrt{NT}\left(\widehat{\beta} - \beta^0\right) \to_d \overline{W}_\infty^{-1}\mathcal{N}(\kappa\overline{B}_\infty + \kappa^{-1}\overline{D}_\infty, \overline{W}_\infty).$$

**Sketch of the Proof.** The detailed proof of Theorem 4.1 is provided in the appendix. Here we include a summary of the main ideas behind the proof.

We start by noting that the existing results for large $N, T$ panels, which were developed for models with only individual effects, cannot be sequentially applied to the two dimensions of the panel to derive the asymptotic distribution of our estimators. These results usually start with a consistency proof that relies on partitioning the log-likelihood in the sum of individual log-likelihoods that depend on a fixed number of parameters, the model parameter $\beta$ and the corresponding individual effect $\alpha_i$. Then, the maximizers of the individual log-likelihood are shown to be consistent estimators of all the parameters as $T$ becomes large using standard arguments . In the presence of time effects there is no partition of the data that is only affected by a fixed number of parameters, and whose size grows with the sample size. We thus require a new approach.

Our approach consists of deriving an asymptotic approximation to the score of the profile log-likelihood, $\partial_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta))$, which is valid locally around $\beta = \beta^0$. We use this approximation to show that there is a solution to the first order condition $\partial_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta)) = 0$ that is close to $\beta^0$ asymptotically, and to characterize the asymptotic properties of this solution. Under the assumption that the log-likelihood is strictly concave, the solution to $\partial_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta)) = 0$ uniquely determines the maximizer $\widehat{\beta}$, so that we do not need a separate proof of consistency to obtain the asymptotic distribution of our estimators.

We derive the asymptotic approximation to $\partial_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta))$ using a second-order Taylor stochastic expansion. This expansion does not rely on the panel structure of the model, but it requires sufficient differentiability of $\mathcal{L}(\beta, \phi)$ and that each incidental parameter affects a subset (namely all observations from individual $i$ for each individual effect $\alpha_i$, and all observations from time period $t$ for each time effect $\gamma_t$) whose size grows with the sample size. For our panel model, the latter implies that the score of the incidental parameters,

$$\mathcal{S}(\beta, \phi) = \partial_\phi \mathcal{L}(\beta, \phi) = \begin{pmatrix} \left[\frac{1}{\sqrt{NT}} \sum_{t=1}^T \partial_\pi \ell_{it}(\beta, \alpha_i + \gamma_t)\right]_{i=1,\ldots,N} \\ \left[\frac{1}{\sqrt{NT}} \sum_{i=1}^N \partial_\pi \ell_{it}(\beta, \alpha_i + \gamma_t)\right]_{t=1,\ldots,T} \end{pmatrix},$$

is of a smaller order at the true parameters $(\beta^0, \phi^0)$ than it is at other parameter values. The entries of $\mathcal{S}(\beta, \phi)$ are of order one generically as $N$ and $T$ grow at the same rate, while the entries of $\mathcal{S} := \mathcal{S}(\beta^0, \phi^0)$ are of order $1/\sqrt{N}$ or $1/\sqrt{T}$. This allows us to bound the higher-order terms in a expansion of $\partial_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta))$ in $\beta$ and $\mathcal{S}(\beta, \phi)$ around $\beta_0$ and $\mathcal{S}$.

The stochastic expansion of $\partial_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta))$ can be obtained in different ways. We find convenient to do it through the Legendre-transformed objective function $\mathcal{L}^*(\beta, S) = \max_\phi [\mathcal{L}(\beta, \phi) - \phi' S]$. This function has the properties: $\mathcal{L}^*(\beta, 0) = \mathcal{L}(\beta, \widehat{\phi}(\beta))$, $\mathcal{L}^*(\beta, \mathcal{S}) = \mathcal{L}(\beta, \phi^0) - \phi^{0\prime}\mathcal{S}$, and $\partial_\beta \mathcal{L}(\beta, \phi^0) = \partial_\beta \mathcal{L}^*(\beta, \mathcal{S})$. The expansion of $\partial_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta)) = \partial_\beta \mathcal{L}^*(\beta, 0)$ can therefore be obtained as a Taylor stochastic expansion of $\partial_\beta \mathcal{L}^*(\beta, S)$ in $(\beta, S)$ around $(\beta^0, \mathcal{S})$ and evaluated at $(\beta, 0)$, see Appendix B for details.

Theorem B.1 gives the stochastic expansion. To obtain the asymptotic distribution of $\widehat{\beta}$ from the expansion, we need to analyze the expected Hessian of the incidental parameters $\overline{\mathcal{H}}$, which is defined in (4.2) for our panel model. More precisely, we need to characterize the asymptotic properties of the inverse of $\overline{\mathcal{H}}$, because this inverse features prominently in the expansion. For models with only individual effects, $\overline{\mathcal{H}}$ is diagonal and its inversion poses no difficulty. In our case $\overline{\mathcal{H}}$ has strong diagonal elements of order 1 and off-diagonal elements of order $(NT)^{-1/2}$. The off-diagonal elements reflect that the individual and time effects are compounded in a non-trivial way. They are of smaller order than the strong diagonal elements, but cannot simply be ignored in the inversion because the number of them is very large and grows with the sample size. For example, the Hessian $\overline{\mathcal{H}}^*$ without penalty has the same structure as $\overline{\mathcal{H}}$, but is not invertible. Lemma D.8 shows that $\overline{\mathcal{H}}$ is invertible, and that $\overline{\mathcal{H}}^{-1}$ has the same structure as $\overline{\mathcal{H}}$, namely strong diagonal elements of order 1 and off-diagonal elements of order $(NT)^{-1/2}$. This result explains why the double incidental parameter problem due to the individual and time effects decouples asymptotically, so we get that the bias has two leading terms of orders $T^{-1}$ and $N^{-1}$. This result agrees with the intuition that one would draw from analyzing separately the incidental parameter problem in each dimension, but without a formal derivation it was not clear that the asymptotic bias also has the simple additive structure in the joint analysis. ■

**Remark 1** (Bias expressions for Conditional Moment Models)**.** *In the derivation of the asymptotic bias, we apply Bartlett identities implied by Assumption 4.1(iii) to simplify the expressions. The following expressions of the bias do not make use of these identities and therefore remain valid in conditional moment settings that do not specify the entire conditional distribution of* $Y_{it}$*:*

$$\overline{B}_\infty = -\overline{\mathbb{E}}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{t=1}^{T}\sum_{\tau=t}^{T}\mathbb{E}_\phi\left(\partial_\pi \ell_{it} D_{\beta\pi}\ell_{i\tau}\right)}{\sum_{t=1}^{T}\mathbb{E}_\phi\left(\partial_{\pi^2}\ell_{it}\right)}\right] + \frac{1}{2}\,\overline{\mathbb{E}}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{t=1}^{T}\mathbb{E}_\phi[(\partial_\pi\ell_{it})^2]\sum_{t=1}^{T}\mathbb{E}_\phi(D_{\beta\pi^2}\ell_{it})}{\left[\sum_{t=1}^{T}\mathbb{E}_\phi\left(\partial_{\pi^2}\ell_{it}\right)\right]^2}\right],$$

$$\overline{D}_\infty = -\overline{\mathbb{E}}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\sum_{i=1}^{N}\mathbb{E}_\phi\left[\partial_\pi \ell_{it} D_{\beta\pi}\ell_{it}\right]}{\sum_{i=1}^{N}\mathbb{E}_\phi\left(\partial_{\pi^2}\ell_{it}\right)}\right] + \frac{1}{2}\,\overline{\mathbb{E}}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\sum_{i=1}^{N}\mathbb{E}_\phi[(\partial_\pi\ell_{it})^2]\sum_{i=1}^{N}\mathbb{E}_\phi(D_{\beta\pi^2}\ell_{it})}{\left[\sum_{i=1}^{N}\mathbb{E}_\phi\left(\partial_{\pi^2}\ell_{it}\right)\right]^2}\right].$$

*For example, consider the least squares fixed effects estimator in a linear model* $Y_{it} = X_{it}'\beta + \alpha_i + \gamma_t + \varepsilon_{it}$ *with* $\mathbb{E}[\varepsilon_{it} \mid X_i^t, \phi, \beta] = 0$*. Applying the previous expressions to* $\ell_{it}(\beta,\pi) = -(Y_{it} - X_{it}'\beta - \alpha_i - \gamma_t)^2$ *yields*

$$\overline{B}_\infty = -\overline{\mathbb{E}}\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{\tau=t+1}^{T}\mathbb{E}_\phi\left(X_{i\tau}\varepsilon_{it}\right)\right]$$

*and* $\overline{D}_\infty = 0$*. The expression for* $\overline{B}_\infty$ *corresponds to Nickell (1981) bias formula when* $X_{it} = Y_{i,t-1}$*. If* $\mathbb{E}[\varepsilon_{it} \mid X_i^T, \phi, \beta] = 0$*, i.e.* $X_{it}$ *is strictly exogenous with respect to* $\varepsilon_{it}$*, then we get the well-known result for linear models of no asymptotic bias,* $\overline{B}_\infty = \overline{D}_\infty = 0$*.*

It is instructive to evaluate the expressions of the bias in our running examples.

**Example 1** (Binary response model). *In this case*

$$\ell_{it}(\beta,\pi) = Y_{it}\log F(X_{it}'\beta + \pi) + (1 - Y_{it})\log[1 - F(X_{it}'\beta + \pi)],$$

*so that* $\partial_\pi\ell_{it} = H_{it}(Y_{it} - F_{it})$*,* $\partial_\beta\ell_{it} = \partial_\pi\ell_{it}X_{it}$*,* $\partial_{\pi^2}\ell_{it} = -H_{it}\partial F_{it} + \partial H_{it}(Y_{it} - F_{it})$*,* $\partial_{\beta\beta'}\ell_{it} = \partial_{\pi^2}\ell_{it}X_{it}X_{it}'$*,* $\partial_{\beta\pi}\ell_{it} = \partial_{\pi^2}\ell_{it}X_{it}$*,* $\partial_{\pi^3}\ell_{it} = -H_{it}\partial^2 F_{it} - 2\partial H_{it}\partial F_{it} + \partial^2 H_{it}(Y_{it} - F_{it})$*, and* $\partial_{\beta\pi^2}\ell_{it} = \partial_{\pi^3}\ell_{it}X_{it}$*, where* $H_{it} = \partial F_{it}/[F_{it}(1 - F_{it})]$*, and* $\partial^j G_{it} := \partial^j G(Z)|_{Z = X_{it}'\beta^0 + \pi_{it}^0}$ *for any function* $G$ *and*

$j = 0, 1, 2$. *Substituting these values in the expressions of the bias of Theorem 4.1 yields*

$$
\begin{aligned}
\overline{B}_\infty &= -\overline{\mathbb{E}} \left[ \frac{1}{2N} \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} \left\{ \mathbb{E}_\phi [H_{it} \partial^2 F_{it} \tilde{X}_{it}] + 2 \sum_{\tau=t+1}^{T} \mathbb{E}_\phi \left[ H_{it}(Y_{it} - F_{it}) \omega_{i\tau} \tilde{X}_{i\tau} \right] \right\}}{\sum_{t=1}^{T} \mathbb{E}_\phi (\omega_{it})} \right], \\
\overline{D}_\infty &= -\overline{\mathbb{E}} \left[ \frac{1}{2T} \sum_{t=1}^{T} \frac{\sum_{i=1}^{N} \mathbb{E}_\phi [H_{it} \partial^2 F_{it} \tilde{X}_{it}]}{\sum_{i=1}^{N} \mathbb{E}_\phi (\omega_{it})} \right], \\
\overline{W}_\infty &= \overline{\mathbb{E}} \left[ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}_\phi [\omega_{it} \tilde{X}_{it} \tilde{X}_{it}'] \right],
\end{aligned}
$$

*where* $\omega_{it} = H_{it} \partial F_{it}$ *and* $\tilde{X}_{it}$ *is the residual of the population projection of* $X_{it}$ *on the space spanned by the incidental parameters under a metric weighted by* $\mathbb{E}_\phi(\omega_{it})$*. For the probit model with all the components of* $X_{it}$ *strictly exogenous,*

$$
\overline{B}_\infty = \overline{\mathbb{E}} \left[ \frac{1}{2N} \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} \mathbb{E}_\phi [\omega_{it} \tilde{X}_{it} \tilde{X}_{it}']}{\sum_{t=1}^{T} \mathbb{E}_\phi (\omega_{it})} \right] \beta^0, \quad \overline{D}_\infty = \overline{\mathbb{E}} \left[ \frac{1}{2T} \sum_{t=1}^{T} \frac{\sum_{i=1}^{N} \mathbb{E}_\phi [\omega_{it} \tilde{X}_{it} \tilde{X}_{it}']}{\sum_{i=1}^{N} \mathbb{E}_\phi (\omega_{it})} \right] \beta^0.
$$

*The asymptotic bias is therefore a positive definite matrix weighted average of the true parameter value as in the case of the probit model with only individual effects (Fernández-Val, 2009).*

**Example 2** (Count response model). *In this case*

$$
\ell_{it}(\beta, \pi) = (X_{it}'\beta + \pi) Y_{it} - \exp(X_{it}'\beta + \pi) - \log Y_{it}!,
$$

*so that* $\partial_\pi \ell_{it} = Y_{it} - \omega_{it}$, $\partial_\beta \ell_{it} = \partial_\pi \ell_{it} X_{it}$, $\partial_{\pi^2} \ell_{it} = \partial_{\pi^3} \ell_{it} = -\omega_{it}$, $\partial_{\beta\beta'} \ell_{it} = \partial_{\pi^2} \ell_{it} X_{it} X_{it}'$, *and* $\partial_{\beta\pi} \ell_{it} = \partial_{\beta\pi^2} \ell_{it} = \partial_{\pi^3} \ell_{it} X_{it}$, *where* $\omega_{it} = \exp(X_{it}'\beta^0 + \pi_{it}^0)$*. Substituting these values in the expressions of the bias of Theorem 4.1 yields*

$$
\begin{aligned}
\overline{B}_\infty &= -\overline{\mathbb{E}} \left[ \frac{1}{N} \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} \sum_{\tau=t+1}^{T} \mathbb{E}_\phi \left[ (Y_{it} - \omega_{it}) \omega_{i\tau} \tilde{X}_{i\tau} \right]}{\sum_{t=1}^{T} \mathbb{E}_\phi (\omega_{it})} \right], \\
\overline{W}_\infty &= \overline{\mathbb{E}} \left[ \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}_\phi [\omega_{it} \tilde{X}_{it} \tilde{X}_{it}'] \right],
\end{aligned}
$$

*and* $\overline{D}_\infty = 0$*, where* $\tilde{X}_{it}$ *is the residual of the population projection of* $X_{it}$ *on the space spanned by the incidental parameters under a metric weighted by* $\mathbb{E}_\phi(\omega_{it})$*. If in addition all the components of* $X_{it}$ *are strictly exogenous, then we get the no asymptotic bias result* $\overline{B}_\infty = \overline{D}_\infty = 0$*.*

## 4.2 Asymptotic distribution of APEs

In nonlinear models we are often interested in APEs, in addition to the model parameters. These effects are averages of the data, parameters and unobserved effects; see expression (2.2). For the panel models of Assumption 4.1 we specify the partial effects as $\Delta(X_{it}, \beta, \alpha_i, \gamma_t) = \Delta_{it}(\beta, \pi_{it})$. The restriction that the partial effects depend on $\alpha_i$ and $\gamma_t$ through $\pi_{it}$ is natural in our panel models since

$$
\mathbb{E}[Y_{it} \mid X_i^t, \alpha_i, \gamma_t, \beta] = \int y \exp[\ell_{it}(\beta, \pi_{it})] dy,
$$

and the partial effects are usually defined as differences or derivatives of this conditional expectation with respect to the components of $X_{it}$. For example, the partial effects for the probit and Poisson models described in Section 2 satisfy this restriction.

The distribution of the unobserved individual and time effects is not ancillary for the APEs, unlike for model parameters. We therefore need to make assumptions on this distribution to define and interpret the APEs, and to derive the asymptotic distribution of their estimators. Here, there are several possibilities depending on whether we define the APE conditional or unconditional on the unobserved effects. For conditional APEs, we treat the unobserved effects as deterministic. In this case $\mathbb{E}[\Delta_{it}] = \mathbb{E}_\phi[\Delta_{it}]$ and $\delta^0_{NT} = (NT)^{-1}\sum_{i,t}\mathbb{E}_\phi[\Delta_{it}]$ can change over $T$ and $N$ in a deterministic fashion. For unconditional APEs, we control the heterogeneity of the partial effects assuming that the individual effects and explanatory variables are identically distributed cross sectionally and/or stationary over time. If $(X_{it}, \alpha_i, \gamma_t)$ is identically distributed over $i$ and can be heterogeneously distributed over $t$, $\mathbb{E}[\Delta_{it}] = \delta^0_t$ and $\delta^0_{NT} = T^{-1}\sum_{t=1}^T \delta^0_t$ changes only with $T$. If $(X_{it}, \alpha_i, \gamma_t)$ is stationary over $t$ and can be heterogeneously distributed over $i$, $\mathbb{E}[\Delta_{it}] = \delta^0_i$ and $\delta^0_{NT} = N^{-1}\sum_{i=1}^N \delta^0_i$ changes only with $N$. Finally, if $(X_{it}, \alpha_i, \gamma_t)$ is identically distributed over $i$ and stationary over $t$, $\mathbb{E}[\Delta_{it}] = \delta^0_{NT}$ and $\delta^0_{NT} = \delta^0$ does not change with $N$ and $T$.

We also impose smoothness and moment conditions on the function $\Delta$ that defines the partial effects. We use these conditions to derive higher-order stochastic expansions for the fixed effect estimator of the APEs and to bound the remainder terms in these expansions. Let $\{\alpha_i\}_N := \{\alpha_i : 1 \le i \le N\}$, $\{\gamma_t\}_T := \{\gamma_t : 1 \le t \le T\}$, and $\{X_{it}, \alpha_i, \gamma_t\}_{NT} := \{(X_{it}, \alpha_i, \gamma_t) : 1 \le i \le N, 1 \le t \le T\}$.

**Assumption 4.2** (Partial effects). *Let $\nu > 0$, $\epsilon > 0$, and $\mathcal{B}^0_\varepsilon$ all be as in Assumption 4.1.*

*(i) Sampling: for all $N, T$, (a) $\{\alpha_i\}_N$ and $\{\gamma_t\}_T$ are deterministic; or (b) $\{X_{it}, \alpha_i, \gamma_t\}_{NT}$ is identically distributed across $i$ and/or stationary across $t$.*

*(ii) Model: for all $i, t, N, T$, the partial effects depend on $\alpha_i$ and $\gamma_t$ through $\alpha_i + \gamma_t$:*

$$\Delta(X_{it}, \beta, \alpha_i, \gamma_t) = \Delta_{it}(\beta, \alpha_i + \gamma_t).$$

*The realizations of the partial effects are denoted by $\Delta_{it} := \Delta_{it}(\beta^0, \alpha^0_i + \gamma^0_t)$.*

*(iii) Smoothness and moments: The function $(\beta, \pi) \mapsto \Delta_{it}(\beta, \pi)$ is four times continuously differentiable over $\mathcal{B}^0_\varepsilon$ a.s. The partial derivatives of $\Delta_{it}(\beta, \pi)$ with respect to the elements of $(\beta, \pi)$ up to fourth order are bounded in absolute value uniformly over $(\beta, \pi) \in \mathcal{B}^0_\varepsilon$ by a function $M(Z_{it}) > 0$ a.s., and $\max_{i,t}\mathbb{E}_\phi[M(Z_{it})^{8+\nu}]$ is a.s. uniformly bounded over $N, T$. Also, $\min_{i,t}[\mathbb{E}(\Delta^2_{it}) - \mathbb{E}(\Delta_{it})^2] > 0$, uniformly over $N, T$.*

Analogous to $\Xi_{it}$ in equation (4.3) we define

$$\Psi_{it} = -\frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{\tau=1}^{T}\left(\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)ij} + \overline{\mathcal{H}}^{-1}_{(\gamma\alpha)tj} + \overline{\mathcal{H}}^{-1}_{(\alpha\gamma)i\tau} + \overline{\mathcal{H}}^{-1}_{(\gamma\gamma)t\tau}\right)\partial_\pi\Delta_{j\tau}, \tag{4.4}$$

which is the population projection of $\partial_\pi\Delta_{j\tau}/\mathbb{E}_\phi[\partial_{\pi^2}\ell_{it}]$ on the space spanned by the incidental parameters under the metric given by $\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{it}]$. We use analogous notation to the previous section for the derivatives with respect to $\beta$ and higher order derivatives with respect to $\pi$.

Let $\delta^0_{NT}$ and $\widehat{\delta}$ be the APE and its fixed effects estimator, defined as in equations (2.2) and (2.8) with $\Delta(X_{it}, \beta, \alpha_i, \gamma_t) = \Delta_{it}(\beta, \alpha_i + \gamma_t)$.[8] The following theorem establishes the asymptotic distribution of $\widehat{\delta}$.

**Theorem 4.2** (Asymptotic distribution of $\widehat{\delta}$)**.** *Suppose that the assumptions of Theorem 4.1 and Assumption 4.2 hold, and that the following limits exist:*

$$\overline{(D_\beta \Delta)}_\infty = \overline{\mathbb{E}}\left[\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \mathbb{E}_\phi(\partial_\beta \Delta_{it} - \Xi_{it}\partial_\pi \Delta_{it})\right],$$

$$\overline{B}^\delta_\infty = \overline{(D_\beta \Delta)}'_\infty \overline{W}^{-1}_\infty \overline{B}_\infty + \overline{\mathbb{E}}\left[\frac{1}{N}\sum_{i=1}^N \frac{\sum_{t=1}^T\sum_{\tau=t}^T \mathbb{E}_\phi\left(\partial_\pi \ell_{it}\partial_{\pi^2}\ell_{i\tau}\Psi_{i\tau}\right)}{\sum_{t=1}^T \mathbb{E}_\phi\left(\partial_{\pi^2}\ell_{it}\right)}\right]$$
$$- \overline{\mathbb{E}}\left[\frac{1}{2\,N}\sum_{i=1}^N \frac{\sum_{t=1}^T\left[\mathbb{E}_\phi(\partial_{\pi^2}\Delta_{it}) - \mathbb{E}_\phi(\partial_{\pi^3}\ell_{it})\mathbb{E}_\phi(\Psi_{it})\right]}{\sum_{t=1}^T \mathbb{E}_\phi\left(\partial_{\pi^2}\ell_{it}\right)}\right],$$

$$\overline{D}^\delta_\infty = \overline{(D_\beta \Delta)}'_\infty \overline{W}^{-1}_\infty \overline{D}_\infty + \overline{\mathbb{E}}\left[\frac{1}{T}\sum_{t=1}^T \frac{\sum_{i=1}^N \mathbb{E}_\phi\left(\partial_\pi \ell_{it}\partial_{\pi^2}\ell_{it}\Psi_{it}\right)}{\sum_{i=1}^N \mathbb{E}_\phi\left(\partial_{\pi^2}\ell_{it}\right)}\right]$$
$$- \overline{\mathbb{E}}\left[\frac{1}{2T}\sum_{t=1}^T \frac{\sum_{i=1}^N\left[\mathbb{E}_\phi(\partial_{\pi^2}\Delta_{it}) - \mathbb{E}_\phi(\partial_{\pi^3}\ell_{it})\mathbb{E}_\phi(\Psi_{it})\right]}{\sum_{i=1}^N \mathbb{E}_\phi\left(\partial_{\pi^2}\ell_{it}\right)}\right],$$

$$\overline{V}^\delta_\infty = \overline{\mathbb{E}}\left\{\frac{r^2_{NT}}{N^2T^2}\mathbb{E}\left[\left(\sum_{i=1}^N\sum_{t=1}^T \widetilde{\Delta}_{it}\right)\left(\sum_{i=1}^N\sum_{t=1}^T \widetilde{\Delta}_{it}\right)' + \sum_{i=1}^N\sum_{t=1}^T \Gamma_{it}\Gamma'_{it}\right]\right\},$$

*for some deterministic sequence $r_{NT} \to \infty$ such that $r_{NT} = \mathcal{O}(\sqrt{NT})$ and $\overline{V}^\delta_\infty > 0$, where $\widetilde{\Delta}_{it} = \Delta_{it} - \mathbb{E}(\Delta_{it})$ and $\Gamma_{it} = \overline{(D_\beta \Delta)}'_\infty \overline{W}^{-1}_\infty D_\beta \ell_{it} - \mathbb{E}_\phi(\Psi_{it})\partial_\pi \ell_{it}$. Then,*

$$r_{NT}(\widehat{\delta} - \delta^0_{NT} - T^{-1}\overline{B}^\delta_\infty - N^{-1}\overline{D}^\delta_\infty) \to_d \mathcal{N}(0,\ \overline{V}^\delta_\infty).$$

**Remark 2** (Convergence rate, bias and variance)**.** *Under Assumption 4.2(i)(a), $\{\Delta_{it} : 1 \leq i \leq N, 1 \leq t \leq\}$ is independent across $i$ and $\alpha$-mixing across $t$ by Assumption 4.1(ii), so that $r_{NT} = \sqrt{NT}$ and*

$$\overline{V}^\delta_\infty = \overline{\mathbb{E}}\left\{\frac{r^2_{NT}}{N^2T^2}\sum_{i=1}^N\left[\sum_{t,\tau=1}^T \mathbb{E}(\widetilde{\Delta}_{it}\widetilde{\Delta}'_{i\tau}) + \sum_{t=1}^T \mathbb{E}(\Gamma_{it}\Gamma'_{it})\right]\right\}.$$

*Bias and variance are of the same order asymptotically under the asymptotic sequences of Assumption 4.1(i). Under Assumption 4.2(i)(b), the rate of convergence depends on the sampling properties of the unobserved effects. For example, if $\{\alpha_i\}_N$ and $\{\gamma_t\}_T$ are independent sequences, and $\alpha_i$ and $\gamma_t$ are independent for all $i, t$, then $r_{NT} = \sqrt{NT/(N+T)}$,*

$$\overline{V}^\delta_\infty = \overline{\mathbb{E}}\left\{\frac{r^2_{NT}}{N^2T^2}\sum_{i=1}^N\left[\sum_{t,\tau=1}^T \mathbb{E}(\widetilde{\Delta}_{it}\widetilde{\Delta}'_{i\tau}) + \sum_{j\neq i}\sum_{t=1}^T \mathbb{E}(\widetilde{\Delta}_{it}\widetilde{\Delta}'_{jt}) + \sum_{t=1}^T \mathbb{E}(\Gamma_{it}\Gamma'_{it})\right]\right\},$$

*and the asymptotic bias is of order $T^{-1/2} + N^{-1/2}$. The bias and the last term of $\overline{V}^\delta_\infty$ are asymptotically negligible in this case under the asymptotic sequences of Assumption 4.1(i).*

[8] We keep the dependence of $\delta^0_{NT}$ on $NT$ to distinguish $\delta^0_{NT}$ from $\delta^0 = \lim_{N,T\to\infty}\delta^0_{NT}$.

**Remark 3** (Average effects from bias corrected estimators)**.** *The first term in the expressions of the biases* $\overline{B}_\infty^\delta$ *and* $\overline{D}_\infty^\delta$ *comes from the bias of the estimator of* $\beta$*. It drops out when the APEs are constructed from asymptotically unbiased or bias corrected estimators of the parameter* $\beta$*, i.e.*

$$\widetilde{\delta} = \Delta(\widetilde{\beta}, \widehat{\phi}(\widetilde{\beta})),$$

*where* $\widetilde{\beta}$ *is such that* $\sqrt{NT}(\widetilde{\beta} - \beta^0) \to_d N(0, \overline{W}_\infty^{-1})$*. The asymptotic variance of* $\widetilde{\delta}$ *is the same as in Theorem 4.2.*

In the following examples we assume that the APEs are constructed from asymptotically unbiased estimators of the model parameters.

**Example 1** (Binary response model). *Consider the partial effects defined in (2.3) and (2.3) with*

$$\Delta_{it}(\beta, \pi) = F(\beta_k + X_{it,-k}'\beta_{-k} + \pi) - F(X_{it,-k}'\beta_{-k} + \pi) \text{ and } \Delta_{it}(\beta, \pi) = \beta_k \partial F(X_{it}'\beta + \pi).$$

*Using the notation previously introduced for this example, the components of the asymptotic bias of* $\widetilde{\delta}$ *are*

$$\overline{B}_\infty^\delta = \overline{\mathbb{E}}\left[\frac{1}{2N}\sum_{i=1}^{N} \frac{\sum_{t=1}^{T}\left[2\sum_{\tau=t+1}^{T}\mathbb{E}_\phi\left(H_{it}(Y_{it}-F_{it})\omega_{i\tau}\tilde{\Psi}_{i\tau}\right) - \mathbb{E}_\phi(\Psi_{it})\mathbb{E}_\phi(H_{it}\partial^2 F_{it}) + \mathbb{E}_\phi(\partial_{\pi^2}\Delta_{it})\right]}{\sum_{t=1}^{T}\mathbb{E}_\phi(\omega_{it})}\right],$$

$$\overline{D}_\infty^\delta = \overline{\mathbb{E}}\left[\frac{1}{2T}\sum_{t=1}^{T} \frac{\sum_{i=1}^{N}\left[-\mathbb{E}_\phi(\Psi_{it})\mathbb{E}_\phi(H_{it}\partial^2 F_{it}) + \mathbb{E}_\phi(\partial_{\pi^2}\Delta_{it})\right]}{\sum_{i=1}^{N}\mathbb{E}_\phi(\omega_{it})}\right],$$

*where* $\tilde{\Psi}_{i\tau}$ *is the residual of the population regression of* $\partial_\pi \Delta_{i\tau}/\mathbb{E}_\phi[\omega_{i\tau}]$ *on the space spanned by the incidental parameters under the metric given by* $\mathbb{E}_\phi[\omega_{i\tau}]$*. If all the components of* $X_{it}$ *are strictly exogenous, the first term of* $\overline{B}_\infty^\delta$ *is zero.*

**Example 2** (Count response model). *Consider the partial effect*

$$\Delta_{it}(\beta, \pi) = g_{it}(\beta)\exp(X_{it}'\beta + \pi),$$

*where* $g_{it}$ *does not depend on* $\pi$*. For example,* $g_{it}(\beta) = \beta_k + \beta_j h(Z_{it})$ *in (2.5). Using the notation previously introduced for this example, the components of the asymptotic bias are*

$$\overline{B}_\infty = \overline{\mathbb{E}}\left[\frac{1}{N}\sum_{i=1}^{N} \frac{\sum_{t=1}^{T}\sum_{\tau=t+1}^{T}\mathbb{E}_\phi\left[(Y_{it} - \omega_{it})\omega_{i\tau}\tilde{g}_{i\tau}\right]}{\sum_{t=1}^{T}\mathbb{E}_\phi(\omega_{it})}\right],$$

*and* $\overline{D}_\infty = 0$*, where* $\tilde{g}_{it}$ *is the residual of the population projection of* $g_{it}$ *on the space spanned by the incidental parameters under a metric weighted by* $\mathbb{E}_\phi[\omega_{it}]$*. The asymptotic bias is zero if all the components of* $X_{it}$ *are strictly exogenous or* $g_{it}(\beta)$ *is constant. The latter arises in the leading case of the partial effect of the k-th component of* $X_{it}$ *since* $g_{it}(\beta) = \beta_k$*. This no asymptotic bias result applies to any type of regressor, strictly exogenous or predetermined.*

### 4.3 Bias corrected estimators

The results of the previous sections show that the asymptotic distributions of the fixed effects estimators of the model parameters and APEs can have biases of the same order as the variances 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 describe how to construct analytical bias corrections for panel models and give conditions for the asymptotic validity of analytical and jackknife bias corrections.

The jackknife correction for the model parameter $\beta$ in equation (3.4) is generic and applies to the panel model. For the APEs, the jackknife correction is formed similarly as

$$\widetilde{\delta}^J_{NT} = 3\widehat{\delta}_{NT} - \widetilde{\delta}_{N,T/2} - \widetilde{\delta}_{N/2,T},$$

where $\widetilde{\delta}_{N,T/2}$ is the average of the 2 split jackknife estimators of the APE that leave out the first and second halves of the time periods, and let $\widetilde{\delta}_{N/2,T}$ is the average of the 2 split jackknife estimators of the APE that leave out half of the individuals.

The analytical corrections are constructed using sample analogs of the expressions in Theorems 4.1 and 4.2, replacing the true values of $\beta$ and $\phi$ by the fixed effects estimators. To describe these corrections, we introduce some additional notation. For any function of the data, unobserved effects and parameters $g_{itj}(\beta, \alpha_i + \gamma_t, \alpha_i + \gamma_{t-j})$ with $0 \leq j < t$, let $\widehat{g}_{itj} = g_{it}(\widehat{\beta}, \widehat{\alpha}_i + \widehat{\gamma}_t, \widehat{\alpha}_i + \widehat{\gamma}_{t-j})$ denote the fixed effects estimator, e.g., $\mathbb{E}_\phi\widehat{[\partial_{\pi^2}\ell_{it}]}$ denotes the fixed effects estimator of $\mathbb{E}_\phi[\partial_{\pi^2}\ell_{it}]$. Let $\widehat{\mathcal{H}}^{-1}_{(\alpha\alpha)}$, $\widehat{\mathcal{H}}^{-1}_{(\alpha\gamma)}$, $\widehat{\mathcal{H}}^{-1}_{(\gamma\alpha)}$, and $\widehat{\mathcal{H}}^{-1}_{(\gamma\gamma)}$ denote the blocks of the matrix $\widehat{\mathcal{H}}^{-1}$, where

$$\widehat{\mathcal{H}} = \begin{pmatrix} \widehat{\mathcal{H}}^*_{(\alpha\alpha)} & \widehat{\mathcal{H}}^*_{(\alpha\gamma)} \\ [\widehat{\mathcal{H}}^*_{(\alpha\gamma)}]' & \widehat{\mathcal{H}}^*_{(\gamma\gamma)} \end{pmatrix} + \frac{b}{\sqrt{NT}} vv',$$

$\widehat{\mathcal{H}}^*_{(\alpha\alpha)} = \mathrm{diag}(-\sum_t \mathbb{E}_\phi\widehat{[\partial_{\pi^2}\ell_{it}]})/\sqrt{NT}$, $\widehat{\mathcal{H}}^*_{(\alpha\alpha)} = \mathrm{diag}(-\sum_i \mathbb{E}_\phi\widehat{[\partial_{\pi^2}\ell_{it}]})/\sqrt{NT}$, and $\widehat{\mathcal{H}}^*_{(\alpha\gamma)it} = -\mathbb{E}_\phi\widehat{[\partial_{\pi^2}\ell_{it}]}/\sqrt{NT}$. Let

$$\widehat{\Xi}_{it} = -\frac{1}{\sqrt{NT}} \sum_{j=1}^{N} \sum_{\tau=1}^{T} \left( \widehat{\mathcal{H}}^{-1}_{(\alpha\alpha)ij} + \widehat{\mathcal{H}}^{-1}_{(\gamma\alpha)tj} + \widehat{\mathcal{H}}^{-1}_{(\alpha\gamma)i\tau} + \widehat{\mathcal{H}}^{-1}_{(\gamma\gamma)t\tau} \right) \mathbb{E}_\phi \widehat{(\partial_{\beta\pi}\ell_{j\tau})}.$$

The $k$-th component of $\widehat{\Xi}_{it}$ corresponds to a least squares regression of $\mathbb{E}_\phi\widehat{(\partial_{\beta_k\pi}\ell_{j\tau})}/\mathbb{E}_\phi\widehat{(\partial_{\pi^2}\ell_{it})}$ on the space spanned by the incidental parameters weighted by $\mathbb{E}_\phi\widehat{(-\partial_{\pi^2}\ell_{it})}$.

The analytical bias corrected estimator of $\beta^0$ is

$$\widetilde{\beta}^A = \widehat{\beta} - \widehat{B}/T - \widehat{D}/N,$$

where

$$\widehat{B} = -\frac{1}{N} \sum_{i=1}^{N} \frac{\sum_{j=0}^{L} [T/(T-j)] \sum_{t=j+1}^{T} \mathbb{E}_\phi \widehat{(\partial_\pi \ell_{i,t-j} D_{\beta\pi}\ell_{it})} + \frac{1}{2} \sum_{t=1}^{T} \mathbb{E}_\phi \widehat{(D_{\beta\pi^2}\ell_{it})}}{\sum_{t=1}^{T} \mathbb{E}_\phi \widehat{(\partial_{\pi^2}\ell_{it})}},$$

$$\widehat{D} = -\frac{1}{T} \sum_{t=1}^{T} \frac{\sum_{i=1}^{N} \left[ \mathbb{E}_\phi \widehat{(\partial_\pi \ell_{it} D_{\beta\pi}\ell_{it})} + \frac{1}{2}\mathbb{E}_\phi \widehat{(D_{\beta\pi^2}\ell_{it})} \right]}{\sum_{i=1}^{N} \mathbb{E}_\phi \widehat{(\partial_{\pi^2}\ell_{it})}},$$

and $L$ is a trimming parameter for estimation of spectral expectations such that $L \to \infty$ and $L/T \to 0$ (Hahn and Kuersteiner, 2011). The factor $T/(T-j)$ is a degrees of freedom adjustment that rescales the time series averages $T^{-1}\sum_{t=j+1}^{T}$ by the number of observations instead of by $T$. Unlike for variance estimation, we do not need to use a kernel function because the bias estimator does not need to be positive. Asymptotic $(1-p)$–confidence intervals for the components of $\beta^0$ can be formed as

$$\widetilde{\beta}_k^A \pm z_{1-p}\sqrt{\widehat{W}_{kk}^{-1}/(NT)}, \quad k = \{1, ..., \dim \beta^0\},$$

where $z_{1-p}$ is the $(1-p)$–quantile of the standard normal distribution, and $\widehat{W}_{kk}^{-1}$ is the $(k,k)$-element of the matrix $\widehat{W}^{-1}$ with

$$\widehat{W} = -(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbb{E}_\phi\,\widehat{(\partial_{\beta\beta'}\ell_{it})} - \mathbb{E}_\phi\,\widehat{(\partial_{\pi^2}\ell_{it}\Xi_{it}\Xi_{it}')}\right]. \tag{4.5}$$

The analytical bias corrected estimator of $\delta_{NT}^0$ is

$$\widetilde{\delta}^A = \widetilde{\delta} - \widehat{B}^\delta/T - \widehat{D}^\delta/N,$$

where $\widetilde{\delta}$ is the APE constructed from a bias corrected estimator of $\beta$. Let

$$\widehat{\Psi}_{it} = -\frac{1}{\sqrt{NT}}\sum_{j=1}^{N}\sum_{\tau=1}^{T}\left(\widehat{\mathcal{H}}_{(\alpha\alpha)ij}^{-1} + \widehat{\mathcal{H}}_{(\gamma\alpha)tj}^{-1} + \widehat{\mathcal{H}}_{(\alpha\gamma)i\tau}^{-1} + \widehat{\mathcal{H}}_{(\gamma\gamma)t\tau}^{-1}\right)\widehat{\partial_\pi\Delta_{j\tau}}.$$

The fixed effects estimators of the components of the asymptotic bias are

$$\begin{aligned}
\widehat{B}^\delta &= \frac{1}{N}\sum_{i=1}^{N}\frac{\sum_{j=0}^{L}[T/(T-j)]\sum_{t=j+1}^{T}\mathbb{E}_\phi\,\widehat{(\partial_\pi\ell_{i,t-j}\partial_{\pi^2}\ell_{it}\Psi_{it})}}{\sum_{t=1}^{T}\mathbb{E}_\phi\,\widehat{(\partial_{\pi^2}\ell_{it})}} \\
&\quad - \frac{1}{2\,N}\sum_{i=1}^{N}\frac{\sum_{t=1}^{T}\left[\mathbb{E}_\phi\widehat{(\partial_{\pi^2}\Delta_{it})} - \mathbb{E}_\phi\widehat{(\partial_{\pi^3}\ell_{it})}\mathbb{E}_\phi\widehat{(\Psi_{it})}\right]}{\sum_{t=1}^{T}\mathbb{E}_\phi\,\widehat{(\partial_{\pi^2}\ell_{it})}}, \\
\widehat{D}^\delta &= \frac{1}{T}\sum_{t=1}^{T}\frac{\sum_{i=1}^{N}\left[\mathbb{E}_\phi\,\widehat{(\partial_\pi\ell_{it}\partial_{\pi^2}\ell_{it}\Psi_{it})} - \frac{1}{2}\mathbb{E}_\phi\widehat{(\partial_{\pi^2}\Delta_{it})} + \frac{1}{2}\mathbb{E}_\phi\widehat{(\partial_{\pi^3}\ell_{it})}\mathbb{E}_\phi\widehat{(\Psi_{it})}\right]}{\sum_{i=1}^{N}\mathbb{E}_\phi\,\widehat{(\partial_{\pi^2}\ell_{it})}}.
\end{aligned}$$

The estimator of the asymptotic variance depends on the assumptions about the distribution of the unobserved effects and explanatory variables. Under Assumption 4.2(i)(a) we need to impose an homogeneity assumption on the distribution of the explanatory variables to estimate the first term of the asymptotic variance. For example, if $\{X_{it} : 1 \le i \le N, 1 \le t \le T\}$ is identically distributed over $i$, we can form

$$\widehat{V}^\delta = \frac{r_{NT}^2}{N^2T^2}\sum_{i=1}^{N}\left[\sum_{t,\tau=1}^{T}\widehat{\widetilde{\Delta}}_{it}\widehat{\widetilde{\Delta}}_{i\tau}' + \sum_{t=1}^{T}\mathbb{E}_\phi\widehat{(\Gamma_{it}\Gamma_{it}')}\right], \tag{4.6}$$

for $\widehat{\widetilde{\Delta}}_{it} = \widehat{\Delta}_{it} - N^{-1}\sum_{i=1}^{N}\widehat{\Delta}_{it}$. Under Assumption 4.2(i)(b) and the independence assumption on the unobserved effects of Remark 2,

$$\widehat{V}^\delta = \frac{r_{NT}^2}{N^2T^2}\sum_{i=1}^{N}\left[\sum_{t,\tau=1}^{T}\widehat{\widetilde{\Delta}}_{it}\widehat{\widetilde{\Delta}}_{i\tau}' + \sum_{t=1}^{T}\sum_{j\neq i}\widehat{\widetilde{\Delta}}_{it}\widehat{\widetilde{\Delta}}_{jt}' + \sum_{t=1}^{T}\mathbb{E}_\phi\widehat{(\Gamma_{it}\Gamma_{it}')}\right], \tag{4.7}$$

where $\widehat{\widetilde{\Delta}}_{it} = \widehat{\Delta}_{it} - N^{-1}\sum_{i=1}^{N}\widehat{\Delta}_{it}$ under identical distribution over $i$, $\widehat{\widetilde{\Delta}}_{it} = \widehat{\Delta}_{it} - T^{-1}\sum_{t=1}^{T}\widehat{\Delta}_{it}$ under stationarity over $t$, and $\widehat{\widetilde{\Delta}}_{it} = \widehat{\Delta}_{it} - \widehat{\delta}$ under both. Note that we do not need to specify the convergence rate to make inference because the standard errors $\sqrt{\widehat{V}^{\delta}}/r_{NT}$ do not depend on $r_{NT}$. Bias corrected estimators and confidence intervals can be constructed in the same fashion as for the model parameter.

We use the following homogeneity assumption to show the validity of the jackknife corrections for the model parameters and APEs. It ensures that the asymptotic bias is the same in all the partitions of the panel. The analytical corrections *do not* require this assumption.

**Assumption 4.3** (Unconditional homogeneity). *The sequence $\{(Y_{it}, X_{it}, \alpha_i, \gamma_t) : 1 \le i \le N, 1 \le t \le T\}$ is identically distributed across $i$ and strictly stationary across $t$, for each $N, T$.*

This assumption might seem restrictive for dynamic models where $X_{it}$ includes lags of the dependent variable because in this case it restricts the unconditional distribution of the initial conditions of $Y_{it}$. Note, however, that Assumption 4.3 allows the initial conditions to depend on the unobserved effects. In other words, it does not impose that the initial conditions are generated from the stationary distribution of $Y_{it}$ conditional on $X_{it}$ and $\phi$. Assumption 4.3 rules out structural breaks in the processes for the unobserved effects and observed variables. For APEs, it also imposes that these effects do not change with $T$ and $N$, i.e. $\delta^0_{NT} = \delta^0$.

The following theorems are the main result of this section. They show that the analytical and jackknife bias corrections eliminate the bias from the asymptotic distribution of the fixed effects estimators of the model parameters and APEs without increasing variance, and that the estimators of the asymptotic variances are consistent.

**Theorem 4.3** (Bias corrections for $\widehat{\beta}$). *Under the conditions of Theorems 4.1,*

$$\widehat{W} \to_P \overline{W}_{\infty},$$

*and, if $L \to \infty$ and $L/T \to 0$,*

$$\sqrt{NT}(\widetilde{\beta}^A - \beta^0) \to_d \mathcal{N}(0, \overline{W}_{\infty}^{-1}).$$

*Under the conditions of Theorems 4.1 and Assumption 4.3,*

$$\sqrt{NT}(\widetilde{\beta}^J - \beta^0) \to_d \mathcal{N}(0, \overline{W}_{\infty}^{-1}).$$

**Theorem 4.4** (Bias corrections for $\widehat{\delta}$). *Under the conditions of Theorems 4.1 and 4.2,*

$$\widehat{V}^{\delta} \to_P \overline{V}^{\delta}_{\infty},$$

*and, if $L \to \infty$ and $L/T \to 0$,*

$$r_{NT}(\widetilde{\delta}^A - \delta^0_{NT}) \to_d \mathcal{N}(0, \overline{V}^{\delta}_{\infty}).$$

*Under the conditions of Theorems 4.1 and 4.2, and Assumption 4.3,*

$$r_{NT}(\widetilde{\delta}^J - \delta^0) \to_d \mathcal{N}(0, \overline{V}^{\delta}_{\infty}).$$

**Remark 4** (Rate of convergence). *The rate of convergence $r_{NT}$ depends on the properties of the sampling process for the explanatory variables and unobserved effects (see remark 2).*

# 5 Monte Carlo Experiments

This section reports evidence on the finite sample behavior of fixed effects estimators of model parameters and APEs in static models with strictly exogenous regressors and dynamic models with predetermined regressors such as lags of the dependent variable. We analyze the performance of uncorrected and bias-corrected fixed effects estimators in terms of bias and inference accuracy of their asymptotic distribution. In particular we compute the biases, standard deviations, and root mean squared errors of the estimators, the ratio of average standard errors to the simulation standard deviations (SE/SD); and the empirical coverages of confidence intervals with 95% nominal value (p; .95).[9] Overall, we find that the analytically corrected estimators dominate the uncorrected and jackknife corrected estimators. All the results are based on 500 replications.

## 5.1 Example 1: binary response models

The designs correspond to static and dynamic probit models. We consider panels with a cross sectional size of 52 individuals, motivated by applications to U.S. states.

### 5.1.1 Static probit model

The data generating process is

$$Y_{it} = \mathbf{1}\left\{X_{it}\beta + \alpha_i + \gamma_t > \varepsilon_{it}\right\}, \quad (i = 1, ..., N;\ t = 1, ..., T),$$

where $\alpha_i \sim \mathcal{N}(0, 1/16)$, $\gamma_t \sim \mathcal{N}(0, 1/16)$, $\varepsilon_{it} \sim \mathcal{N}(0, 1)$, and $\beta = 1$. We consider two alternative designs for $X_{it}$: correlated and uncorrelated with the individual and time effects. In the first design, $X_{it} = X_{i,t-1}/2 + \alpha_i + \gamma_t + \upsilon_{it}$, $\upsilon_{it} \sim \mathcal{N}(0, 1/2)$, and $X_{i0} \sim \mathcal{N}(0, 1)$. In the second design, $X_{it} = X_{i,t-1}/2 + \upsilon_{it}$, $\upsilon_{it} \sim \mathcal{N}(0, 3/4)$, and $X_{i0} \sim \mathcal{N}(0, 1)$. In both designs $X_{it}$ is strictly exogenous with respect to $\varepsilon_{it}$ conditional to the individual and time effects, and has an unconditional variance equal to one. The variables $\alpha_i$, $\gamma_t$, $\varepsilon_{it}$, $\upsilon_{it}$, and $X_{i0}$ are independent and *i.i.d.* across individuals and time periods. We generate panel data sets with $N = 52$ individuals and three different numbers of time periods $T$: 14, 26 and 52.

Table 3 reports the results for the probit coefficient $\beta$, and the APE of $X_{it}$. We compute the APE using (2.4). Throughout the table, MLE-FETE corresponds to the probit maximum likelihood estimator with individual and time fixed effects, Analytical is the bias corrected estimator that uses the analytical correction, and Jackknife is the bias corrected estimator that uses SPJ in both the individual and time dimensions. The cross-sectional division in the jackknife follows the order of the observations. All the results are reported in percentage of the true parameter value.

We find that the bias is of the same order of magnitude as the standard deviation for the uncorrected estimator of the probit coefficient causing severe undercoverage of the confidence intervals. This result holds for both designs and all the sample sizes considered. The bias corrections, specially Analytical,

[9] The standard errors are computed using the expressions (4.5), (4.6) and (4.7) evaluated at uncorrected estimates of the parameters.

remove the bias without increasing dispersion, and produce substantial improvements in rmse and coverage probabilities. For example, Analytical reduces rmse by more than 40 % and increases coverage by 20% in the correlated design with $T = 14$. As in Hahn and Newey (2004) and Fernandez-Val (2009), we find very little bias in the uncorrected estimates of the APE, despite the large bias in the probit coefficients.

### 5.1.2 Dynamic probit model

The data generating process is

$$\begin{aligned} Y_{it} &= \mathbf{1}\{Y_{i,t-1}\beta_Y + Z_{it}\beta_Z + \alpha_i + \gamma_t > \varepsilon_{it}\}, \quad (i = 1, ..., N; \ t = 1, ..., T), \\ Y_{i0} &= \mathbf{1}\{Z_{i0}\beta_Z + \alpha_i + \gamma_0 > \varepsilon_{i0}\}, \end{aligned}$$

where $\alpha_i \sim \mathcal{N}(0, 1/16)$, $\gamma_t \sim \mathcal{N}(0, 1/16)$, $\varepsilon_{it} \sim \mathcal{N}(0, 1)$, $\beta_Y = 0.5$, and $\beta_Z = 1$. We consider two alternative designs for $Z_{it}$: correlated an uncorrelated with the individual and time effects. In the first design, $Z_{it} = Z_{i,t-1}/2 + \alpha_i + \gamma_t + \upsilon_{it}$, $\upsilon_{it} \sim \mathcal{N}(0, 1/2)$, and $Z_{i0} \sim \mathcal{N}(0, 1)$. In the second design, $Z_{it} = Z_{i,t-1}/2 + \upsilon_{it}$, $\upsilon_{it} \sim \mathcal{N}(0, 3/4)$, and $Z_{i0} \sim \mathcal{N}(0, 1)$. The unconditional variance of $Z_{it}$ is one in both designs. The variables $\alpha_i$, $\gamma_t$, $\varepsilon_{it}$, $\upsilon_{it}$, and $Z_{i0}$ are independent and *i.i.d.* across individuals and time periods. We generate panel data sets with $N = 52$ individuals and three different numbers of time periods $T$: 14, 26 and 52.

Table 4 reports the simulation results for the probit coefficient $\beta_Y$ and the APE of $Y_{i,t-1}$. We compute the partial effect of $Y_{i,t-1}$ using the expression in equation (2.3) with $X_{it,k} = Y_{i,t-1}$. This effect is commonly reported as a measure of state dependence for dynamic binary processes. Table 5 reports the simulation results for the estimators of the probit coefficient $\beta_Z$ and the APE of $Z_{it}$. We compute the partial effect using (2.4) with $X_{it,k} = Z_{it}$. Throughout the tables, we compare the same estimators as for the static model. For the analytical correction we consider two versions, Analytical (L=1) sets the trimming parameter to estimate spectral expectations $L$ to one, whereas Analytical (L=2) sets $L$ to two. Again, all the results in the tables are reported in percentage of the true parameter value.

The results in table 4 show important biases toward zero for *both* the probit coefficient and the APE of $Y_{i,t-1}$ in the two designs. This bias can indeed be substantially larger than the corresponding standard deviation for short panels yielding coverage probabilities below 70% for $T = 14$. The analytical corrections significantly reduce biases and rmse, bring coverage probabilities close to their nominal level, and have little sensitivity to the trimming parameter $L$. The jackknife corrections reduce bias but increase dispersion, producing less drastic improvements in rmse and coverage than the analytical corrections. The results for $Z_{it}$ in table 5 are similar to the static probit model. There are significant bias and undercoverage of confidence intervals for the coefficient, which are removed by the corrections, whereas there are little bias and undercoverage in the APEs.

## 5.2 Example 2: count response models

The designs correspond to static and dynamic Poisson models with additive individual and time effects. Motivated by the empirical example in next section, we calibrate all the parameters and exogenous variables using the dataset from Aghion, Bloom, Blundell, Griffith and Howitt (2005) (ABBGH). They estimated the relationship between competition and innovation using an an unbalanced panel dataset of 17 industries over the 22 years period 1973–1994. The dependent variable is number of patents.

### 5.2.1 Static Poisson model

The data generating process is

$$Y_{it} \mid Z_i^T, \alpha, \gamma \sim \mathcal{P}(\exp[Z_{it}\beta_1 + Z_{it}^2\beta_2 + \alpha_i + \gamma_t]), \quad (i = 1, ..., N;\ t = 1, ..., T),$$

where $\mathcal{P}$ denotes the Poisson distribution. The variable $Z_{it}$ is fixed to the values of the competition variable in the dataset and all the parameters are set to the fixed effect estimates of the model. We generate unbalanced panel data sets with $T = 22$ years and three different numbers of industries $N$: 17, 34, and 51. In the second (third) case, we double (triple) the cross-sectional size by merging two (three) independent realizations of the panel.

Table 6 reports the simulation results for the coefficients $\beta_1$ and $\beta_2$, and the APE of $Z_{it}$. We compute the APE using the expression (2.5) with $H(Z_{it}) = Z_{it}^2$. Throughout the table, MLE corresponds to the pooled Poisson maximum likelihood estimator (without individual and time effects), MLE-TE corresponds to the Poisson estimator with only time effects, MLE-FETE corresponds to the Poisson maximum likelihood estimator with individual and time fixed effects, Analytical (L=l) is the bias corrected estimator that uses the analytical correction with $L = l$, and Jackknife is the bias corrected estimator that uses SPJ in both the individual and time dimensions. The analytical corrections are different from the uncorrected estimator because they do not use that the regressor $Z_{it}$ is strictly exogenous. The cross-sectional division in the jackknife follows the order of the observations. The choice of these estimators is motivated by the empirical analysis of ABBGH. All the results in the table are reported in percentage of the true parameter value.

The results of the table agree with the no asymptotic bias result for the Poisson model with exogenous regressors. Thus, the bias of MLE-FETE for the coefficients and APE is negligible relative to the standard deviation and the coverage probabilities get close to the nominal level as $N$ grows. The analytical corrections preserve the performance of the estimators and have very little sensitivity to the trimming parameter. The jackknife correction increases dispersion and rmse, specially for the small cross-sectional size of the application. The estimators that do not control for individual effects are clearly biased.

### 5.2.2 Dynamic Poisson model

The data generating process is

$$Y_{it} \mid Y_i^{t-1}, Z_i^t, \alpha, \gamma \sim \mathcal{P}(\exp[\beta_Y \log(1 + Y_{i,t-1}) + Z_{it}\beta_1 + Z_{it}^2\beta_2 + \alpha_i + \gamma_t]), \quad (i = 1, ..., N; t = 1, ..., T).$$

The competition variable $Z_{it}$ and the initial condition for the number of patents $Y_{i0}$ are fixed to the values in the dataset and all the parameters are set to the fixed effect estimates of the model. To generate panels, we first impute values to the missing observations of $Z_{it}$ using forward and backward predictions from a panel AR(1) linear model with individual and time effects. We then draw panel data sets with $T = 21$ years and three different numbers of industries $N$: 17, 34, and 51. As in the static model, we double (triple) the cross-sectional size by merging two (three) independent realizations of the panel. We make the generated panels unbalanced by dropping the values corresponding to the missing observations in the original dataset.

Table 7 reports the simulation results for the coefficient $\beta_Y^0$ and the APE of $Y_{i,t-1}$. The estimators considered are the same as for the static Poisson model above. We compute the partial effect of $Y_{i,t-1}$ using (2.5) with $Z_{it} = Y_{i,t-1}$, $H(Z_{it}) = \log(1 + Z_{it})$, and dropping the linear term. Table 8 reports the simulation results for the coefficients $\beta_1^0$ and $\beta_2^0$, and the APE of $Z_{it}$. We compute the partial effect using (2.5) with $H(Z_{it}) = Z_{it}^2$. Again, all the results in the tables are reported in percentage of the true parameter value.

The results in table 7 show biases of the same order of magnitude as the standard deviation for the fixed effects estimators of the coefficient and APE of $Y_{i,t-1}$, which cause severe undercoverage of confidence intervals. Note that in this case the rate of convergence for the estimator of the APE is $r_{NT} = \sqrt{NT}$, because the individual and time effects are hold fixed across the simulations. The analytical corrections reduce bias by more than half without increasing dispersion, substantially reducing rmse and bringing coverage probabilities closer to their nominal levels. The jackknife corrections reduce bias and increase dispersion leading to lower improvements in rmse and coverage probability than the analytical corrections. The results for the coefficient of $Z_{it}$ in table 8 are similar to the static model. The results for the APE of $Z_{it}$ are imprecise, because the true value of the effect is close to zero.

# 6 Empirical Example

To illustrate the bias corrections with real data, we revisit the empirical application of Aghion, Bloom, Blundell, Griffith and Howitt (2005) that estimated a count data model to analyze the relationship between innovation and competition. They used an unbalanced panel of seventeen U.K. industries followed over the 22 years between 1973 and 1994. The dependent variable, $Y_{it}$, is innovation as measured by a citation-weighted number of patents, and the explanatory variable of interest, $Z_{it}$, is competition as measured by one minus the Lerner index in the industry-year. Following ABBGH we consider a quadratic static Poisson model with industry and year effects where

$$Y_{it} \mid Z_i^T, \alpha_i, \gamma_t \sim \mathcal{P}(\exp[\beta_1 Z_{it} + \beta_2 Z_{it}^2 + \alpha_i + \gamma_t]),$$

for $(i = 1, ..., 17; t = 1973, ..., 1994)$, and extend the analysis to a dynamic Poisson model with industry and year effects where

$$Y_{it} \mid Y_i^{t-1}, Z_i^t, \alpha_i, \gamma^t \sim \mathcal{P}(\exp[\beta_Y \log(1 + Y_{i,t-1}) + \beta_1 Z_{it} + \beta_2 Z_{it}^2 + \alpha_i + \gamma_t]),$$

for $(i = 1, ..., 17; t = 1974, ..., 1994)$. In the dynamic model we use the year 1973 as the initial condition for $Y_{it}$.

Table 9 reports the results of the analysis. Columns (2) and (3) for the static model replicate the empirical results of Table I in ABBGH (p. 708), adding estimates of the APEs. Columns (4) and (5) report estimates of the analytical corrections that do not assume that competition is strictly exogenous with $L = 1$ and $L = 2$, and column (6) reports estimates of the jackknife bias corrections described in

(3.4). Overall, the corrected estimates, while numerically different from the uncorrected estimates in column (3), agree with the inverted-U pattern in the relationship between innovation and competition found by ABBGH. The close similarity between the uncorrected and bias corrected estimates gives some evidence in favor of the strict exogeneity of competition with respect to the innovation process.

The results for the dynamic model show substantial positive state dependence on the innovation process that is not explained by industry heterogeneity. Uncorrected fixed effects underestimates the coefficient and APE of lag patents relative to the bias corrections, specially relative to the jackknife. The pattern of the differences between the estimates is consistent with the biases that we find in the numerical example in Table 7. Accounting for state dependence does not change the inverted-U pattern, but flattens the relationship between innovation and competition.

# 7 Concluding remarks

In this paper we develop analytical and jackknife corrections for fixed effects estimators of model parameters and APEs in semi parametric nonlinear panel models with additive individual and time effects. Our analysis applies to conditional maximum likelihood estimators with concave log-likelihood functions, and therefore covers logit, probit, ordered probit, ordered logit, Poisson, negative binomial, and Tobit estimators, which are the most popular nonlinear estimators in empirical economics.

We are currently developing similar corrections for nonlinear models with interactive individual and time effects (Chen, Fernández-Val, and Weidner (2013)). Another interesting avenue of future research is to derive higher-order expansions for fixed effects estimators with individual and time effects. These expansions are needed to justify theoretically the validity of alternative corrections based on the leave-one-observation-out panel jackknife method of Hahn and Newey (2004).

# References


Aghion, P., Bloom, N., Blundell, R., Griffith, R., and Howitt, P. (2005). Competition and innovation: an inverted-u relationship. *The Quarterly Journal of Economics*, 120(2):701.

Arellano, M. and Bonhomme, S. (2009). Robust priors in nonlinear panel data models. *Econometrica*, 77(2):489–536.

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.

Charbonneau, K. (2012). Multiple fixed effects in nonlinear panel data models. *Unpublished manuscript*.

Chen, M., Fernández-Val, I., and Weidner, M. (2013). Interactive fixed effects in nonlinear panel data models with large $n$, $t$.

Cox, D. D. and Kim, T. Y. (1995). Moment bounds for mixing random variables useful in nonparametric function estimation. *Stochastic processes and their applications*, 56(1):151–158.

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

Fan, J. and Yao, Q. (2003). Nonlinear time series: nonparametric and parametric methods.

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

Fernández-Val, I. and Vella, F. (2011). Bias corrections for two-step fixed effects panel data estimators. *Journal of Econometrics*.

Galvao, A. and Kato, K. (2013). Estimation and inference for linear panel data models under misspecification when both n and t are large.

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. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory*, 1(1):1–40.

Hahn, J. and Moon, H. (2006). Reducing bias of mle in a dynamic panel model. *Econometric Theory*, 22(03):499–512.

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.

Higham, N. J. (1992). Estimating the matrixp-norm. *Numerische Mathematik*, 62(1):539–555.

Horn, R. A. and Johnson, C. R. (1985). *Matrix analysis*. Cambridge university press.

Hu, L. (2002). Estimation of a censored dynamic panel data model. *Econometrica*, 70(6):2499–2517.

Kato, K., Galvao, A., and Montes-Rojas, G. (2012). Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics*, 170(1):76–91.

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.

McLeish, D. (1974). Dependent central limit theorems and invariance principles. *the Annals of Probability*, pages 620–628.

Moon, H. and Weidner, M. (2010). Dynamic Linear Panel Regression Models with Interactive Fixed Effects. *Manuscript*.

Moon, H. and Weidner, M. (2013). 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.

Nickell, S. J. (1981). Biases in dynamic models with fixed effects. *Econometrica*, 49(6):1417–26.

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.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, 74(4):967–1012.

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.

White, H. (2001). *Asymptotic theory for econometricians*. Academic press New York.

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

# Appendix

## A Notation and Choice of Norms

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 square $n \times n$ matrices $B$, $C$, we use $B > C$ (or $B \geq C$) to indicate that $B - C$ is positive (semi) definite. We write wpa1 for "with probability approaching one" and wrt for "with respect to". All the limits are taken as $N, T \to \infty$ jointly.

As in the main text, we usually suppress the dependence on $NT$ of all the sequences of functions and parameters to lighten the notation, e.g. we write $\mathcal{L}$ for $\mathcal{L}_{NT}$ and $\phi$ for $\phi_{NT}$. Let

$$\mathcal{S}(\beta, \phi) = \partial_\phi \mathcal{L}(\beta, \phi), \qquad \mathcal{H}(\beta, \phi) = -\partial_{\phi\phi'} \mathcal{L}(\beta, \phi),$$

where $\partial_x f$ denotes the partial derivative of $f$ with respect to $x$, and additional subscripts denote higher-order partial derivatives. We refer to the $\dim \phi$-vector $\mathcal{S}(\beta, \phi)$ as the incidental parameter score, and to the $\dim \phi \times \dim \phi$ matrix $\mathcal{H}(\beta, \phi)$ as the incidental parameter Hessian. We omit the arguments 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 expectations conditional on $\phi$, e.g. $\partial_\beta \overline{\mathcal{L}} = \mathbb{E}_\phi[\partial_\beta \mathcal{L}]$, and a tilde to denote variables in deviations with respect to expectations, e.g. $\partial_\beta \widetilde{\mathcal{L}} = \partial_\beta \mathcal{L} - \partial_\beta \overline{\mathcal{L}}$.

We use the Euclidian norm $\|.\|$ for vectors of dimension $\dim \beta$, and we use the norm induced by the Euclidian norm for the corresponding matrices and tensors, which we also denote by $\|.\|$. For matrices of dimension $\dim \beta \times \dim \beta$ this induced norm is the spectral norm. The generalization of the spectral norm to higher order tensors is straightforward, e.g. the induced norm of the $\dim \beta \times \dim \beta \times \dim \beta$ tensor of third partial derivatives of $\mathcal{L}(\beta, \phi)$ wrt $\beta$ is given by

$$\|\partial_{\beta\beta\beta} \mathcal{L}(\beta, \phi)\| = \max_{\{u, v \in \mathbb{R}^{\dim \beta},\, \|u\|=1,\, \|v\|=1\}} \left\| \sum_{k,l=1}^{\dim \beta} u_k \, v_l \, \partial_{\beta \beta_k \beta_l} \mathcal{L}(\beta, \phi) \right\|.$$

This choice of norm is immaterial for the asymptotic analysis because $\dim \beta$ is fixed with the sample size.

In contrast, it is important what norms we choose for vectors of dimension $\dim \phi$, and their corresponding matrices and tensors, because $\dim \phi$ is increasing with the sample size. For vectors of dimension $\dim \phi$, we use the $\ell_q$-norm

$$\|\phi\|_q = \left( \sum_{g=1}^{\dim \phi} |\phi_g|^q \right)^{1/q},$$

where $2 \leq q \leq \infty$.[10] The particular value $q = 8$ will be chosen later.[11] We use the norms that are induced by the $\ell_q$-norm for the corresponding matrices and tensors, e.g. the induced $q$-norm of the

[10] We use the letter $q$ instead of $p$ to avoid confusion with the use of $p$ for probability.

[11] The main reason not to choose $q = \infty$ is the assumption $\|\widetilde{\mathcal{H}}\|_q = o_P(1)$ below, which is used to guarantee that $\|\mathcal{H}^{-1}\|_q$ is of the same order as $\|\overline{\mathcal{H}}^{-1}\|_q$. If we assume $\|\mathcal{H}^{-1}\|_q = \mathcal{O}_P(1)$ directly instead of $\|\overline{\mathcal{H}}^{-1}\|_q = \mathcal{O}_P(1)$, then we can set $q = \infty$.

$\dim \phi \times \dim \phi \times \dim \phi$ tensor of third partial derivatives of $\mathcal{L}(\beta, \phi)$ wrt $\phi$ is

$$\|\partial_{\phi\phi\phi}\mathcal{L}(\beta,\phi)\|_q = \max_{\{u,v\in\mathbb{R}^{\dim\phi},\, \|u\|_q=1,\, \|v\|_q=1\}} \left\| \sum_{g,h=1}^{\dim\phi} u_g\, v_h\, \partial_{\phi\phi_g\phi_h}\mathcal{L}(\beta,\phi) \right\|_q . \tag{A.1}$$

Note that in general the ordering of the indices of the tensor would matter in the definition of this norm, with the first index having a special role. However, since partial derivatives like $\partial_{\phi_g\phi_h\phi_l}\mathcal{L}(\beta,\phi)$ are fully symmetric in the indices $g$, $h$, $l$, the ordering is not important in their case.

For mixed partial derivatives of $\mathcal{L}(\beta,\phi)$ wrt $\beta$ and $\phi$, we use the norm that is induced by the Euclidian norm on $\dim\beta$-vectors and the $q$-norm on $\dim\phi$-indices, e.g.

$$\begin{aligned}\|\partial_{\beta\beta\phi\phi\phi}\mathcal{L}(\beta,\phi)\|_q = &\max_{\{u,v\in\mathbb{R}^{\dim\beta},\, \|u\|=1,\, \|v\|=1\}} \max_{\{w,x\in\mathbb{R}^{\dim\phi},\, \|w\|_q=1,\, \|x\|_q=1\}} \\ &\left\| \sum_{k,l=1}^{\dim\beta}\sum_{g,h=1}^{\dim\phi} u_k\, v_l\, w_g\, x_h\, \partial_{\beta_k\beta_l\phi\phi_g\phi_h}\mathcal{L}(\beta,\phi) \right\|_q ,\end{aligned} \tag{A.2}$$

where we continue to use the notation $\|.\|_q$, even though this is a mixed norm.

Note that for $w, x \in \mathbb{R}^{\dim\phi}$ and $q \geq 2$,

$$|w'x| \leq \|w\|_q \|x\|_{q/(q-1)} \leq (\dim\phi)^{(q-2)/q}\|w\|_q\|x\|_q.$$

Thus, whenever we bound a scalar product of vectors, matrices and tensors in terms of the above norms we have to account for this additional factor $(\dim\phi)^{(q-2)/q}$. For example,

$$\left| \sum_{k,l=1}^{\dim\beta}\sum_{f,g,h=1}^{\dim\phi} u_k\, v_l\, w_f\, x_h\, y_f\, \partial_{\beta_k\beta_l\phi_f\phi_g\phi_h}\mathcal{L}(\beta,\phi) \right| \leq (\dim\phi)^{(q-2)/q}\|u\|\,\|v\|\,\|w\|_q\,\|x\|_q\,\|y\|_q\,\|\partial_{\beta\beta\phi\phi\phi}\mathcal{L}(\beta,\phi)\|_q .$$

For higher-order tensors, we use the notation $\partial_{\phi\phi\phi}\mathcal{L}(\beta,\phi)$ inside the $q$-norm $\|.\|_q$ defined above, while we rely on standard index and matrix notation for all other expressions involving those partial derivatives, e.g. $\partial_{\phi\phi'\phi_g}\mathcal{L}(\beta,\phi)$ is a $\dim\phi \times \dim\phi$ matrix for every $g = 1,\ldots,\dim\phi$. Occasionally, e.g. in Assumption B.1($vi$) below, we use the Euclidian norm for $\dim\phi$-vectors, and the spectral norm for $\dim\phi \times \dim\phi$-matrices, denoted by $\|.\|$, and defined as $\|.\|_q$ with $q = 2$. Moreover, we employ the matrix infinity norm $\|A\|_\infty = \max_i \sum_j |A_{ij}|$, and the matrix maximum norm $\|A\|_{\max} = \max_{ij}|A_{ij}|$ to characterize the properties of the inverse of the expected Hessian of the incidental parameters in Section D.4.

For $r \geq 0$, we define the sets $\mathcal{B}(r,\beta^0) = \left\{\beta : \|\beta-\beta^0\| \leq r\right\}$, and $\mathcal{B}_q(r,\phi^0) = \left\{\phi : \|\phi-\phi^0\|_q \leq r\right\}$, which are closed balls of radius $r$ around the true parameter values $\beta^0$ and $\phi^0$, respectively.

# B Asymptotic Expansions

In this section, we derive asymptotic expansions for the score of the profile objective function, $\mathcal{L}(\beta,\widehat{\phi}(\beta))$, and for the fixed effects estimators of the parameters and APEs, $\widehat{\beta}$ and $\widehat{\delta}$. We do not employ the panel structure of the model, nor the particular form of the objective function given in Section 4. Instead, we consider the estimation of an unspecified model based on a sample of size $NT$ and a generic objective function $\mathcal{L}(\beta,\phi)$, which depends on the parameter of interest $\beta$ and the incidental parameter $\phi$. The estimators $\widehat{\phi}(\beta)$ and $\widehat{\beta}$ are defined in (2.6) and (2.7).

We make the following high-level assumptions. These assumptions might appear somewhat abstract, but will be justified by more primitive conditions in the context of panel models.

**Assumption B.1** (Regularity conditions for asymptotic expansion of $\widehat{\beta}$)**.** *Let $q > 4$ and $0 \leq \epsilon < 1/8 - 1/(2q)$. Let $r_\beta = r_{\beta,NT} > 0$, $r_\phi = r_{\phi,NT} > 0$, with $r_\beta = o\left[(NT)^{-1/(2q)-\epsilon}\right]$ and $r_\phi = o\left[(NT)^{-\epsilon}\right]$. We assume that*

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

*(ii)* $(\beta,\phi) \mapsto \mathcal{L}(\beta,\phi)$ *is four times continuously differentiable in* $\mathcal{B}(r_\beta,\beta^0) \times \mathcal{B}_q(r_\phi,\phi^0)$, *wpa1.*

*(iii)* $\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \left\|\widehat{\phi}(\beta) - \phi^0\right\|_q = o_P(r_\phi)$.

*(iv)* $\overline{\mathcal{H}} > 0$, *and* $\left\|\overline{\mathcal{H}}^{-1}\right\|_q = \mathcal{O}_P(1)$.

*(v) For the q-norm defined in Appendix A,*

$$\|\mathcal{S}\|_q = \mathcal{O}_P\left((NT)^{-1/4+1/(2q)}\right), \qquad \|\partial_\beta \mathcal{L}\| = \mathcal{O}_P(1), \qquad \|\widetilde{\mathcal{H}}\|_q = o_P(1),$$
$$\|\partial_{\beta\phi'}\mathcal{L}\|_q = \mathcal{O}_P\left((NT)^{1/(2q)}\right), \qquad \|\partial_{\beta\beta'}\mathcal{L}\| = \mathcal{O}_P(\sqrt{NT}), \qquad \|\partial_{\beta\phi\phi}\mathcal{L}\|_q = \mathcal{O}_P((NT)^{\epsilon}),$$
$$\|\partial_{\phi\phi\phi}\mathcal{L}\|_q = \mathcal{O}_P\left((NT)^{\epsilon}\right),$$

*and*

$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\beta\beta}\mathcal{L}(\beta,\phi)\| = \mathcal{O}_P\left(\sqrt{NT}\right),$$
$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\beta\phi}\mathcal{L}(\beta,\phi)\|_q = \mathcal{O}_P\left((NT)^{1/(2q)}\right),$$
$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\beta\phi\phi}\mathcal{L}(\beta,\phi)\|_q = \mathcal{O}_P\left((NT)^{\epsilon}\right),$$
$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\phi\phi\phi}\mathcal{L}(\beta,\phi)\|_q = \mathcal{O}_P\left((NT)^{\epsilon}\right),$$
$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\phi\phi\phi\phi}\mathcal{L}(\beta,\phi)\|_q = \mathcal{O}_P\left((NT)^{\epsilon}\right).$$

*(vi) For the spectral norm* $\|.\| = \|.\|_2$,

$$\|\widetilde{\mathcal{H}}\| = o_P\left((NT)^{-1/8}\right), \quad \left\|\partial_{\beta\beta'}\widetilde{\mathcal{L}}\right\| = o_P(\sqrt{NT}), \quad \left\|\partial_{\beta\phi\phi}\widetilde{\mathcal{L}}\right\| = o_P\left((NT)^{-1/8}\right),$$
$$\left\|\partial_{\beta\phi'}\widetilde{\mathcal{L}}\right\| = \mathcal{O}_P(1), \quad \left\|\sum_{g,h=1}^{\dim\phi} \partial_{\phi\phi_g\phi_h}\widetilde{\mathcal{L}}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g[\overline{\mathcal{H}}^{-1}\mathcal{S}]_h\right\| = o_P\left((NT)^{-1/4}\right).$$

Let $\partial_\beta\mathcal{L}(\beta,\widehat{\phi}(\beta))$ be the score of the profile objective function.[12] The following theorem is the main result of this appendix.

[12]Note that $\frac{d}{d\beta}\mathcal{L}(\beta,\widehat{\phi}(\beta)) = \partial_\beta\mathcal{L}(\beta,\widehat{\phi}(\beta))$ by the envelope theorem.

**Theorem B.1** (Asymptotic expansions of $\widehat{\phi}(\beta)$ and $\partial_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta))$)**.** *Let Assumption B.1 hold. Then*

$$\widehat{\phi}(\beta) - \phi^0 = \mathcal{H}^{-1}\mathcal{S} + \mathcal{H}^{-1}[\partial_{\phi\beta'}\mathcal{L}](\beta-\beta^0) + \tfrac{1}{2}\mathcal{H}^{-1}\sum_{g=1}^{\dim\phi}[\partial_{\phi\phi'\phi_g}\mathcal{L}]\mathcal{H}^{-1}\mathcal{S}[\mathcal{H}^{-1}\mathcal{S}]_g + R^{\phi}(\beta),$$

*and*

$$\partial_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta)) = U - \overline{W}\,\sqrt{NT}(\beta-\beta^0) + R(\beta),$$

*where* $U = U^{(0)} + U^{(1)}$*, and*

$$\overline{W} = -\frac{1}{\sqrt{NT}}\left(\partial_{\beta\beta'}\overline{\mathcal{L}} + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}\,[\partial_{\phi\beta'}\overline{\mathcal{L}}]\right),$$

$$U^{(0)} = \partial_\beta\mathcal{L} + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}\mathcal{S},$$

$$U^{(1)} = [\partial_{\beta\phi'}\widetilde{\mathcal{L}}]\overline{\mathcal{H}}^{-1}\mathcal{S} - [\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}\,\widetilde{\mathcal{H}}\,\overline{\mathcal{H}}^{-1}\,\mathcal{S} + \frac{1}{2}\sum_{g=1}^{\dim\phi}\left(\partial_{\beta\phi'\phi_g}\overline{\mathcal{L}} + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}[\partial_{\phi\phi'\phi_g}\overline{\mathcal{L}}]\right)[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g\overline{\mathcal{H}}^{-1}\mathcal{S}.$$

*The remainder terms of the expansions satisfy*

$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\frac{(NT)^{1/2-1/(2q)}\,\left\|R^{\phi}(\beta)\right\|_q}{1+\sqrt{NT}\|\beta-\beta^0\|} = o_P(1)\ , \qquad \sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\frac{\|R(\beta)\|}{1+\sqrt{NT}\|\beta-\beta^0\|} = o_P(1)\ .$$

**Remark 5.** *The result for* $\widehat{\phi}(\beta)-\phi^0$ *does not rely on Assumption B.1(vi). Without this assumption we can also show that*

$$\begin{aligned}\partial_\beta\mathcal{L}(\beta,\widehat{\phi}(\beta)) &= \partial_\beta\mathcal{L} + \left[\partial_{\beta\beta'}\mathcal{L} + (\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi'\beta}\mathcal{L})\right](\beta-\beta^0) + (\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1}\mathcal{S}\\ &\quad + \frac{1}{2}\sum_g\left(\partial_{\beta\phi'\phi_g}\mathcal{L} + [\partial_{\beta\phi'}\mathcal{L}]\,\mathcal{H}^{-1}[\partial_{\phi\phi'\phi_g}\mathcal{L}]\right)[\mathcal{H}^{-1}\mathcal{S}]_g\mathcal{H}^{-1}\mathcal{S} + R_1(\beta),\end{aligned}$$

*with* $R_1(\beta)$ *satisfying the same bound as* $R(\beta)$*. Thus, the spectral norm bounds in Assumption B.1(vi) for* $\dim\phi$*-vectors, matrices and tensors are only used after separating expectations from deviations of expectations for certain partial derivatives. Otherwise, the derivation of the bounds is purely based on the q-norm for* $\dim\phi$*-vectors, matrices and tensors.*

The proofs are given in Section B.1. Theorem B.1 characterizes asymptotic expansions for the incidental parameter estimator and the score of the profile objective function in the incidental parameter score $\mathcal{S}$ up to quadratic order. The theorem provides bounds on the the remainder terms $R^{\phi}(\beta)$ and $R(\beta)$, which make the expansions applicable to estimators of $\beta$ that take values within a shrinking $r_\beta$-neighborhood of $\beta^0$ wpa1. Given such an $r_\beta$-consistent estimator $\widehat{\beta}$ that solves the first order condition $\partial_\beta\mathcal{L}(\beta,\widehat{\phi}(\beta)) = 0$, we can use the expansion of the profile objective score to obtain an asymptotic expansion for $\widehat{\beta}$. This gives rise to the following corollary of Theorem B.1 . Let $\overline{W}_\infty := \lim_{N,T\to\infty}\overline{W}$.

**Corollary B.2** (Asymptotic expansion of $\widehat{\beta}$)**.** *Let Assumption B.1 be satisfied. In addition, let* $U = \mathcal{O}_P(1)$*, let* $\overline{W}_\infty$ *exist with* $\overline{W}_\infty > 0$*, and let* $\|\widehat{\beta}-\beta^0\| = o_P(r_\beta)$*. Then*

$$\sqrt{NT}(\widehat{\beta}-\beta^0) = \overline{W}_\infty^{-1}U + o_P(1).$$

The following theorem states that for strictly concave objective functions no separate consistency proof is required for $\widehat{\phi}(\beta)$ and for $\widehat{\beta}$.

**Theorem B.3** (Consistency under Concavity)**.** *Let Assumption B.1(i), (ii), (iv), (v) and (vi) hold, and let $(\beta, \phi) \mapsto \mathcal{L}(\beta, \phi)$ be strictly concave over $(\beta, \phi) \in \mathbb{R}^{\dim \beta + \dim \phi}$, wpa1. Assume furthermore that $(NT)^{-1/4+1/(2q)} = o_P(r_\phi)$ and $(NT)^{1/(2q)} r_\beta = o_P(r_\phi)$. Then,*

$$\sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \left\| \widehat{\phi}(\beta) - \phi^0 \right\|_q = o_P(r_\phi),$$

*i.e. Assumption B.1(iii) is satisfied. If, in addition, $\overline{W}_\infty$ exists with $\overline{W}_\infty > 0$, then $\|\widehat{\beta} - \beta^0\| = \mathcal{O}_P\left((NT)^{-1/4}\right)$.*

In the application of Theorem B.1 to panel models, we focus on estimators with strictly concave objective functions. By Theorem B.3, we only need to check Assumption B.1(*i*), (*ii*), (*iv*), (*v*) and (*vi*), as well as $U = \mathcal{O}_P(1)$ and $\overline{W}_\infty > 0$, when we apply Corollary B.2 to derive the limiting distribution of $\widehat{\beta}$. We give the proofs of Corollary B.2 and Theorem B.3 in Section B.1.

### Expansion for Average Effects

We invoke the following high-level assumption, which is verified under more primitive conditions for panel data models in the next section.

**Assumption B.2** (Regularity conditions for asymptotic expansion of $\widehat{\delta}$)**.** *Let $q$, $\epsilon$, $r_\beta$ and $r_\phi$ be defined as in Assumption B.1. We assume that*

*(i)* $(\beta, \phi) \mapsto \Delta(\beta, \phi)$ *is three times continuously differentiable in* $\mathcal{B}(r_\beta, \beta^0) \times \mathcal{B}_q(r_\phi, \phi^0)$, *wpa1.*

*(ii)* $\|\partial_\beta \Delta\| = \mathcal{O}_P(1)$, $\|\partial_\phi \Delta\|_q = \mathcal{O}_P\left((NT)^{1/(2q)-1/2}\right)$, $\|\partial_{\phi\phi} \Delta\|_q = \mathcal{O}_P((NT)^{\epsilon-1/2})$, *and*

$$\sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \|\partial_{\beta\beta} \Delta(\beta, \phi)\| = \mathcal{O}_P(1),$$
$$\sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \|\partial_{\beta\phi'} \Delta(\beta, \phi)\|_q = \mathcal{O}_P\left((NT)^{1/(2q)-1/2}\right),$$
$$\sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \|\partial_{\phi\phi\phi} \Delta(\beta, \phi)\|_q = \mathcal{O}_P\left((NT)^{\epsilon-1/2}\right).$$

*(iii)* $\left\|\partial_\beta \widetilde{\Delta}\right\| = o_P(1)$, $\left\|\partial_\phi \widetilde{\Delta}\right\| = \mathcal{O}_P\left((NT)^{-1/2}\right)$, *and* $\left\|\partial_{\phi\phi} \widetilde{\Delta}\right\| = o_P\left((NT)^{-5/8}\right)$.

The following result gives the asymptotic expansion for the estimator, $\widehat{\delta} = \Delta(\beta, \widehat{\phi}(\beta))$, wrt $\delta := \Delta(\beta^0, \phi^0)$.

**Theorem B.4** (Asymptotic expansion of $\hat{\delta}$)**.** *Let Assumptions B.1 and B.2 hold and let $\|\widehat{\beta} - \beta^0\| = \mathcal{O}_P\left((NT)^{-1/2}\right) = o_P(r_\beta)$. Then*

$$\widehat{\delta} - \delta = \left[\partial_{\beta'} \overline{\Delta} + (\partial_{\phi'} \overline{\Delta}) \overline{\mathcal{H}}^{-1} (\partial_{\phi\beta'} \overline{\mathcal{L}})\right] (\widehat{\beta} - \beta^0) + U_\Delta^{(0)} + U_\Delta^{(1)} + o_P\left(1/\sqrt{NT}\right),$$

*where*

$$U_{\Delta}^{(0)} = (\partial_{\phi'}\overline{\Delta})\overline{\mathcal{H}}^{-1}\mathcal{S},$$
$$U_{\Delta}^{(1)} = (\partial_{\phi'}\widetilde{\Delta})\overline{\mathcal{H}}^{-1}\mathcal{S} - (\partial_{\phi'}\overline{\Delta})\overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\mathcal{S} + \tfrac{1}{2}\mathcal{S}'\overline{\mathcal{H}}^{-1}\left[\partial_{\phi\phi'}\overline{\Delta} + \sum_{g=1}^{\dim\phi}\left[\partial_{\phi\phi'\phi_g}\overline{\mathcal{L}}\right]\left[\overline{\mathcal{H}}^{-1}(\partial_{\phi}\overline{\Delta})\right]_g\right]\overline{\mathcal{H}}^{-1}\mathcal{S}.$$

**Remark 6.** *The expansion of the profile score $\partial_{\beta_k}\mathcal{L}(\beta,\widehat{\phi}(\beta))$ in Theorem B.1 is a special case of the expansion in Theorem B.4, for $\Delta(\beta,\phi) = \frac{1}{\sqrt{NT}}\partial_{\beta_k}\mathcal{L}(\beta,\phi)$. Assumptions B.2 also exactly match with the corresponding subset of Assumption B.1.*

## B.1 Proofs for Appendix B (Asymptotic Expansions)

The following Lemma contains some statements that are not explicitly assumed in Assumptions B.1, but that are implied by it.

**Lemma B.5.** *Let Assumptions B.1 be satisfied. Then*

*(i)* $\mathcal{H}(\beta,\phi) > 0$ *for all* $\beta \in \mathcal{B}(r_\beta,\beta^0)$ *and* $\phi \in \mathcal{B}_q(r_\phi,\phi^0)$ *wpa1,*

$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\ \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\beta'}\mathcal{L}(\beta,\,\phi)\| = \mathcal{O}_P\left(\sqrt{NT}\right),$$
$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\ \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\phi'}\mathcal{L}(\beta,\,\phi)\|_q = \mathcal{O}_P\left((NT)^{1/(2q)}\right),$$
$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\ \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\phi\phi\phi}\mathcal{L}(\beta,\,\phi)\|_q = \mathcal{O}_P\left((NT)^{\epsilon}\right),$$
$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\ \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\phi\phi}\mathcal{L}(\beta,\phi)\|_q = \mathcal{O}_P((NT)^{\epsilon}),$$
$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\ \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \left\|\mathcal{H}^{-1}(\beta,\phi)\right\|_q = \mathcal{O}_P(1).$$

*(ii) Moreover,* $\|\mathcal{S}\| = \mathcal{O}_P(1)$, $\left\|\mathcal{H}^{-1}\right\| = \mathcal{O}_P(1)$, $\left\|\overline{\mathcal{H}}^{-1}\right\| = \mathcal{O}_P(1)$, $\left\|\mathcal{H}^{-1} - \overline{\mathcal{H}}^{-1}\right\| = o_P\left((NT)^{-1/8}\right)$, $\left\|\mathcal{H}^{-1} - \left(\overline{\mathcal{H}}^{-1} - \overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\right)\right\| = o_P\left((NT)^{-1/4}\right)$, $\|\partial_{\beta\phi'}\mathcal{L}\| = \mathcal{O}_P\left((NT)^{1/4}\right)$, $\|\partial_{\beta\phi\phi}\mathcal{L}\| = \mathcal{O}_P\left((NT)^{\epsilon}\right)$, $\left\|\sum_g \partial_{\phi\phi'\phi_g}\mathcal{L}\,[\mathcal{H}^{-1}\mathcal{S}]_g\right\| = \mathcal{O}_P\left((NT)^{-1/4+1/(2q)+\epsilon}\right)$, *and* $\left\|\sum_g \partial_{\phi\phi'\phi_g}\mathcal{L}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g\right\| = \mathcal{O}_P\left((NT)^{-1/4+1/(2q)+\epsilon}\right)$.

**Proof of Lemma B.5.** # Part $(i)$: Let $v \in \mathbb{R}^{\dim\beta}$ and $w, u \in \mathbb{R}^{\dim\phi}$. By a Taylor expansion of $\partial_{\beta\phi'\phi_g}\mathcal{L}(\beta,\,\phi)$ around $(\beta^0,\phi^0)$

$$\sum_g u_g v'\left[\partial_{\beta\phi'\phi_g}\mathcal{L}(\beta,\,\phi)\right]w = \sum_g u_g v'\left[\partial_{\beta\phi'\phi_g}\mathcal{L} + \sum_k(\beta_k-\beta_k^0)\partial_{\beta_k\beta\phi'\phi_g}\mathcal{L}(\tilde\beta,\tilde\phi) - \sum_h(\phi_h-\phi_h^0)\partial_{\beta\phi'\phi_g\phi_h}\mathcal{L}(\tilde\beta,\tilde\phi)\right]w,$$

with $(\tilde\beta,\tilde\phi)$ between $(\beta^0,\phi^0)$ and $(\beta,\phi)$. Thus

$$\|\partial_{\beta\phi\phi}\mathcal{L}(\beta,\,\phi)\|_q = \sup_{\|v\|=1}\ \sup_{\|u\|_q=1}\ \sup_{\|w\|_{q/(q-1)}=1}\ \sum_g u_g v'\left[\partial_{\beta\phi'\phi_g}\mathcal{L}(\beta,\,\phi)\right]w$$
$$\leq \|\partial_{\beta\phi\phi}\mathcal{L}\|_q + \|\beta-\beta^0\|\sup_{(\tilde\beta,\tilde\phi)}\left\|\partial_{\beta\beta\phi\phi}\mathcal{L}(\tilde\beta,\tilde\phi)\right\|_q + \|\phi-\phi^0\|_q\sup_{(\tilde\beta,\tilde\phi)}\left\|\partial_{\beta\phi\phi\phi}\mathcal{L}(\tilde\beta,\tilde\phi)\right\|_q,$$

where the supremum over $(\tilde{\beta}, \tilde{\phi})$ is necessary, because those parameters depend on $v$, $w$, $u$. By Assumption B.1, for large enough $N$ and $T$,

$$\begin{aligned}
\sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \|\partial_{\beta\phi\phi}\mathcal{L}(\beta,\, \phi)\|_q &\leq \|\partial_{\beta\phi\phi}\mathcal{L}\| + r_\beta \sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \|\partial_{\beta\beta\phi\phi}\mathcal{L}(\beta, \phi)\|_q \\
&\quad + r_\phi \sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \|\partial_{\beta\phi\phi\phi}\mathcal{L}(\beta, \phi)\|_q \\
&= \mathcal{O}_P\left[(NT)^\epsilon + r_\beta (NT)^\epsilon + r_\phi (NT)^\epsilon\right] = \mathcal{O}_P\left((NT)^\epsilon\right).
\end{aligned}$$

The proofs for the bounds on $\|\partial_{\beta\beta'}\mathcal{L}(\beta,\, \phi)\|$, $\|\partial_{\beta\phi'}\mathcal{L}(\beta,\, \phi)\|_q$ and $\|\partial_{\phi\phi\phi}\mathcal{L}(\beta,\, \phi)\|_q$ are analogous.

Next, we show that $\mathcal{H}(\beta, \phi)$ is non-singular for all $\beta \in \mathcal{B}(r_\beta, \beta^0)$ and $\phi \in \mathcal{B}_q(r_\phi, \phi^0)$ wpa1. By a Taylor expansion and Assumption B.1, for large enough $N$ and $T$,

$$\begin{aligned}
\sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \|\mathcal{H}(\beta, \phi) - \mathcal{H}\|_q &\leq r_\beta \sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \|\partial_{\beta\phi\phi}\mathcal{L}(\beta, \phi)\|_q \\
&\quad + r_\phi \sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \|\partial_{\phi\phi\phi}\mathcal{L}(\beta, \phi)\|_q = o_P(1). \qquad \text{(B.1)}
\end{aligned}$$

Define $\Delta\mathcal{H}(\beta, \phi) = \overline{\mathcal{H}} - \mathcal{H}(\beta, \phi)$. Then $\|\Delta\mathcal{H}(\beta, \phi)\|_q \leq \|\mathcal{H}(\beta, \phi) - \mathcal{H}\|_q + \left\|\widetilde{\mathcal{H}}\right\|_q$, and therefore

$$\sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \|\Delta\mathcal{H}(\beta, \phi)\|_q = o_P(1),$$

by Assumption B.1 and equation (B.1).

For any square matrix with $\|A\|_q < 1$, $\left\|(\mathbb{1} - A)^{-1}\right\|_q \leq (1 - \|A\|_q)^{-1}$, see e.g. p.301 in Horn and Johnson (1985). Then

$$\begin{aligned}
\sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \left\|\mathcal{H}^{-1}(\beta, \phi)\right\|_q &= \sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \left\|\left(\overline{\mathcal{H}} - \Delta\mathcal{H}(\beta, \phi)\right)^{-1}\right\|_q \\
&= \sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \left\|\overline{\mathcal{H}}^{-1}\left(\mathbb{1} - \Delta\mathcal{H}(\beta, \phi)\overline{\mathcal{H}}^{-1}\right)^{-1}\right\|_q \\
&\leq \left\|\overline{\mathcal{H}}^{-1}\right\|_q \sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \left\|\left(\mathbb{1} - \Delta\mathcal{H}(\beta, \phi)\overline{\mathcal{H}}^{-1}\right)^{-1}\right\|_q \\
&\leq \left\|\overline{\mathcal{H}}^{-1}\right\|_q \sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \left(1 - \left\|\Delta\mathcal{H}(\beta, \phi)\overline{\mathcal{H}}^{-1}\right\|_q\right)^{-1} \\
&\leq \left\|\overline{\mathcal{H}}^{-1}\right\|_q (1 - o_P(1))^{-1} = \mathcal{O}_P(1).
\end{aligned}$$

#Part $(ii)$: By the properties of the $\ell_q$-norm and Assumption B.1$(v)$,

$$\|\mathcal{S}\| = \|\mathcal{S}\|_2 \leq (\dim\phi)^{1/2 - 1/q}\|\mathcal{S}\|_q = \mathcal{O}_p(1).$$

Analogously,

$$\|\partial_{\beta\phi'}\mathcal{L}\| \leq (\dim\phi)^{1/2-1/q}\, \|\partial_{\beta\phi'}\mathcal{L}\|_q = \mathcal{O}_P\left((NT)^{1/4}\right).$$

By Lemma D.4, $\|\overline{\mathcal{H}}^{-1}\|_{q/(q-1)} = \|\overline{\mathcal{H}}^{-1}\|_q$ because $\overline{\mathcal{H}}^{-1}$ is symmetric, and

$$\left\|\overline{\mathcal{H}}^{-1}\right\| = \left\|\overline{\mathcal{H}}^{-1}\right\|_2 \leq \sqrt{\|\overline{\mathcal{H}}^{-1}\|_{q/(q-1)}\|\overline{\mathcal{H}}^{-1}\|_q} = \|\overline{\mathcal{H}}^{-1}\|_q = \mathcal{O}_P(1). \qquad \text{(B.2)}$$

Analogously,

$$\begin{aligned}
\|\partial_{\beta\phi\phi}\mathcal{L}\| &\leq \|\partial_{\beta\phi\phi}\mathcal{L}\|_q = \mathcal{O}_P\left((NT)^{\epsilon}\right), \\
\left\|\sum_g \partial_{\phi\phi'\phi_g}\mathcal{L}\,[\mathcal{H}^{-1}\mathcal{S}]_g\right\| &\leq \left\|\sum_g \partial_{\phi\phi'\phi_g}\mathcal{L}\,[\mathcal{H}^{-1}\mathcal{S}]_g\right\|_q \\
&\leq \|\partial_{\phi\phi\phi}\mathcal{L}\|_q \left\|\mathcal{H}^{-1}\right\|_q \|\mathcal{S}\|_q = \mathcal{O}_P\left((NT)^{-1/4+1/(2q)+\epsilon}\right), \\
\left\|\sum_g \partial_{\phi\phi'\phi_g}\mathcal{L}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g\right\| &\leq \left\|\sum_g \partial_{\phi\phi'\phi_g}\mathcal{L}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g\right\|_q \\
&\leq \|\partial_{\phi\phi\phi}\mathcal{L}\|_q \left\|\overline{\mathcal{H}}^{-1}\right\|_q \|\mathcal{S}\|_q = \mathcal{O}_P\left((NT)^{-1/4+1/(2q)+\epsilon}\right).
\end{aligned}$$

Assumption B.1 guarantees that $\left\|\overline{\mathcal{H}}^{-1}\right\| \left\|\widetilde{\mathcal{H}}\right\| < 1$ wpa1. Therefore,

$$\mathcal{H}^{-1} = \overline{\mathcal{H}}^{-1}\left(\mathbb{1} + \widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\right)^{-1} = \overline{\mathcal{H}}^{-1}\sum_{s=0}^{\infty}(-\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1})^s = \overline{\mathcal{H}}^{-1} - \overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1} + \overline{\mathcal{H}}^{-1}\sum_{s=2}^{\infty}(-\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1})^s.$$

Note that $\left\|\overline{\mathcal{H}}^{-1}\sum_{s=2}^{\infty}(-\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1})^s\right\| \leq \left\|\overline{\mathcal{H}}^{-1}\right\| \sum_{s=2}^{\infty}\left(\left\|\overline{\mathcal{H}}^{-1}\right\| \left\|\widetilde{\mathcal{H}}\right\|\right)^s$, and therefore

$$\left\|\mathcal{H}^{-1} - \left(\overline{\mathcal{H}}^{-1} - \overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\right)\right\| \leq \frac{\left\|\overline{\mathcal{H}}^{-1}\right\|^3 \left\|\widetilde{\mathcal{H}}\right\|^2}{1 - \left\|\overline{\mathcal{H}}^{-1}\right\| \left\|\widetilde{\mathcal{H}}\right\|} = o_P\left((NT)^{-1/4}\right),$$

by Assumption B.1($vi$) and equation (B.2).

The results for $\|\mathcal{H}^{-1}\|$ and $\left\|\mathcal{H}^{-1} - \overline{\mathcal{H}}^{-1}\right\|$ follow immediately. ■

### B.1.1 Legendre Transformed Objective Function

We consider the shrinking neighborhood $\mathcal{B}(r_\beta, \beta^0) \times \mathcal{B}_q(r_\phi, \phi^0)$ of the true parameters $(\beta^0, \phi^0)$. Statement ($i$) of Lemma B.5 implies that the objective function $\mathcal{L}(\beta, \phi)$ is strictly concave in $\phi$ in this shrinking neighborhood wpa1. We define

$$\mathcal{L}^*(\beta, S) = \max_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \left[\mathcal{L}(\beta, \phi) - \phi' S\right], \qquad \Phi(\beta, S) = \operatorname*{argmax}_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \left[\mathcal{L}(\beta, \phi) - \phi' S\right], \tag{B.3}$$

where $\beta \in \mathcal{B}(r_\beta, \beta^0)$ and $S \in \mathbb{R}^{\dim \phi}$. The function $\mathcal{L}^*(\beta, S)$ is the Legendre transformation of the objective function $\mathcal{L}(\beta, \phi)$ in the incidental parameter $\phi$. We denote the parameter $S$ as the dual parameter to $\phi$, and $\mathcal{L}^*(\beta, S)$ as the dual function to $\mathcal{L}(\beta, \phi)$. We only consider $\mathcal{L}^*(\beta, S)$ and $\Phi(\beta, S)$ for parameters $\beta \in \mathcal{B}(r_\beta, \beta^0)$ and $S \in \mathcal{S}(\beta, \mathcal{B}_q(r_\phi, \phi^0))$, where the optimal $\phi$ is defined by the first order conditions, i.e. is not a boundary solution. We define the corresponding set of pairs $(\beta, \mathcal{S})$ that is dual to $\mathcal{B}(r_\beta, \beta^0) \times \mathcal{B}_q(r_\phi, \phi^0)$ by

$$\mathcal{SB}_r(\beta^0, \phi^0) = \left\{(\beta, \mathcal{S}) \in \mathbb{R}^{\dim \beta + \dim \phi} \;:\; (\beta, \Phi(\beta, S)) \in \mathcal{B}(r_\beta, \beta^0) \times \mathcal{B}_q(r_\phi, \phi^0)\right\}.$$

Assumption B.1 guarantees that for $\beta \in \mathcal{B}(r_\beta, \beta^0)$ the domain $\mathcal{S}(\beta, \mathcal{B}_q(r_\phi, \phi^0))$ includes $S = 0$, the origin of $\mathbb{R}^{\dim \phi}$, as an interior point, wpa1, and that $\mathcal{L}^*(\beta, S)$ is four times differentiable in a neighborhood of $S = 0$ (see Lemma B.6 below). The optimal $\phi = \Phi(\beta, S)$ in equation (B.3) satisfies the first order condition $S = \mathcal{S}(\beta, \phi)$. Thus, for given $\beta$, the functions $\Phi(\beta, S)$ and $\mathcal{S}(\beta, \phi)$ are inverse to each other, and the relationship between $\phi$ and its dual $S$ is one-to-one. This is a consequence of strict concavity of $\mathcal{L}(\beta,\, \phi)$ in the neighborhood of the true parameter value that we consider here.[13] One can show that

$$\Phi(\beta, S) = -\frac{\partial \mathcal{L}^*(\beta,\, S)}{\partial S} \ ,$$

which shows the dual nature of the functions $\mathcal{L}(\beta,\, \phi)$ and $\mathcal{L}^*(\beta,\, S)$. For $S = 0$ the optimization in (B.3) is just over the objective function $\mathcal{L}(\beta, \phi)$, so that $\Phi(\beta, 0) = \widehat{\phi}(\beta)$ and $\mathcal{L}^*(\beta, 0) = \mathcal{L}(\beta, \widehat{\phi}(\beta))$, the profile objective function. We already introduced $\mathcal{S} = \mathcal{S}(\beta^0, \phi^0)$, i.e. at $\beta = \beta^0$ the dual of $\phi^0$ is $\mathcal{S}$, and vica versa. We can write the profile objective function $\mathcal{L}(\beta, \widehat{\phi}(\beta)) = \mathcal{L}^*(\beta, 0)$ as a Taylor series expansion of $\mathcal{L}^*(\beta,\, S)$ around $(\beta, S) = (\beta^0, \mathcal{S})$, namely

$$\mathcal{L}(\beta, \widehat{\phi}(\beta)) = \mathcal{L}^*(\beta^0, \mathcal{S}) + (\partial_{\beta'} \mathcal{L}^*)\Delta\beta - \Delta\beta'(\partial_{\beta S'} \mathcal{L}^*)\mathcal{S} + \frac{1}{2}\Delta\beta'(\partial_{\beta\beta'} \mathcal{L}^*)\Delta\beta + \dots \ ,$$

where $\Delta\beta = \beta - \beta^0$, and here and in the following we omit the arguments of $\mathcal{L}^*(\beta, S)$ and of its partial derivatives when they are evaluated at $(\beta^0, \mathcal{S})$. Analogously, we can obtain Taylor expansions for the profile score $\partial_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta)) = \partial_\beta \mathcal{L}^*(\beta, 0)$ and the estimated nuisance parameter $\widehat{\phi}(\beta) = -\partial_S \mathcal{L}^*(\beta, 0)$ in $\Delta\beta$ and $\mathcal{S}$, see the proof of Theorem B.1 below. Apart from combinatorial factors those expansions feature the same coefficients as the expansion of $\mathcal{L}(\beta, \widehat{\phi}(\beta))$ itself. They are standard Taylor expansions that can be truncated at a certain order, and the remainder term can be bounded by applying the mean value theorem.

The functions $\mathcal{L}(\beta,\, \phi)$ and its dual $\mathcal{L}^*(\beta,\, S)$ are closely related. In particular, for given $\beta$ their first derivatives with respect to the second argument $\mathcal{S}(\beta, \phi)$ and $\Phi(\beta, S)$ are inverse functions of each other. We can therefore express partial derivatives of $\mathcal{L}^*(\beta,\, S)$ in terms of partial derivatives of $\mathcal{L}(\beta,\, \phi)$. This is done in Lemma B.6. The norms $\|\partial_{\beta SSS} \mathcal{L}^*(\beta, S)\|_q$, $\|\partial_{SSSS} \mathcal{L}^*(\beta, S)\|_q$, etc., are defined as in equation (A.1) and (A.2).

**Lemma B.6.** *Let assumption B.1 be satisfied.*

*(i) The function $\mathcal{L}^*(\beta, S)$ is well-defined and is four times continuously differentiable in $\mathcal{SB}_r(\beta^0, \phi^0)$, wpa1.*

[13] Another consequence of strict concavity of $\mathcal{L}(\beta,\, \phi)$ is that the dual function $\mathcal{L}^*(\beta,\, S)$ is strictly convex in $S$. The original $\mathcal{L}(\beta,\, \phi)$ can be recovered from $\mathcal{L}^*(\beta,\, S)$ by again performing a Legendre transformation, namely

$$\mathcal{L}(\beta,\, \phi) \ = \ \min_{S \in \mathbb{R}^{\dim \phi}} \left[ \mathcal{L}^*(\beta,\, S) + \phi' S \right] \ .$$

*(ii)* *For* $\mathcal{L}^* = \mathcal{L}^*(\beta^0, \mathcal{S})$,

$$\partial_S \mathcal{L}^* = -\phi^0, \quad \partial_\beta \mathcal{L}^* = \partial_\beta \mathcal{L}, \quad \partial_{SS'} \mathcal{L}^* = -(\partial_{\phi\phi'}\mathcal{L})^{-1} = \mathcal{H}^{-1}, \quad \partial_{\beta S'} \mathcal{L}^* = -(\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1},$$

$$\partial_{\beta\beta'}\mathcal{L}^* = \partial_{\beta\beta'}\mathcal{L} + (\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi'\beta}\mathcal{L}), \quad \partial_{SS'S_g}\mathcal{L}^* = -\sum_h \mathcal{H}^{-1}(\partial_{\phi\phi'\phi_h}\mathcal{L})\mathcal{H}^{-1}(\mathcal{H}^{-1})_{gh},$$

$$\partial_{\beta_k SS'}\mathcal{L}^* = \mathcal{H}^{-1}(\partial_{\beta_k\phi'\phi}\mathcal{L})\mathcal{H}^{-1} + \sum_g \mathcal{H}^{-1}(\partial_{\phi_g\phi'\phi}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}\partial_{\beta_k\phi}\mathcal{L}]_g,$$

$$\begin{aligned}\partial_{\beta_k\beta_l S'}\mathcal{L}^* = &-(\partial_{\beta_k\beta_l\phi'}\mathcal{L})\mathcal{H}^{-1} - (\partial_{\beta_l\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta_k\phi\phi'}\mathcal{L})\mathcal{H}^{-1} - (\partial_{\beta_k\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta_l\phi'\phi}\mathcal{L})\mathcal{H}^{-1} \\ &- \sum_g (\partial_{\beta_k\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi_g\phi'\phi}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}\partial_{\beta_l\phi}\mathcal{L}]_g,\end{aligned}$$

$$\begin{aligned}\partial_{\beta_k\beta_l\beta_m}\mathcal{L}^* = &\,\partial_{\beta_k\beta_l\beta_m}\mathcal{L} + \sum_g (\partial_{\beta_k\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi_g\phi'\phi}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta_l\phi}\mathcal{L})[\mathcal{H}^{-1}\partial_{\phi\beta_m}\mathcal{L}]_g \\ &+ (\partial_{\beta_k\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta_l\phi'\phi}\mathcal{L})\mathcal{H}^{-1}\partial_{\phi\beta_m}\mathcal{L} + (\partial_{\beta_m\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta_k\phi'\phi}\mathcal{L})\mathcal{H}^{-1}\partial_{\phi\beta_l}\mathcal{L} \\ &+ (\partial_{\beta_l\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta_m\phi'\phi}\mathcal{L})\mathcal{H}^{-1}\partial_{\phi\beta_k}\mathcal{L} \\ &+ (\partial_{\beta_k\beta_l\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi'\beta_m}\mathcal{L}) + (\partial_{\beta_k\beta_m\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi'\beta_l}\mathcal{L}) + (\partial_{\beta_l\beta_m\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi'\beta_k}\mathcal{L}),\end{aligned}$$

*and*

$$\begin{aligned}\partial_{SS'S_gS_h}\mathcal{L}^* = &\sum_{f,e} \mathcal{H}^{-1}(\partial_{\phi\phi'\phi_f\phi_e}\mathcal{L})\mathcal{H}^{-1}(\mathcal{H}^{-1})_{gf}(\mathcal{H}^{-1})_{he} \\ &+ 3\sum_{f,e} \mathcal{H}^{-1}(\partial_{\phi\phi'\phi_e}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi\phi'\phi_f}\mathcal{L})\mathcal{H}^{-1}(\mathcal{H}^{-1})_{gf}(\mathcal{H}^{-1})_{he},\end{aligned}$$

$$\begin{aligned}\partial_{\beta_k SS'S_g}\mathcal{L}^* = &-\sum_h \mathcal{H}^{-1}(\partial_{\beta_k\phi'\phi}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi\phi'\phi_h}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}]_{gh} \\ &- \sum_h \mathcal{H}^{-1}(\partial_{\phi\phi'\phi_h}\mathcal{L})\mathcal{H}^{-1}(\partial_{\beta_k\phi'\phi}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}]_{gh} \\ &- \sum_h \mathcal{H}^{-1}(\partial_{\phi\phi'\phi_h}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}(\partial_{\beta_k\phi'\phi}\mathcal{L})\mathcal{H}^{-1}]_{gh} \\ &- \sum_{h,f} \mathcal{H}^{-1}(\partial_{\phi_f\phi'\phi}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi\phi'\phi_h}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}]_{gh}[\mathcal{H}^{-1}\partial_{\beta_k\phi}\mathcal{L}]_f \\ &- \sum_{h,f} \mathcal{H}^{-1}(\partial_{\phi\phi'\phi_h}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi_f\phi'\phi}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}]_{gh}[\mathcal{H}^{-1}\partial_{\beta_k\phi}\mathcal{L}]_f \\ &- \sum_{h,f} \mathcal{H}^{-1}(\partial_{\phi\phi'\phi_h}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}(\partial_{\phi_f\phi'\phi}\mathcal{L})\mathcal{H}^{-1}]_{gh}[\mathcal{H}^{-1}\partial_{\beta_k\phi}\mathcal{L}]_f \\ &- \sum_h \mathcal{H}^{-1}(\partial_{\beta_k\phi\phi'\phi_h}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}]_{gh} \\ &- \sum_{h,f} \mathcal{H}^{-1}(\partial_{\phi\phi'\phi_h\phi_f}\mathcal{L})\mathcal{H}^{-1}[\mathcal{H}^{-1}]_{gh}[\mathcal{H}^{-1}(\partial_{\beta_k\phi}\mathcal{L})]_f.\end{aligned}$$

*(iii) Moreover,*

$$
\begin{aligned}
\sup_{(\beta,S)\in\mathcal{SB}_r(\beta^0,\phi^0)} \|\partial_{\beta\beta\beta}\mathcal{L}^*(\beta,S)\| &= \mathcal{O}_P\left((NT)^{1/2+1/(2q)+\epsilon}\right), \\
\sup_{(\beta,S)\in\mathcal{SB}_r(\beta^0,\phi^0)} \|\partial_{\beta\beta S}\mathcal{L}^*(\beta,S)\|_q &= \mathcal{O}_P\left((NT)^{1/q+\epsilon}\right), \\
\sup_{(\beta,S)\in\mathcal{SB}_r(\beta^0,\phi^0)} \|\partial_{\beta SS}\mathcal{L}^*(\beta,S)\|_q &= \mathcal{O}_P\left((NT)^{1/(2q)+\epsilon}\right), \\
\sup_{(\beta,S)\in\mathcal{SB}_r(\beta^0,\phi^0)} \|\partial_{\beta SSS}\mathcal{L}^*(\beta,S)\|_q &= \mathcal{O}_P\left((NT)^{1/(2q)+2\epsilon}\right), \\
\sup_{(\beta,S)\in\mathcal{SB}_r(\beta^0,\phi^0)} \|\partial_{SSSS}\mathcal{L}^*(\beta,S)\|_q &= \mathcal{O}_P\left((NT)^{2\epsilon}\right).
\end{aligned}
$$

**Proof of Lemma B.6.** #Part $(i)$: According to the definition (B.3), $\mathcal{L}^*(\beta,S) = \mathcal{L}(\beta,\Phi(\beta,S)) - \Phi(\beta,S)'S$, where $\Phi(\beta,S)$ solves the FOC, $\mathcal{S}(\beta,\Phi(\beta,S)) = S$, i.e. $\mathcal{S}(\beta,.)$ and $\Phi(\beta,.)$ are inverse functions for every $\beta$. Taking the derivative of $\mathcal{S}(\beta,\Phi(\beta,S)) = S$ wrt to both $S$ and $\beta$ yields

$$
\begin{aligned}
[\partial_S\Phi(\beta,S)'][\partial_\phi\mathcal{S}(\beta,\Phi(\beta,S))'] &= \mathbb{1}, \\
[\partial_\beta\mathcal{S}(\beta,\Phi(\beta,S))'] + [\partial_\beta\Phi(\beta,S)'][\partial_\phi\mathcal{S}(\beta,\Phi(\beta,S))'] &= 0.
\end{aligned} \tag{B.4}
$$

By definition, $\mathcal{S} = \mathcal{S}(\beta^0,\phi^0)$. Therefore, $\Phi(\beta,S)$ is the unique function that satisfies the boundary condition $\Phi(\beta^0,\mathcal{S}) = \phi^0$ and the system of partial differential equations (PDE) in (B.4). Those PDE's can equivalently be written as

$$
\begin{aligned}
\partial_S\Phi(\beta,S)' &= -[\mathcal{H}(\beta,\Phi(\beta,S))]^{-1}, \\
\partial_\beta\Phi(\beta,S)' &= [\partial_{\beta\phi'}\mathcal{L}(\beta,\Phi(\beta,S))][\mathcal{H}(\beta,\Phi(\beta,S))]^{-1}.
\end{aligned} \tag{B.5}
$$

This shows that $\Phi(\beta,S)$ (and thus $\mathcal{L}^*(\beta,S)$) are well-defined in any neighborhood of $(\beta,S) = (\beta^0,\mathcal{S})$ in which $\mathcal{H}(\beta,\Phi(\beta,S))$ is invertible (inverse function theorem). Lemma B.5 shows that $\mathcal{H}(\beta,\phi)$ is invertible in $\mathcal{B}(r_\beta,\beta^0)\times\mathcal{B}_q(r_\phi,\phi^0)$, wpa1. The inverse function theorem thus guarantee that $\Phi(\beta,S)$ and $\mathcal{L}^*(\beta,S)$ are well-defined in $\mathcal{SB}_r(\beta^0,\phi^0)$. The partial derivatives of $\mathcal{L}^*(\beta,S)$ of up to fourth order can be expressed as continuous transformations of the partial derivatives of $\mathcal{L}(\beta,\phi)$ up to fourth order (see e.g. proof of part $(ii)$ of the lemma). Hence, $\mathcal{L}^*(\beta,S)$ is four times continuously differentiable because $\mathcal{L}(\beta,\phi)$ is four times continuously differentiable.

#Part $(ii)$: Differentiating $\mathcal{L}^*(\beta,S) = \mathcal{L}(\beta,\Phi(\beta,S)) - \Phi(\beta,S)'S$ wrt $\beta$ and $S$ and using the FOC of the maximization over $\phi$ in the definition of $\mathcal{L}^*(\beta,S)$ gives $\partial_\beta\mathcal{L}^*(\beta,S) = \partial_\beta\mathcal{L}(\beta,\Phi(\beta,S))$ and $\partial_S\mathcal{L}^*(\beta,S) = -\Phi(\beta,S)$, respectively. Evaluating this expression at $(\beta,S) = (\beta^0,\mathcal{S})$ gives the first two statements of part $(ii)$.

Using $\partial_S\mathcal{L}^*(\beta,S) = -\Phi(\beta,S)$, the PDE (B.5) can be written as

$$
\begin{aligned}
\partial_{SS'}\mathcal{L}^*(\beta,S) &= \mathcal{H}^{-1}(\beta,\Phi(\beta,S)), \\
\partial_{\beta S'}\mathcal{L}^*(\beta,S) &= -[\partial_{\beta\phi'}\mathcal{L}(\beta,\Phi(\beta,S))]\mathcal{H}^{-1}(\beta,\Phi(\beta,S)).
\end{aligned}
$$

Evaluating this expression at $(\beta,S) = (\beta^0,\mathcal{S})$ gives the next two statements of part $(ii)$.

Taking the derivative of $\partial_\beta \mathcal{L}^*(\beta,S) = \partial_\beta \mathcal{L}(\beta,\Phi(\beta,S))$ wrt to $\beta$ and using the second equation of (B.5) gives the next statement when evaluated at $(\beta,S)=(\beta^0,\mathcal{S})$.

Taking the derivative of $\partial_{SS'}\mathcal{L}^*(\beta,S) = -[\partial_{\phi\phi'}\mathcal{L}(\beta,\Phi(\beta,S))]^{-1}$ wrt to $S_g$ and using the first equation of (B.5) gives the next statement when evaluated at $(\beta,S)=(\beta^0,\mathcal{S})$.

Taking the derivative of $\partial_{SS'}\mathcal{L}^*(\beta,S) = -[\partial_{\phi\phi'}\mathcal{L}(\beta,\Phi(\beta,S))]^{-1}$ wrt to $\beta_k$ and using the second equation of (B.5) gives

$$\begin{aligned}\partial_{\beta_k SS'}\mathcal{L}^*(\beta,S) &= \mathcal{H}^{-1}(\beta,\phi)[\partial_{\beta_k\phi'\phi}\mathcal{L}(\beta,\phi)]\mathcal{H}^{-1}(\beta,\phi)\\ &\quad+\sum_g \mathcal{H}^{-1}(\beta,\phi)[\partial_{\phi_g\phi'\phi}\mathcal{L}(\beta,\phi)]\mathcal{H}^{-1}(\beta,\phi)\{\mathcal{H}^{-1}(\beta,\phi)[\partial_{\beta_k\phi}\mathcal{L}(\beta,\phi)]\}_g,\end{aligned} \tag{B.6}$$

where $\phi=\Phi(\beta,S)$. This becomes the next statement when evaluated at $(\beta,S)=(\beta^0,\mathcal{S})$.

We omit the proofs for $\partial_{\beta_k\beta_l S'}\mathcal{L}^*$, $\partial_{\beta_k\beta_l S}\mathcal{L}^*$, $\partial_{SS'S_gS_h}\mathcal{L}^*$ and $\partial_{\beta_k SS'S_g}\mathcal{L}^*$ because they are analogous.

#Part $(iii)$: We only show the result for $\left\|\partial_{\beta SS}\mathcal{L}^*(\beta,S)\right\|_q$, the proof of the other statements is analogous. By equation (B.6)

$$\left\|\partial_{\beta SS}\mathcal{L}^*(\beta,S)\right\|_q \le \left\|\mathcal{H}^{-1}(\beta,\phi)\right\|_q^2 \left\|\partial_{\beta\phi\phi}\mathcal{L}(\beta,\phi)\right\|_q + \left\|\mathcal{H}^{-1}(\beta,\phi)\right\|_q^3 \left\|\partial_{\phi\phi\phi}\mathcal{L}(\beta,\phi)\right\|_q \left\|\partial_{\beta\phi'}\mathcal{L}(\beta,\phi)\right\|_q,$$

where $\phi=\Phi(\beta,S)$. Then, by Lemma B.5

$$\begin{aligned}\sup_{(\beta,S)\in\mathcal{SB}_r(\beta^0,\phi^0)}\left\|\partial_{\beta SS}\mathcal{L}^*(\beta,S)\right\|_q &\le \sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\ \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\Big[\left\|\mathcal{H}^{-1}(\beta,\phi)\right\|_q^2\left\|\partial_{\beta\phi\phi}\mathcal{L}(\beta,\phi)\right\|_q\\ &\quad+\left\|\mathcal{H}^{-1}(\beta,\phi)\right\|_q^3\left\|\partial_{\phi\phi\phi}\mathcal{L}(\beta,\phi)\right\|_q\left\|\partial_{\beta\phi'}\mathcal{L}(\beta,\phi)\right\|_q\Big] = \mathcal{O}\left((NT)^{1/(2q)+\epsilon}\right).\end{aligned}$$

To derive the rest of the bounds we can use that the expressions from part $(ii)$ hold not only for $(\beta^0,\mathcal{S})$, but also for other values $(\beta,S)$, provided that $(\beta,\Phi(\beta,S)$ is used as the argument on the rhs expressions. ■

### B.1.2 Proofs of Theorem B.1, Corollary B.2, and Theorem B.3

**Proof of Theorem B.1, Part 1: Expansion of $\widehat{\phi}(\beta)$.** Let $\beta=\beta_{NT}\in\mathcal{B}(\beta^0,r_\beta)$. A Taylor expansion of $\partial_S\mathcal{L}^*(\beta,0)$ around $(\beta^0,\mathcal{S})$ gives

$$\widehat{\phi}(\beta) = -\partial_S\mathcal{L}^*(\beta,0) = -\partial_S\mathcal{L}^* - (\partial_{S\beta'}\mathcal{L}^*)\Delta\beta + (\partial_{SS'}\mathcal{L}^*)\mathcal{S} - \frac{1}{2}\sum_g(\partial_{SS'S_g}\mathcal{L}^*)\mathcal{S}\mathcal{S}_g + R^\phi(\beta),$$

where we first expand in $\beta$ holding $S=\mathcal{S}$ fixed, and then expand in $S$. For any $v\in\mathbb{R}^{\dim\phi}$ the remainder term satisfies

$$\begin{aligned}v'R^\phi(\beta) = v'\Big\{&-\frac{1}{2}\sum_k[\partial_{S\beta'\beta_k}\mathcal{L}^*(\tilde\beta,\mathcal{S})](\Delta\beta)(\Delta\beta_k) + \sum_k[\partial_{SS'\beta_k}\mathcal{L}^*(\beta^0,\tilde S)]\mathcal{S}(\Delta\beta_k)\\ &+\frac{1}{6}\sum_{g,h}[\partial_{SS'S_gS_h}\mathcal{L}^*(\beta^0,\bar S)]\mathcal{S}\mathcal{S}_g\mathcal{S}_h\Big\},\end{aligned}$$

where $\tilde\beta$ is between $\beta^0$ and $\beta$, and $\tilde S$ and $\bar S$ are between 0 and $\mathcal{S}$. By part $(ii)$ of Lemma B.6,

$$\widehat{\phi}(\beta)-\phi^0 = \mathcal{H}^{-1}(\partial_{\phi\beta'}\mathcal{L})\Delta\beta + \mathcal{H}^{-1}\mathcal{S} + \tfrac{1}{2}\mathcal{H}^{-1}\sum_g(\partial_{\phi\phi'\phi_g}\mathcal{L})\mathcal{H}^{-1}\mathcal{S}(\mathcal{H}^{-1}\mathcal{S})_g + R^\phi(\beta).$$

Using that the vector norm $\|.\|_{q/(q-1)}$ is the dual to the vector norm $\|.\|_q$, Assumption B.1, and Lemmas B.5 and B.6 yields

$$
\begin{aligned}
\left\|R^{\phi}(\beta)\right\|_q &= \sup_{\|v\|_{q/(q-1)}=1} v' R^{\phi}(\beta) \\
&\leq \frac{1}{2}\left\|\partial_{S\beta\beta}\mathcal{L}^*(\tilde{\beta},\mathcal{S})\right\|_q \|\Delta\beta\|^2 + \left\|\partial_{SS\beta}\mathcal{L}^*(\beta^0,\tilde{S})\right\|_q \|\mathcal{S}\|_q\|\Delta\beta\| + \frac{1}{6}\left\|\partial_{SSSS}\mathcal{L}^*(\beta^0,\bar{S})\right\|_q \|\mathcal{S}\|_q^3 \\
&= \mathcal{O}_P\left[(NT)^{1/q+\epsilon} r_\beta \|\Delta\beta\| + (NT)^{-1/4+1/q+\epsilon}\|\Delta\beta\| + (NT)^{-3/4+3/(2q)+2\epsilon}\right] \\
&= o_P\left((NT)^{-1/2+1/(2q)}\right) + o_P\left((NT)^{1/(2q)}\|\beta-\beta^0\|\right),
\end{aligned}
$$

uniformly over $\beta \in \mathcal{B}(\beta^0, r_\beta)$ by Lemma B.6. ■

**Proof of Theorem B.1, Part 2: Expansion of profile score.** Let $\beta = \beta_{NT} \in \mathcal{B}(\beta^0, r_\beta)$. A Taylor expansion of $\partial_\beta \mathcal{L}^*(\beta, 0)$ around $(\beta^0, \mathcal{S})$ gives

$$
\partial_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta)) = \partial_\beta \mathcal{L}^*(\beta, 0) = \partial_\beta \mathcal{L}^* + (\partial_{\beta\beta'}\mathcal{L}^*)\Delta\beta - (\partial_{\beta S'}\mathcal{L}^*)\mathcal{S} + \frac{1}{2}\sum_g (\partial_{\beta S' S_g}\mathcal{L}^*)\mathcal{S}\mathcal{S}_g + R_1(\beta),
$$

where we first expand in $\beta$ for fixed $S = \mathcal{S}$, and then expand in $S$. For any $v \in \mathbb{R}^{\dim \beta}$ the remainder term satisfies

$$
\begin{aligned}
v' R_1(\beta) = v' \Bigg\{ &\frac{1}{2}\sum_k [\partial_{\beta\beta'\beta_k}\mathcal{L}^*(\tilde{\beta},\mathcal{S})](\Delta\beta)(\Delta\beta_k) - \sum_k [\partial_{\beta\beta_k S'}\mathcal{L}^*(\beta^0,\tilde{S})]\mathcal{S}(\Delta\beta_k) \\
&- \frac{1}{6}\sum_{g,h}[\partial_{\beta S' S_g S_h}\mathcal{L}^*(\beta^0,\bar{S})]\mathcal{S}\mathcal{S}_g\mathcal{S}_h \Bigg\},
\end{aligned}
$$

where $\tilde{\beta}$ is between $\beta^0$ and $\beta$, and $\tilde{S}$ and $\bar{S}$ are between 0 and $\mathcal{S}$. By Lemma B.6,

$$
\begin{aligned}
\partial_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta)) = \ &\partial_\beta \mathcal{L} + \left[\partial_{\beta\beta'}\mathcal{L} + (\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi'\beta}\mathcal{L})\right](\beta - \beta^0) + (\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1}\mathcal{S} \\
&+ \frac{1}{2}\sum_g \left(\partial_{\beta\phi'\phi_g}\mathcal{L} + [\partial_{\beta\phi'}\mathcal{L}]\,\mathcal{H}^{-1}[\partial_{\phi\phi'\phi_g}\mathcal{L}]\right)[\mathcal{H}^{-1}\mathcal{S}]_g \mathcal{H}^{-1}\mathcal{S} + R_1(\beta),
\end{aligned}
$$

where for any $v \in \mathbb{R}^{\dim \beta}$,

$$
\begin{aligned}
\|R_1(\beta)\| &= \sup_{\|v\|=1} v' R_1(\beta) \\
&\leq \frac{1}{2}\left\|\partial_{\beta\beta\beta}\mathcal{L}^*(\tilde{\beta},\mathcal{S})\right\| \|\Delta\beta\|^2 + (NT)^{1/2-1/q}\left\|\partial_{\beta\beta S}\mathcal{L}^*(\beta^0,\tilde{S})\right\|_q \|\mathcal{S}\|_q \|\Delta\beta\| \\
&\quad + \frac{1}{6}(NT)^{1/2-1/q}\left\|\partial_{\beta SSS}\mathcal{L}^*(\beta^0,\bar{S})\right\|_q \|\mathcal{S}\|_q^3 \\
&= \mathcal{O}_P\left[(NT)^{1/2+1/(2q)+\epsilon} r_\beta \|\Delta\beta\| + (NT)^{1/4+1/(2q)+\epsilon}\|\Delta\beta\| + (NT)^{-1/4+1/q+2\epsilon}\right] \\
&= o_P(1) + o_P(\sqrt{NT}\|\beta-\beta^0\|),
\end{aligned}
$$

uniformly over $\beta \in \mathcal{B}(\beta^0, r_\beta)$ by Lemma B.6. We can also write

$$
\begin{aligned}
d_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta)) &= \partial_\beta \mathcal{L} - \sqrt{NT}\,\overline{W}(\Delta\beta) + (\partial_{\beta\phi'}\overline{\mathcal{L}})\overline{\mathcal{H}}^{-1}\mathcal{S} + (\partial_{\beta\phi'}\widetilde{\mathcal{L}})\overline{\mathcal{H}}^{-1}\mathcal{S} - (\partial_{\beta\phi'}\overline{\mathcal{L}})\overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\mathcal{S} \\
&\quad + \frac{1}{2}\sum_g \left(\partial_{\beta\phi'\phi_g}\overline{\mathcal{L}} + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}[\partial_{\phi\phi'\phi_g}\overline{\mathcal{L}}]\right)[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g \overline{\mathcal{H}}^{-1}\mathcal{S} + R(\beta), \\
&= U - \sqrt{NT}\,\overline{W}(\Delta\beta) + R(\beta),
\end{aligned}
$$

where we decompose the term linear in $\mathcal{S}$ into multiple terms by using that

$$-(\partial_{\beta S'}\mathcal{L}^*) = (\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1} = \left[(\partial_{\beta\phi'}\overline{\mathcal{L}}) + (\partial_{\beta\phi'}\widetilde{\mathcal{L}})\right]\left[\overline{\mathcal{H}}^{-1} - \overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1} + \dots\right].$$

The new remainder term is

$$\begin{aligned}
R(\beta) = R_1(\beta) &+ (\partial_{\beta\beta'}\widetilde{\mathcal{L}})\Delta\beta + \left[(\partial_{\beta\phi'}\mathcal{L})\mathcal{H}^{-1}(\partial_{\phi'\beta}\mathcal{L}) - (\partial_{\beta\phi'}\overline{\mathcal{L}})\overline{\mathcal{H}}^{-1}(\partial_{\phi'\beta}\overline{\mathcal{L}})\right]\Delta\beta \\
&+ (\partial_{\beta\phi'}\mathcal{L})\left[\mathcal{H}^{-1} - \left(\overline{\mathcal{H}}^{-1} - \overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\right)\right]\mathcal{S} - (\partial_{\beta\phi'}\widetilde{\mathcal{L}})\overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\mathcal{S} \\
&+ \frac{1}{2}\left[\sum_g \partial_{\beta\phi'\phi_g}\mathcal{L}[\mathcal{H}^{-1}\mathcal{S}]_g\mathcal{H}^{-1}\mathcal{S} - \sum_g \partial_{\beta\phi'\phi_g}\overline{\mathcal{L}}[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g\overline{\mathcal{H}}^{-1}\mathcal{S}\right] \\
&+ \frac{1}{2}\left[\sum_g [\partial_{\beta\phi'}\mathcal{L}]\,\mathcal{H}^{-1}[\partial_{\phi\phi'\phi_g}\mathcal{L}][\mathcal{H}^{-1}\mathcal{S}]_g\mathcal{H}^{-1}\mathcal{S} - \sum_g [\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}[\partial_{\phi\phi'\phi_g}\overline{\mathcal{L}}][\overline{\mathcal{H}}^{-1}\mathcal{S}]_g\overline{\mathcal{H}}^{-1}\mathcal{S}\right].
\end{aligned}$$

By Assumption B.1 and Lemma B.5,

$$\begin{aligned}
\|R(\beta)\| \leq \|R_1(\beta)\| &+ \left\|\partial_{\beta\beta'}\widetilde{\mathcal{L}}\right\|\|\Delta\beta\| + \|\partial_{\beta\phi'}\mathcal{L}\|\left\|\mathcal{H}^{-1} - \overline{\mathcal{H}}^{-1}\right\|\|\partial_{\phi'\beta}\mathcal{L}\|\,\|\Delta\beta\| \\
&+ \left\|\partial_{\beta\phi'}\widetilde{\mathcal{L}}\right\|\left\|\overline{\mathcal{H}}^{-1}\right\|\left(\|\partial_{\phi'\beta}\mathcal{L}\| + \|\partial_{\phi'\beta}\overline{\mathcal{L}}\|\right)\|\Delta\beta\| \\
&+ \|\partial_{\beta\phi'}\mathcal{L}\|\left\|\mathcal{H}^{-1} - \left(\overline{\mathcal{H}}^{-1} - \overline{\mathcal{H}}^{-1}\widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\right)\right\|\|\mathcal{S}\| + \left\|\overline{\mathcal{H}}^{-1}\right\|^2\left\|\partial_{\beta\phi'}\widetilde{\mathcal{L}}\right\|\left\|\widetilde{\mathcal{H}}\right\|\|\mathcal{S}\| \\
&+ \frac{1}{2}\|\partial_{\beta\phi\phi}\mathcal{L}\|\left(\|\mathcal{H}^{-1}\| + \left\|\overline{\mathcal{H}}^{-1}\right\|\right)\left\|\mathcal{H}^{-1} - \overline{\mathcal{H}}^{-1}\right\|\|\mathcal{S}\|^2 \\
&+ \frac{1}{2}\left\|\overline{\mathcal{H}}^{-1}\right\|^2\left\|\partial_{\beta\phi\phi}\widetilde{\mathcal{L}}\right\|\|\mathcal{S}\|^2 \\
&+ \frac{1}{2}\left\|\sum_g [\partial_{\beta\phi'}\mathcal{L}]\,\mathcal{H}^{-1}[\partial_{\phi\phi'\phi_g}\mathcal{L}][\mathcal{H}^{-1}\mathcal{S}]_g\mathcal{H}^{-1}\mathcal{S} - \sum_g [\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}[\partial_{\phi\phi'\phi_g}\overline{\mathcal{L}}][\overline{\mathcal{H}}^{-1}\mathcal{S}]_g\overline{\mathcal{H}}^{-1}\mathcal{S}\right\| \\
&= \|R_1(\beta)\| + o_P(1) + o_P(\sqrt{NT}\|\beta - \beta^0\|) + \mathcal{O}_P\left[(NT)^{-1/8+\epsilon+1/(2q)}\right] \\
&= o_P(1) + o_P(\sqrt{NT}\|\beta - \beta^0\|),
\end{aligned}$$

uniformly over $\beta \in \mathcal{B}(\beta^0, r_\beta)$. Here we use that

$$\begin{aligned}
&\left\|\sum_g [\partial_{\beta\phi'}\mathcal{L}]\,\mathcal{H}^{-1}[\partial_{\phi\phi'\phi_g}\mathcal{L}][\mathcal{H}^{-1}\mathcal{S}]_g\mathcal{H}^{-1}\mathcal{S} - \sum_g [\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}[\partial_{\phi\phi'\phi_g}\overline{\mathcal{L}}][\overline{\mathcal{H}}^{-1}\mathcal{S}]_g\overline{\mathcal{H}}^{-1}\mathcal{S}\right\| \\
&\leq \|\partial_{\beta\phi'}\mathcal{L}\|\left\|\mathcal{H}^{-1} - \overline{\mathcal{H}}^{-1}\right\|\left(\|\mathcal{H}^{-1}\| + \left\|\overline{\mathcal{H}}^{-1}\right\|\right)\|\mathcal{S}\|\left\|\sum_g \partial_{\phi\phi'\phi_g}\mathcal{L}\,[\mathcal{H}^{-1}\mathcal{S}]_g\right\| \\
&\quad + \|\partial_{\beta\phi'}\mathcal{L}\|\left\|\mathcal{H}^{-1} - \overline{\mathcal{H}}^{-1}\right\|\left\|\overline{\mathcal{H}}^{-1}\right\|\|\mathcal{S}\|\left\|\sum_g \partial_{\phi\phi'\phi_g}\mathcal{L}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g\right\| \\
&\quad + \left\|\partial_{\beta\phi'}\widetilde{\mathcal{L}}\right\|\left\|\overline{\mathcal{H}}^{-1}\right\|^2\|\mathcal{S}\|\left\|\sum_g \partial_{\phi\phi'\phi_g}\mathcal{L}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g\right\| \\
&\quad + \|\partial_{\beta\phi'}\overline{\mathcal{L}}\|\left\|\overline{\mathcal{H}}^{-1}\right\|\left\|\sum_{g,h} \partial_{\phi\phi_g\phi_h}\widetilde{\mathcal{L}}\,[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g[\overline{\mathcal{H}}^{-1}\mathcal{S}]_h\right\|.
\end{aligned}$$

■

**Proof of Corollary B.2.** $\widehat{\beta}$ solves the FOC

$$\partial_\beta \mathcal{L}(\widehat{\beta}, \widehat{\phi}(\widehat{\beta})) = 0.$$

By $\left\|\widehat{\beta} - \beta^0\right\| = o_P(r_\beta)$ and Theorem B.1,

$$0 = \partial_\beta \mathcal{L}(\widehat{\beta}, \widehat{\phi}(\widehat{\beta})) = U - \overline{W}\sqrt{NT}(\widehat{\beta} - \beta^0) + o_P(1) + o_P(\sqrt{NT}\|\widehat{\beta} - \beta^0\|).$$

Thus, $\sqrt{NT}(\widehat{\beta} - \beta^0) = \overline{W}^{-1}U + o_P(1) + o_P(\sqrt{NT}\|\widehat{\beta} - \beta^0\|) = \overline{W}_\infty^{-1}U + o_P(1) + o_P(\sqrt{NT}\|\widehat{\beta} - \beta^0\|)$, where we use that $\overline{W} = \overline{W}_\infty + o_P(1)$ is invertible wpa1 and that $\overline{W}^{-1} = \overline{W}_\infty^{-1} + o_P(1)$. We conclude that $\sqrt{NT}(\widehat{\beta} - \beta^0) = \mathcal{O}_P(1)$ because $U = \mathcal{O}_P(1)$, and therefore $\sqrt{NT}(\widehat{\beta} - \beta^0) = \overline{W}_\infty^{-1}U + o_P(1)$. ■

**Proof of Theorem B.3.** # Consistency of $\widehat{\phi}(\beta)$: Let $\eta = \eta_{NT} > 0$ be such that $\eta = o_P(r_\phi)$, $(NT)^{-1/4+1/(2q)} = o_P(\eta)$, and $(NT)^{1/(2q)}r_\beta = o_P(\eta)$. For $\beta \in \mathcal{B}(r_\beta, \beta^0)$, define

$$\widehat{\phi}^*(\beta) := \underset{\{\phi:\, \|\phi - \phi^0\|_q \leq \eta\}}{\operatorname{argmin}} \|\mathcal{S}(\beta, \phi)\|_q. \tag{B.7}$$

Then, $\|\mathcal{S}(\beta, \widehat{\phi}^*(\beta))\|_q \leq \|\mathcal{S}(\beta, \phi^0)\|_q$, and therefore by a Taylor expansion of $\mathcal{S}(\beta, \phi^0)$ around $\beta = \beta^0$,

$$\begin{aligned}\|\mathcal{S}(\beta, \widehat{\phi}^*(\beta)) - \mathcal{S}(\beta, \phi^0)\|_q &\leq \|\mathcal{S}(\beta, \widehat{\phi}^*(\beta))\|_q + \|\mathcal{S}(\beta, \phi^0)\|_q \leq 2\|\mathcal{S}(\beta, \phi^0)\|_q \\ &\leq 2\|\mathcal{S}\|_q + 2\left\|\partial_{\phi\beta'}\mathcal{L}(\tilde{\beta}, \phi^0)\right\|_q \|\beta - \beta^0\| \\ &= \mathcal{O}_P\left[(NT)^{-1/4+1/(2q)} + (NT)^{1/(2q)}\|\beta - \beta^0\|\right],\end{aligned}$$

uniformly over $\beta \in \mathcal{B}(r_\beta, \beta^0)$, where $\tilde{\beta}$ is between $\beta^0$ and $\beta$, and we use Assumption B.1($v$) and Lemma B.5. Thus,

$$\sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \|\mathcal{S}(\beta, \widehat{\phi}^*(\beta)) - \mathcal{S}(\beta, \phi^0)\|_q = \mathcal{O}_P\left[(NT)^{-1/4+1/(2q)} + (NT)^{1/(2q)}r_\beta\right].$$

By a Taylor expansion of $\Phi(\beta, S)$ around $S = \mathcal{S}(\beta, \phi^0)$,

$$\begin{aligned}\left\|\widehat{\phi}^*(\beta) - \phi^0\right\|_q &= \left\|\Phi(\beta, \mathcal{S}(\beta, \widehat{\phi}^*(\beta))) - \Phi(\beta, \mathcal{S}(\beta, \phi^0))\right\|_q \leq \left\|\partial_S \Phi(\beta, \tilde{S})'\right\|_q \left\|\mathcal{S}(\beta, \widehat{\phi}^*(\beta)) - \mathcal{S}(\beta, \phi^0)\right\|_q \\ &= \left\|\mathcal{H}^{-1}(\beta, \Phi(\beta, \tilde{S}))\right\|_q \left\|\mathcal{S}(\beta, \widehat{\phi}^*(\beta)) - \mathcal{S}(\beta, \phi^0)\right\|_q = \mathcal{O}_P(1)\left\|\mathcal{S}(\beta, \widehat{\phi}^*(\beta)) - \mathcal{S}(\beta, \phi^0)\right\|_q,\end{aligned}$$

where $\tilde{S}$ is between $\mathcal{S}(\beta, \widehat{\phi}^*(\beta))$ and $\mathcal{S}(\beta, \phi^0)$ and we use Lemma B.5($i$). Thus,

$$\sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \left\|\widehat{\phi}^*(\beta) - \phi^0\right\|_q = \mathcal{O}_P\left[(NT)^{-1/4+1/(2q)} + (NT)^{1/(2q)}r_\beta\right] = o_P(\eta).$$

This shows that $\widehat{\phi}^*(\beta)$ is an interior solution of the minimization problem (B.7), wpa1. Thus, $\mathcal{S}(\beta, \widehat{\phi}^*(\beta)) = 0$, because the objective function $\mathcal{L}(\beta, \phi)$ is strictly concave and differentiable, and therefore $\widehat{\phi}^*(\beta) = \widehat{\phi}(\beta)$. We conclude that $\sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \left\|\widehat{\phi}(\beta) - \phi^0\right\|_q = \mathcal{O}_P(\eta) = o_P(r_\phi)$.

# Consistency of $\widehat{\beta}$: We have already shown that Assumption B.1($ii$) is satisfied, in addition to the remaining parts of Assumption B.1, which we assume. The bounds on the spectral norm in Assumption B.1($vi$) and in part ($ii$) of Lemma B.5 can be used to show that $U = \mathcal{O}_P((NT)^{1/4})$.

First, we consider the case $\dim(\beta) = 1$ first. The extension to $\dim(\beta) > 1$ is discussed below. Let $\eta = 2(NT)^{-1/2}\overline{W}^{-1}|U|$. Our goal is to show that $\widehat{\beta} \in [\beta^0 - \eta, \beta^0 + \eta]$. By Theorem B.1,

$$\partial_\beta \mathcal{L}(\beta^0 + \eta, \widehat{\phi}(\beta^0 + \eta)) = U - \overline{W}\sqrt{NT}\eta + o_P(1) + o_P(\sqrt{NT}\eta) = o_P(\sqrt{NT}\eta) - \overline{W}\sqrt{NT}\eta,$$
$$\partial_\beta \mathcal{L}(\beta^0 - \eta, \widehat{\phi}(\beta^0 - \eta)) = U + \overline{W}\sqrt{NT}\eta + o_P(1) + o_P(\sqrt{NT}\eta) = o_P(\sqrt{NT}\eta) + \overline{W}\sqrt{NT}\eta,$$

and therefore for sufficiently large $N, T$

$$\partial_\beta \mathcal{L}(\beta^0 + \eta, \widehat{\phi}(\beta^0 + \eta)) \leq 0 \leq \partial_\beta \mathcal{L}(\beta^0 - \eta, \widehat{\phi}(\beta^0 - \eta)).$$

Thus, since $\partial_\beta \mathcal{L}(\widehat{\beta}, \widehat{\phi}(\widehat{\beta})) = 0$, for sufficiently large $N, T$,

$$\partial_\beta \mathcal{L}(\beta^0 + \eta, \widehat{\phi}(\beta^0 + \eta)) \leq \partial_\beta \mathcal{L}(\widehat{\beta}, \widehat{\phi}(\widehat{\beta})) \leq \partial_\beta \mathcal{L}(\beta^0 - \eta, \widehat{\phi}(\beta^0 - \eta)).$$

The profile objective $\mathcal{L}(\beta, \widehat{\phi}(\beta))$ is strictly concave in $\beta$ because $\mathcal{L}(\beta, \phi)$ is strictly concave in $(\beta, \phi)$. Thus, $\partial_\beta \mathcal{L}(\beta, \widehat{\phi}(\beta))$ is strictly decreasing. The previous set of inequalities implies that for sufficiently large $N, T$

$$\beta^0 + \eta \geq \widehat{\beta} \geq \beta^0 - \eta.$$

We conclude that $\|\widehat{\beta} - \beta^0\| \leq \eta = \mathcal{O}_P((NT)^{-1/4})$. This concludes the proof for $\dim(\beta) = 1$.

To generalize the proof to $\dim(\beta) > 1$ we define $\beta_\pm = \beta^0 \pm \eta \frac{\widehat{\beta} - \beta^0}{\|\widehat{\beta} - \beta^0\|}$. Let $\langle \beta_-, \beta_+ \rangle = \{r\beta_- + (1 - r)\beta_+ \mid r \in [0, 1]\}$ be the line segment between $\beta_-$ and $\beta_+$. By restricting attention to values $\beta \in \langle \beta_-, \beta_+ \rangle$ we can repeat the above argument for the case $\dim(\beta) = 1$ and thus show that $\widehat{\beta} \in \langle \beta_-, \beta_+ \rangle$, which implies $\|\widehat{\beta} - \beta^0\| \leq \eta = \mathcal{O}_P((NT)^{-1/4})$. ■

### B.1.3 Proof of Theorem B.4

**Proof of Theorem B.4.** A Taylor expansion of $\Delta(\beta, \phi)$ around $(\beta^0, \phi^0)$ yields

$$\Delta(\beta, \phi) = \Delta + [\partial_{\beta'}\Delta](\beta - \beta^0) + [\partial_{\phi'}\Delta](\phi - \phi^0) + \tfrac{1}{2}(\phi - \phi^0)'[\partial_{\phi\phi'}\Delta](\phi - \phi^0) + R_1^\Delta(\beta, \phi),$$

with remainder term

$$R_1^\Delta(\beta, \phi) = \tfrac{1}{2}(\beta - \beta^0)'[\partial_{\beta\beta'}\Delta(\bar{\beta}, \phi)](\beta - \beta^0) + (\beta - \beta^0)'[\partial_{\beta\phi'}\Delta(\beta^0, \tilde{\phi})](\phi - \phi^0)$$
$$+ \tfrac{1}{6}\sum_g (\phi - \phi^0)'[\partial_{\phi\phi'\phi_g}\Delta(\beta^0, \bar{\phi})](\phi - \phi^0)[\phi - \phi^0]_g,$$

where $\bar{\beta}$ is between $\beta$ and $\beta^0$, and $\tilde{\phi}$ and $\bar{\phi}$ are between $\phi$ and $\phi^0$.

By assumption, $\|\widehat{\beta} - \beta^0\| = o_P((NT)^{-1/4})$, and by the expansion of $\widehat{\phi} = \widehat{\phi}(\widehat{\beta})$ in Theorem B.1,

$$\|\widehat{\phi} - \phi^0\|_q \leq \left\|\mathcal{H}^{-1}\right\|_q \|\mathcal{S}\|_q + \left\|\mathcal{H}^{-1}\right\|_q \|\partial_{\phi\beta'}\mathcal{L}\|_q \left\|\widehat{\beta} - \beta^0\right\|_q + \tfrac{1}{2}\left\|\mathcal{H}^{-1}\right\|_q^3 \|\partial_{\phi\phi\phi}\mathcal{L}\|_q \|\mathcal{S}\|_q^2 + \left\|R^\phi(\widehat{\beta})\right\|_q$$
$$= \mathcal{O}_P((NT)^{-1/4 + 1/(2q)}).$$

Thus, for $\widehat{R}_1^{\Delta} := R_1^{\Delta}(\widehat{\beta}, \widehat{\phi})$,

$$
\begin{aligned}
\left|\widehat{R}_1^{\Delta}\right| &\leq \tfrac{1}{2}\|\widehat{\beta} - \beta^0\|^2 \sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \|\partial_{\beta\beta'}\Delta(\beta, \phi)\| \\
&\quad + (NT)^{1/2-1/q}\|\widehat{\beta} - \beta^0\| \|\widehat{\phi} - \phi^0\|_q \sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \|\partial_{\beta\phi'}\Delta(\beta, \phi)\|_q \\
&\quad + \tfrac{1}{6}(NT)^{1/2-1/q}\|\widehat{\phi} - \phi^0\|_q^3 \sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \|\partial_{\phi\phi\phi}\Delta(\beta, \phi)\|_q \\
&= o_P(1/\sqrt{NT}).
\end{aligned}
$$

Again by the expansion of $\widehat{\phi} = \widehat{\phi}(\widehat{\beta})$ from Theorem B.1,

$$
\begin{aligned}
\widehat{\delta} - \delta &= \Delta(\widehat{\beta}, \widehat{\phi}) - \Delta = \left(\partial_{\beta'}\Delta + [\partial_\phi \Delta]' \mathcal{H}^{-1} [\partial_{\phi\beta'}\mathcal{L}]\right)(\widehat{\beta} - \beta^0) \\
&\quad + [\partial_\phi \Delta]' \mathcal{H}^{-1} \left( \mathcal{S} + \tfrac{1}{2} \sum_{g=1}^{\dim \phi} [\partial_{\phi\phi'\phi_g}\mathcal{L}] \mathcal{H}^{-1} \mathcal{S} [\mathcal{H}^{-1}\mathcal{S}]_g \right) + \tfrac{1}{2}\, \mathcal{S}' \mathcal{H}^{-1} [\partial_{\phi\phi'}\Delta] \mathcal{H}^{-1} \mathcal{S} + R_2^{\Delta},
\end{aligned}
\tag{B.8}
$$

where

$$
\begin{aligned}
\left|R_2^{\Delta}\right| &= \left| R_1^{\Delta} + [\partial_\phi \Delta]' R^{\phi}(\widehat{\beta}) + \tfrac{1}{2}(\widehat{\phi} - \phi^0 + \mathcal{H}^{-1}\mathcal{S})' [\partial_{\phi\phi'}\Delta] (\widehat{\phi} - \phi^0 - \mathcal{H}^{-1}\mathcal{S}) \right| \\
&\leq \left|R_1^{\Delta}\right| + (NT)^{1/2-1/q} \left\|\partial_\phi \Delta\right\|_q \left\| R^{\phi}(\widehat{\beta}) \right\|_q \\
&\quad + \tfrac{1}{2}(NT)^{1/2-1/q} \left\| \widehat{\phi} - \phi^0 + \mathcal{H}^{-1}\mathcal{S} \right\|_q \left\|\partial_{\phi\phi'}\Delta\right\|_q \left\| \widehat{\phi} - \phi^0 - \mathcal{H}^{-1}\mathcal{S} \right\|_q \\
&= o_P(1/\sqrt{NT}),
\end{aligned}
$$

that uses $\left\| \widehat{\phi} - \phi^0 - \mathcal{H}^{-1}\mathcal{S} \right\|_q = \mathcal{O}_P\left((NT)^{-1/2+1/q+\epsilon}\right)$. From equation (B.8), the terms of the expansion for $\widehat{\delta} - \delta$ are analogous to the terms of the expansion for the score in Theorem B.1, with $\Delta(\beta, \phi)$ taking the role of $\frac{1}{\sqrt{NT}} \partial_{\beta_k} \mathcal{L}(\beta, \phi)$. ■

# C Proofs of Section 4

## C.1 Application of General Expansion to Panel Estimators

We now apply the general expansion of appendix B to the panel fixed effects estimators considered in the main text. For the objective function specified in (2.1) and (4.1), the incidental parameter score evaluated at the true parameter value is

$$
\mathcal{S} = \begin{pmatrix} \left[ \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \partial_\pi \ell_{it} \right]_{i=1,\ldots,N} \\ \left[ \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \partial_\pi \ell_{it} \right]_{t=1,\ldots,T} \end{pmatrix}.
$$

The penalty term in the objective function does not contribute to $\mathcal{S}$, because at the true parameter value $v'\phi^0 = 0$. The corresponding expected incidental parameter Hessian $\overline{\mathcal{H}}$ is given in (4.2). Section D.4 discusses the structure of $\overline{\mathcal{H}}$ and $\overline{\mathcal{H}}^{-1}$ in more detail. Define

$$
\Lambda_{it} := -\frac{1}{\sqrt{NT}} \sum_{j=1}^{N} \sum_{\tau=1}^{T} \left( \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)ij} + \overline{\mathcal{H}}^{-1}_{(\gamma\alpha)tj} + \overline{\mathcal{H}}^{-1}_{(\alpha\gamma)i\tau} + \overline{\mathcal{H}}^{-1}_{(\gamma\gamma)t\tau} \right) \partial_\pi \ell_{j\tau},
\tag{C.1}
$$

and the operator $D_\beta \Delta_{it} := \partial_\beta \Delta_{it} - \partial_\pi \Delta_{it} \Xi_{it}$, which are similar to $\Xi_{it}$ and $D_\beta \ell_{it}$ in equation (4.3).

The following theorem shows that Assumption 4.1 and Assumption 4.2 for the panel model are sufficient for Assumption B.1 and Assumption B.2 for the general expansion, and particularizes the terms of the expansion to the panel estimators.

**Theorem C.1.** *Consider an estimator with objective function given by* (2.1) *and* (4.1). *Let Assumption 4.1 be satisfied and suppose that the limit $\overline{W}_\infty$ defined in Theorem 4.1 exists and is positive definite. Let $q = 8$, $\epsilon = 1/(16 + 2\nu)$, $r_{\beta,NT} = \log(NT)(NT)^{-1/8}$ and $r_{\phi,NT} = (NT)^{-1/16}$. Then,*

*(i) Assumption B.1 holds and $\|\widehat{\beta} - \beta^0\| = \mathcal{O}_P((NT)^{-1/4})$.*

*(ii) The approximate Hessian and the terms of the score defined in Theorem B.1 can be written as*

$$\overline{W} = -\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}_\phi \left( \partial_{\beta\beta'} \ell_{it} - \partial_{\pi^2} \ell_{it} \Xi_{it} \Xi'_{it} \right),$$

$$U^{(0)} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} D_\beta \ell_{it},$$

$$U^{(1)} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ -\Lambda_{it} \left[ D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi(D_{\beta\pi} \ell_{it}) \right] + \frac{1}{2} \Lambda_{it}^2 \, \mathbb{E}_\phi(D_{\beta\pi^2} \ell_{it}) \right\}.$$

*(iii) In addition, let Assumption 4.2 hold. Then, Assumption B.2 is satisfied for the partial effects defined in* (2.2). *By Theorem B.4,*

$$\sqrt{NT} \left( \widehat{\delta} - \delta \right) = V_\Delta^{(0)} + V_\Delta^{(1)} + o_P(1),$$

*where*

$$V_\Delta^{(0)} = \left[ \frac{1}{NT} \sum_{i,t} \mathbb{E}_\phi(D_\beta \Delta_{it}) \right]' \overline{W}_\infty^{-1} U^{(0)} - \frac{1}{\sqrt{NT}} \sum_{i,t} \mathbb{E}_\phi(\Psi_{it}) \partial_\pi \ell_{it},$$

$$V_\Delta^{(1)} = \left[ \frac{1}{NT} \sum_{i,t} \mathbb{E}_\phi(D_\beta \Delta_{it}) \right]' \overline{W}_\infty^{-1} U^{(1)} + \frac{1}{\sqrt{NT}} \sum_{i,t} \Lambda_{it} \left[ \Psi_{it} \partial_{\pi^2} \ell_{it} - \mathbb{E}_\phi(\Psi_{it}) \mathbb{E}_\phi(\partial_{\pi^2} \ell_{it}) \right]$$
$$+ \frac{1}{2\sqrt{NT}} \sum_{i,t} \Lambda_{it}^2 \left[ \mathbb{E}_\phi(\partial_{\pi^2} \Delta_{it}) - \mathbb{E}_\phi(\partial_{\pi^3} \ell_{it}) \mathbb{E}_\phi(\Psi_{it}) \right].$$

**Proof of Theorem C.1, Part** $(i)$**.** Assumption B.1$(i)$ is satisfied because $\lim_{N,T\to\infty} \frac{\dim \phi}{\sqrt{NT}} = \lim_{N,T\to\infty} \frac{N+T}{\sqrt{NT}} = \kappa + \kappa^{-1}$.

Assumption B.1$(ii)$ is satisfied because $\ell_{it}(\beta, \pi)$ and $(v'\phi)^2$ are four times continuously differentiable and the same is true for $\mathcal{L}(\beta, \phi)$.

Let $\overline{\mathcal{D}} = \operatorname{diag}\left( \overline{\mathcal{H}}^*_{(\alpha\alpha)}, \overline{\mathcal{H}}^*_{(\gamma\gamma)} \right)$. Then, $\left\| \overline{\mathcal{D}}^{-1} \right\|_\infty = \mathcal{O}_P(1)$ by Assumption 4.1$(v)$. By the properties of the matrix norms and Lemma D.8, $\left\| \overline{\mathcal{H}}^{-1} - \overline{\mathcal{D}}^{-1} \right\|_\infty \leq (N+T) \left\| \overline{\mathcal{H}}^{-1} - \overline{\mathcal{D}}^{-1} \right\|_{\max} = \mathcal{O}_P(1)$. Thus, $\left\| \overline{\mathcal{H}}^{-1} \right\|_q \leq \left\| \overline{\mathcal{H}}^{-1} \right\|_\infty \leq \left\| \overline{\mathcal{D}}^{-1} \right\|_\infty + \left\| \overline{\mathcal{H}}^{-1} - \overline{\mathcal{D}}^{-1} \right\|_\infty = \mathcal{O}_P(1)$ by Lemma D.4 and the triangle inequality. We conclude that Assumption B.1$(iv)$ holds.

We now show that the assumptions of Lemma D.7 are satisfied:

(i) By Lemma D.2, $\chi_i = \frac{1}{\sqrt{T}} \sum_t \partial_{\beta_k} \ell_{it}$ satisfies $\mathbb{E}_\phi(\chi_i^2) \leq B$. Thus, by independence across $i$

$$\mathbb{E}_\phi \left[ \left( \frac{1}{\sqrt{NT}} \sum_{i,t} \partial_{\beta_k} \ell_{it} \right)^2 \right] = \mathbb{E}_\phi \left[ \left( \frac{1}{\sqrt{N}} \sum_i \chi_i \right)^2 \right] = \frac{1}{N} \sum_i \mathbb{E}_\phi \chi_i^2 \leq B,$$

and therefore $\frac{1}{\sqrt{NT}} \sum_{i,t} \partial_{\beta_k} \ell_{it} = \mathcal{O}_P(1)$. Analogously, $\frac{1}{NT} \sum_{i,t} \{\partial_{\beta_k \beta_l} \ell_{it} - \mathbb{E}_\phi [\partial_{\beta_k \beta_l} \ell_{it}]\} = \mathcal{O}_P(1/\sqrt{NT}) = o_P(1)$. Next,

$$\begin{aligned}
&\mathbb{E}_\phi \left( \sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \frac{1}{NT} \sum_{i,t} \partial_{\beta_k \beta_l \beta_m} \ell_{it}(\beta, \pi_{it}) \right)^2 \\
&\leq \mathbb{E}_\phi \left( \sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \frac{1}{NT} \sum_{i,t} |\partial_{\beta_k \beta_l \beta_m} \ell_{it}(\beta, \pi_{it})| \right)^2 \leq \mathbb{E}_\phi \left( \frac{1}{NT} \sum_{i,t} M(Z_{it}) \right)^2 \\
&\leq \mathbb{E}_\phi \frac{1}{NT} \sum_{i,t} M(Z_{it})^2 = \frac{1}{NT} \sum_{i,t} \mathbb{E}_\phi M(Z_{it})^2 = \mathcal{O}_P(1),
\end{aligned}$$

and therefore $\sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \frac{1}{NT} \sum_{i,t} \partial_{\beta_k \beta_l \beta_m} \ell_{it}(\beta, \pi_{it}) = \mathcal{O}_P(1)$. A similar argument gives $\frac{1}{NT} \sum_{i,t} \partial_{\beta_k \beta_l} \ell_{it} = \mathcal{O}_P(1)$.

(ii) For $\xi_{it}(\beta, \phi) = \partial_{\beta_k \pi} \ell_{it}(\beta, \pi_{it})$ or $\xi_{it}(\beta, \phi) = \partial_{\beta_k \beta_l \pi} \ell_{it}(\beta, \pi_{it})$,

$$\begin{aligned}
&\mathbb{E}_\phi \left[ \sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \frac{1}{T} \sum_t \left| \frac{1}{N} \sum_i \xi_{it}(\beta, \phi) \right|^q \right] \\
&\leq \mathbb{E}_\phi \left[ \sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \frac{1}{T} \sum_t \left( \frac{1}{N} \sum_i |\xi_{it}(\beta, \phi)| \right)^q \right] \\
&\leq \mathbb{E}_\phi \left[ \frac{1}{T} \sum_t \left( \frac{1}{N} \sum_i M(Z_{it}) \right)^q \right] \leq \mathbb{E}_\phi \left[ \frac{1}{T} \sum_t \frac{1}{N} \sum_i M(Z_{it})^q \right] \\
&= \frac{1}{T} \sum_t \frac{1}{N} \sum_i \mathbb{E}_\phi M(Z_{it})^q = \mathcal{O}_P(1),
\end{aligned}$$

i.e. $\sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \frac{1}{T} \sum_t \left| \frac{1}{N} \sum_i \xi_{it}(\beta, \phi) \right|^q = \mathcal{O}_P(1)$. Analogously, it follows that $\sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \frac{1}{N} \sum_i \left| \frac{1}{T} \sum_t \xi_{it}(\beta, \phi) \right|^q = \mathcal{O}_P(1)$.

(iii) For $\xi_{it}(\beta, \phi) = \partial_{\pi^r} \ell_{it}(\beta, \pi_{it})$, with $r \in \{3, 4\}$, or $\xi_{it}(\beta, \phi) = \partial_{\beta_k \pi^r} \ell_{it}(\beta, \pi_{it})$, with $r \in \{2, 3\}$, or $\xi_{it}(\beta, \phi) = \partial_{\beta_k \beta_l \pi^2} \ell_{it}(\beta, \pi_{it})$,

$$\begin{aligned}
&\mathbb{E}_\phi \left[ \left( \sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \max_i \frac{1}{T} \sum_t |\xi_{it}(\beta, \phi)| \right)^{(8+\nu)} \right] \\
&= \mathbb{E}_\phi \left[ \max_i \left( \sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \frac{1}{T} \sum_t |\xi_{it}(\beta, \phi)| \right)^{(8+\nu)} \right] \\
&\leq \mathbb{E}_\phi \left[ \sum_i \left( \sup_{\beta \in \mathcal{B}(r_\beta, \beta^0)} \sup_{\phi \in \mathcal{B}_q(r_\phi, \phi^0)} \frac{1}{T} \sum_t |\xi_{it}(\beta, \phi)| \right)^{(8+\nu)} \right] \leq \mathbb{E}_\phi \left[ \sum_i \left( \frac{1}{T} \sum_t M(Z_{it}) \right)^{(8+\nu)} \right] \\
&\leq \mathbb{E}_\phi \left[ \sum_i \frac{1}{T} \sum_t M(Z_{it})^{(8+\nu)} \right] = \sum_i \frac{1}{T} \sum_t \mathbb{E}_\phi M(Z_{it})^{(8+\nu)} = \mathcal{O}_P(N).
\end{aligned}$$

Thus, $\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\max_i \frac{1}{T}\sum_t |\xi_{it}(\beta,\phi)| = \mathcal{O}_P\left(N^{1/(8+\nu)}\right) = \mathcal{O}_P\left(N^{2\epsilon}\right)$. Analogously, it follows that $\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\max_t \frac{1}{N}\sum_i |\xi_{it}(\beta,\phi)| = \mathcal{O}_P\left(N^{2\epsilon}\right)$.

(iv) Let $\chi_t = \frac{1}{\sqrt{N}}\sum_i \partial_\pi \ell_{it}$. By cross-sectional independence and $\mathbb{E}_\phi(\partial_\pi\ell_{it})^8 \leq \mathbb{E}_\phi M(Z_{it})^8 = \mathcal{O}_P(1)$, $\mathbb{E}_\phi\chi_t^8 = \mathcal{O}_P(1)$ uniformly over $t$. Thus, $\mathbb{E}_\phi\frac{1}{T}\sum_t\chi_t^8 = \mathcal{O}_P(1)$ and therefore $\frac{1}{T}\sum_t\left|\frac{1}{\sqrt{N}}\sum_i\partial_\pi\ell_{it}\right|^q = \mathcal{O}_P(1)$, with $q=8$.

Let $\chi_i = \frac{1}{\sqrt{T}}\sum_t\partial_\pi\ell_{it}(\beta^0,\pi^0_{it})$. By Lemma D.2 and $\mathbb{E}_\phi(\partial_\pi\ell_{it})^{8+\nu} \leq \mathbb{E}_\phi M(Z_{it})^{8+\nu} = \mathcal{O}_P(1)$, $\mathbb{E}_\phi\chi_i^8 = \mathcal{O}_P(1)$ uniformly over $i$. Here we use $\mu > 4/[1-8/(8+\nu)] = 4(8+\nu)/\nu$ that is imposed in Assumption B.1. Thus, $\mathbb{E}_\phi\frac{1}{N}\sum_i\chi_i^8 = \mathcal{O}_P(1)$ and therefore $\frac{1}{N}\sum_i\left|\frac{1}{\sqrt{T}}\sum_t\partial_\pi\ell_{it}\right|^q = \mathcal{O}_P(1)$, with $q=8$.

The proofs for $\frac{1}{T}\sum_t\left|\frac{1}{\sqrt{N}}\sum_i\partial_{\beta_k\pi}\ell_{it} - \mathbb{E}_\phi\left[\partial_{\beta_k\pi}\ell_{it}\right]\right|^2 = \mathcal{O}_P(1)$ and $\frac{1}{N}\sum_i\left|\frac{1}{\sqrt{T}}\sum_t\partial_{\beta_k\pi}\ell_{it} - \mathbb{E}_\phi\left[\partial_{\beta_k\pi}\ell_{it}\right]\right|^2 = \mathcal{O}_P(1)$ are analogous.

(v) It follows by the independence of $\{(\ell_{i1},\ldots,\ell_{iT}) : 1 \leq i \leq N\}$ across $i$, conditional on $\phi$, in Assumption B.1$(ii)$.

(vi) Let $\xi_{it} = \partial_{\pi^r}\ell_{it}(\beta^0,\pi^0_{it}) - \mathbb{E}_\phi\left[\partial_{\pi^r}\ell_{it}\right]$, with $r\in\{2,3\}$, or $\xi_{it} = \partial_{\beta_k\pi^2}\ell_{it}(\beta^0,\pi^0_{it}) - \mathbb{E}_\phi\left[\partial_{\beta_k\pi^2}\ell_{it}\right]$. For $\tilde\nu = \nu$, $\max_i\mathbb{E}_\phi\left[\xi_{it}^{8+\tilde\nu}\right] = \mathcal{O}_P(1)$ by assumption. By Lemma D.1,

$$
\begin{aligned}
\left|\sum_s \mathbb{E}_\phi\left[\xi_{it}\xi_{is}\right]\right| &= \sum_s\left|\mathrm{Cov}_\phi\left(\xi_{it},\,\xi_{is}\right)\right| \\
&\leq \sum_s\left[8\,a(|t-s|)\right]^{1-2/(8+\nu)}\left[\mathbb{E}_\phi|\xi_t|^{8+\nu}\right]^{1/(8+\nu)}\left[\mathbb{E}_\phi|\xi_s|^{8+\nu}\right]^{1/(8+\nu)} \\
&= \tilde C\sum_{m=1}^\infty m^{-\mu[1-2/(8+\nu)]} \leq \tilde C\sum_{m=1}^\infty m^{-4} = \tilde C\pi^4/90,
\end{aligned}
$$

where $\tilde C$ is a constant. Here we use that $\mu > 4(8+\nu)/\nu$ implies $\mu[1-2/(8+\nu) > 4$. We thus have shown $\max_i\max_t\sum_s\mathbb{E}_\phi\left[\xi_{it}\xi_{js}\right] \leq \tilde C\pi^4/90 =: C$.

Analogous to the proof of part (iv), we can use Lemma D.2 to obtain $\max_i\mathbb{E}_\phi\left\{\left[\frac{1}{\sqrt{T}}\sum_t\xi_{it}\right]^8\right\} \leq C$, and independence across $i$ to obtain $\max_t\mathbb{E}_\phi\left\{\left[\frac{1}{\sqrt{N}}\sum_i\xi_{it}\right]^8\right\} \leq C$. Similarly, by Lemma D.2

$$
\max_{i,j}\mathbb{E}_\phi\left\{\left[\frac{1}{\sqrt{T}}\sum_t\left[\xi_{it}\xi_{jt} - \mathbb{E}_\phi\left(\xi_{it}\xi_{jt}\right)\right]\right]^4\right\} \leq C,
$$

which requires $\mu > 2/[1-4/(4+\nu/2)]$, which is implied by the assumption that $\mu > 4(8+\nu)/\nu$.

(vii) We have already shown that $\left\|\overline{\mathcal{H}}^{-1}\right\|_q = \mathcal{O}_P(1)$.

Therefore, we can apply Lemma D.7, which shows that Assumption B.1$(v)$ and $(vi)$ hold. We have already shown that Assumption B.1$(i)$, $(ii)$, $(iv)$, $(v)$ and $(vi)$ hold. One can also check that $(NT)^{-1/4+1/(2q)} = o_P(r_\phi)$ and $(NT)^{1/(2q)}r_\beta = o_P(r_\phi)$ are satisfied. In addition, $\mathcal{L}(\beta,\phi)$ is strictly concave. We can therefore invoke Theorem B.3 to show that Assumption B.1$(iii)$ holds and that $\|\widehat\beta - \beta^0\| = \mathcal{O}_P((NT)^{-1/4})$. ■

**Proof of Theorem C.1, Part** $(ii)$**.** For any $N\times T$ matrix $A$ we define the $N\times T$ matrix $\mathbb{P}A$ as follows

$$
(\mathbb{P}A)_{it} = \alpha_i^* + \gamma_t^*, \qquad (\alpha^*,\gamma^*) \in \operatorname*{argmin}_{\alpha,\gamma}\sum_{i,t}\mathbb{E}_\phi(-\partial_{\pi^2}\ell_{it})\left(A_{it}-\alpha_i-\gamma_t\right)^2. \tag{C.2}
$$

Here, the minimization is over $\alpha \in \mathbb{R}^N$ and $\gamma \in \mathbb{R}^T$. The operator $\mathbb{P}$ is a linear projection, i.e. we have $\mathbb{P}\mathbb{P} = \mathbb{P}$. It is also convenient to define

$$\widetilde{\mathbb{P}} A = \mathbb{P} \tilde{A}, \qquad \text{where} \qquad \tilde{A}_{it} = \frac{A_{it}}{\mathbb{E}_\phi(-\partial_{\pi^2}\ell_{it})}. \tag{C.3}$$

$\widetilde{\mathbb{P}}$ is a linear operator, but not a projection. Note that $\Lambda$ and $\Xi$ defined in (C.1) and (4.3) can be written as $\Lambda = \widetilde{\mathbb{P}} A$ and $\Xi_k = \widetilde{\mathbb{P}} B_k$, where $A_{it} = -\partial_\pi \ell_{it}$ and $B_{k,it} = -\mathbb{E}_\phi(\partial_{\beta_k \pi}\ell_{it})$, for $k = 1, \ldots, \dim \beta$.[14]

By Lemma D.11(*ii*),

$$\overline{W} = -\frac{1}{\sqrt{NT}} \left(\partial_{\beta\beta'}\overline{\mathcal{L}} + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\, \overline{\mathcal{H}}^{-1}\, [\partial_{\phi\beta'}\overline{\mathcal{L}}]\right) = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbb{E}_\phi\left(\partial_{\beta\beta'}\ell_{it}\right) + \mathbb{E}_\phi\left(-\partial_{\pi^2}\ell_{it}\right)\Xi_{it}\Xi_{it}'\right].$$

By Lemma D.11(*i*),

$$U^{(0)} = \partial_\beta \mathcal{L} + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\, \overline{\mathcal{H}}^{-1}\mathcal{S} = \frac{1}{\sqrt{NT}}\sum_{i,t}\left(\partial_\beta \ell_{it} - \Xi_{it}\,\partial_\pi \ell_{it}\right) = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T} D_\beta \ell_{it}.$$

We decompose $U^{(1)} = U^{(1a)} + U^{(1b)}$, with

$$\begin{aligned} U^{(1a)} &= [\partial_{\beta\phi'}\widetilde{\mathcal{L}}]\overline{\mathcal{H}}^{-1}\mathcal{S} - [\partial_{\beta\phi'}\overline{\mathcal{L}}]\, \overline{\mathcal{H}}^{-1}\, \widetilde{\mathcal{H}}\, \overline{\mathcal{H}}^{-1}\, \mathcal{S}, \\ U^{(1b)} &= \sum_{g=1}^{\dim\phi}\left(\partial_{\beta\phi'\phi_g}\overline{\mathcal{L}} + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\, \overline{\mathcal{H}}^{-1}[\partial_{\phi\phi'\phi_g}\overline{\mathcal{L}}]\right)\overline{\mathcal{H}}^{-1}\mathcal{S}[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g/2. \end{aligned}$$

By Lemma D.11(*i*) and (*iii*),

$$U^{(1a)} = -\frac{1}{\sqrt{NT}}\sum_{i,t}\Lambda_{it}\left(\partial_{\beta\pi}\tilde{\ell}_{it} + \Xi_{it}\,\partial_{\pi^2}\tilde{\ell}_{it}\right) = -\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\Lambda_{it}\,\left[D_{\beta\pi}\ell_{it} - \mathbb{E}_\phi(D_{\beta\pi}\ell_{it})\right],$$

and

$$U^{(1b)} = \frac{1}{2\sqrt{NT}}\sum_{i,t}\Lambda_{it}^2\left[\mathbb{E}_\phi(\partial_{\beta\pi^2}\ell_{it}) + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\, \overline{\mathcal{H}}^{-1}\mathbb{E}_\phi(\partial_\phi\partial_{\pi^2}\ell_{it})\right],$$

where for each $i, t$, $\partial_\phi\partial_{\pi^2}\ell_{it}$ is a $\dim\phi$-vector, which can be written as $\partial_\phi\partial_{\pi^2}\ell_{it} = \binom{A 1_T}{A' 1_N}$ for an $N \times T$ matrix $A$ with elements $A_{j\tau} = \partial_{\pi^3}\ell_{j\tau}$ if $j = i$ and $\tau = t$, and $A_{j\tau} = 0$ otherwise. Thus, Lemma D.11(*i*) gives $[\partial_{\beta\phi'}\overline{\mathcal{L}}]\, \overline{\mathcal{H}}^{-1}\partial_\phi\partial_{\pi^2}\ell_{it} = -\sum_{j,\tau}\Xi_{j\tau}1(i=j)1(t=\tau)\partial_{\pi^3}\ell_{it} = -\Xi_{it}\partial_{\pi^3}\ell_{it}$. Therefore

$$U^{(1b)} = \frac{1}{2\sqrt{NT}}\sum_{i,t}\Lambda_{it}^2\mathbb{E}_\phi\left(\partial_{\beta\pi^2}\ell_{it} - \Xi_{it}\partial_{\pi^3}\ell_{it}\right) = \frac{1}{2\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\Lambda_{it}^2\,\mathbb{E}_\phi(D_{\beta\pi^2}\ell_{it}).$$

∎

**Proof of Theorem C.1, Part** (*iii*)**.** Showing that Assumption B.2 is satisfied is analogous to the proof of Lemma D.7 and of part (*ii*) of this Theorem.

[14] $B_k$ and $\Xi_k$ are $N \times T$ matrices with entries $B_{k,it}$ and $\Xi_{k,it}$, respectively, while $B_{it}$ and $\Xi_{it}$ are $\dim\beta$-vectors with entries $B_{k,it}$ and $\Xi_{k,it}$.

In the proof of Theorem 4.1 we show that Assumption 4.1 implies that $U = \mathcal{O}_P(1)$. This fact together with part $(i)$ of this theorem show that Corollary B.2 is applicable, so that $\sqrt{NT}(\widehat{\beta} - \beta^0) = \overline{W}_\infty^{-1} U + o_P(1) = \mathcal{O}_P(1)$, and we can apply Theorem B.4.

By Lemma D.11 and the result for $\sqrt{NT}(\widehat{\beta} - \beta^0)$,

$$\sqrt{NT}\left[\partial_{\beta'}\overline{\Delta} + (\partial_{\phi'}\overline{\Delta})\overline{\mathcal{H}}^{-1}(\partial_{\phi\beta'}\overline{\mathcal{L}})\right](\widehat{\beta} - \beta^0) = \left[\frac{1}{NT}\sum_{i,t}\mathbb{E}_\phi(D_\beta\Delta_{it})\right]' \overline{W}_\infty^{-1}\left(U^{(0)} + U^{(1)}\right) + o_P(1). \tag{C.4}$$

We apply Lemma D.11 to $U_\Delta^{(0)}$ and $U_\Delta^{(1)}$ defined in Theorem B.4 to give

$$\begin{aligned}
\sqrt{NT}\, U_\Delta^{(0)} &= -\frac{1}{\sqrt{NT}}\sum_{i,t}\mathbb{E}_\phi(\Psi_{it})\partial_\pi\ell_{it}, \\
\sqrt{NT}\, U_\Delta^{(1)} &= \frac{1}{\sqrt{NT}}\sum_{i,t}\Lambda_{it}\left[\Psi_{it}\partial_{\pi^2}\ell_{it} - \mathbb{E}_\phi(\Psi_{it})\mathbb{E}_\phi(\partial_{\pi^2}\ell_{it})\right] \\
&\quad + \frac{1}{2\sqrt{NT}}\sum_{i,t}\Lambda_{it}^2\left[\mathbb{E}_\phi(\partial_{\pi^2}\Delta_{it}) - \mathbb{E}_\phi(\partial_{\pi^3}\ell_{it})\mathbb{E}_\phi(\Psi_{it})\right].
\end{aligned} \tag{C.5}$$

The derivation of (C.4) and (C.5) is analogous to the proof of the part $(ii)$ of the Theorem. Combining Theorem B.4 with equations (C.4) and (C.5) gives the result. ■

## C.2 Proofs of Theorems 4.1 and 4.2

**Proof of Theorem 4.1.** # First, we want to show that $U^{(0)} \to_d \mathcal{N}(0,\, \overline{W}_\infty)$. In our likelihood setting, $\mathbb{E}_\phi\partial_\beta\mathcal{L} = 0$, $\mathbb{E}_\phi\mathcal{S} = 0$, and, by the Bartlett identities, $\mathbb{E}_\phi(\partial_\beta\mathcal{L}\partial_{\beta'}\mathcal{L}) = -\frac{1}{\sqrt{NT}}\partial_{\beta\beta'}\overline{\mathcal{L}}$, $\mathbb{E}_\phi(\partial_\beta\mathcal{L}\mathcal{S}') = -\frac{1}{\sqrt{NT}}\partial_{\beta\phi'}\overline{\mathcal{L}}$ and $\mathbb{E}_\phi(\mathcal{S}\mathcal{S}') = \frac{1}{\sqrt{NT}}\left(\overline{\mathcal{H}} - \frac{b}{\sqrt{NT}}vv'\right)$. Furthermore, $\mathcal{S}'v = 0$ and $\partial_{\beta\phi'}\overline{\mathcal{L}}v = 0$. Then, by definition of $\overline{W} = -\frac{1}{\sqrt{NT}}\left(\partial_{\beta\beta'}\overline{\mathcal{L}} + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}\,[\partial_{\phi\beta'}\overline{\mathcal{L}}]\right)$ and $U^{(0)} = \partial_\beta\mathcal{L} + [\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}\mathcal{S}$,

$$\mathbb{E}_\phi\left(U^{(0)}\right) = 0, \qquad \mathrm{Var}\left(U^{(0)}\right) = \overline{W},$$

which implies that $\lim_{N,T\to\infty}\mathrm{Var}\left(U^{(0)}\right) = \lim_{N,T\to\infty}\overline{W} = \overline{W}_\infty$. Moreover, part $(ii)$ of Theorem C.1 yields

$$U^{(0)} = \frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T D_\beta\ell_{it},$$

where $D_\beta\ell_{it} = \partial_\beta\ell_{it} - \partial_\pi\ell_{it}\Xi_{it}$ is a martingale difference sequence for each $i$ and independent across $i$, conditional on $\phi$. Thus, by Lemma D.3 and the Cramer-Wold device we conclude that

$$U^{(0)} \to_d \mathcal{N}\left[0,\ \lim_{N,T\to\infty}\mathrm{Var}\left(U^{(0)}\right)\right] \sim \mathcal{N}(0,\, \overline{W}_\infty).$$

# Next, we show that $U^{(1)} \to_P \kappa\overline{B}_\infty + \kappa^{-1}\overline{D}_\infty$. Part $(ii)$ of Theorem C.1 gives $U^{(1)} = U^{(1a)} + U^{(1b)}$, with

$$\begin{aligned}
U^{(1a)} &= -\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T\Lambda_{it}\left[D_{\beta\pi}\ell_{it} - \mathbb{E}_\phi(D_{\beta\pi}\ell_{it})\right], \\
U^{(1b)} &= \frac{1}{2\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T\Lambda_{it}^2\,\mathbb{E}_\phi(D_{\beta\pi^2}\ell_{it}).
\end{aligned}$$

Plugging-in the definition of $\Lambda_{it}$, we decompose $U^{(1a)} = U^{(1a,1)} + U^{(1a,2)} + U^{(1a,3)} + U^{(1a,4)}$, where

$$U^{(1a,1)} = \frac{1}{NT} \sum_{i,j} \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)ij} \left( \sum_{\tau} \partial_\pi \ell_{j\tau} \right) \sum_t \left[ D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi (D_{\beta\pi} \ell_{it}) \right],$$

$$U^{(1a,2)} = \frac{1}{NT} \sum_{j,t} \overline{\mathcal{H}}^{-1}_{(\gamma\alpha)tj} \left( \sum_{\tau} \partial_\pi \ell_{j\tau} \right) \sum_i \left[ D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi (D_{\beta\pi} \ell_{it}) \right],$$

$$U^{(1a,3)} = \frac{1}{NT} \sum_{i,\tau} \overline{\mathcal{H}}^{-1}_{(\alpha\gamma)i\tau} \left( \sum_{j} \partial_\pi \ell_{j\tau} \right) \sum_t \left[ D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi (D_{\beta\pi} \ell_{it}) \right],$$

$$U^{(1a,4)} = \frac{1}{NT} \sum_{t,\tau} \overline{\mathcal{H}}^{-1}_{(\gamma\gamma)t\tau} \left( \sum_{j} \partial_\pi \ell_{j\tau} \right) \sum_i \left[ D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi (D_{\beta\pi} \ell_{it}) \right].$$

By the Cauchy-Schwarz inequality applied to the sum over $t$ in $U^{(1a,2)}$,

$$\left( U^{(1a,2)} \right)^2 \leq \frac{1}{(NT)^2} \left[ \sum_t \left( \sum_{j,\tau} \overline{\mathcal{H}}^{-1}_{(\gamma\alpha)tj} \partial_\pi \ell_{j\tau} \right)^2 \right] \left[ \sum_t \left( \sum_i \left[ D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi (D_{\beta\pi} \ell_{it}) \right] \right)^2 \right].$$

By Lemma D.8, $\overline{\mathcal{H}}^{-1}_{(\gamma\alpha)tj} = \mathcal{O}_P(1/\sqrt{NT})$, uniformly over $t, j$. Using that both $\sqrt{NT}\,\overline{\mathcal{H}}^{-1}_{(\gamma\alpha)tj} \partial_\pi \ell_{j\tau}$ and $D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi (D_{\beta\pi} \ell_{it})$ are mean zero, independence across $i$ and Lemma D.2 across $t$, we obtain

$$\mathbb{E}_\phi \left( \frac{1}{\sqrt{NT}} \sum_{j,\tau} [\sqrt{NT}\,\overline{\mathcal{H}}^{-1}_{(\gamma\alpha)tj}] \partial_\pi \ell_{j\tau} \right)^2 = \mathcal{O}_P(1), \quad \mathbb{E}_\phi \left( \frac{1}{\sqrt{N}} \sum_i \left[ D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi (D_{\beta\pi} \ell_{it}) \right] \right)^2 = \mathcal{O}_P(1),$$

uniformly over $t$. Thus, $\sum_t \left( \sum_{j,\tau} \overline{\mathcal{H}}^{-1}_{(\gamma\alpha)tj} \partial_\pi \ell_{j\tau} \right)^2 = \mathcal{O}_P(T)$ and $\sum_t \left( \sum_i \left[ D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi (D_{\beta\pi} \ell_{it}) \right] \right)^2 = \mathcal{O}_P(NT)$. We conclude that

$$\left( U^{(1a,2)} \right)^2 = \frac{1}{(NT)^2} \mathcal{O}_P(T) \mathcal{O}_P(NT) = \mathcal{O}_P(1/N) = o_P(1),$$

and therefore that $U^{(1a,2)} = o_P(1)$. Analogously one can show that $U^{(1a,3)} = o_P(1)$.

By Lemma D.8, $\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)} = -\text{diag}\left[ \left( \frac{1}{\sqrt{NT}} \sum_{t=1}^T \mathbb{E}_\phi(\partial_{\pi^2} \ell_{it} \right)^{-1} \right] + \mathcal{O}_P(1/\sqrt{NT})$. Analogously to the proof of $U^{(1a,2)} = o_P(1)$, one can show that the $\mathcal{O}_P(1/\sqrt{NT})$ part of $\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}$ has an asymptotically negligible contribution to $U^{(1a,1)}$. Thus,

$$U^{(1a,1)} = -\frac{1}{\sqrt{NT}} \sum_i \underbrace{\frac{\left( \sum_\tau \partial_\pi \ell_{i\tau} \right) \sum_t \left[ D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi (D_{\beta\pi} \ell_{it}) \right]}{\sum_t \mathbb{E}_\phi (\partial_{\pi^2} \ell_{it})}}_{=: U_i^{(1a,1)}} + o_P(1).$$

Our assumptions guarantee that $\mathbb{E}_\phi \left[ \left( U_i^{(1a,1)} \right)^2 \right] = \mathcal{O}_P(1)$, uniformly over $i$. Note that both the denominator and the numerator of $U_i^{(1a,1)}$ are of order $T$. For the denominator this is obvious because of the sum over $T$. For the numerator there are two sums over $T$, but both $\partial_\pi \ell_{i\tau}$ and $D_{\beta\pi} \ell_{it} - \mathbb{E}_\phi (D_{\beta\pi} \ell_{it})$ are mean zero weakly correlated processes, so that their sums are of order $\sqrt{T}$. By the WLLN over $i$,

$N^{-1}\sum_i U_i^{(1a,1)} = N^{-1}\mathbb{E}_\phi U_i^{(1a,1)} + o_P(1)$, and therefore

$$U^{(1a,1)} = \underbrace{-\sqrt{\frac{N}{T}}\frac{1}{N}\sum_{i=1}^N \frac{\sum_{t=1}^T \sum_{\tau=t}^T \mathbb{E}_\phi\left(\partial_\pi \ell_{it} D_{\beta\pi}\ell_{i\tau}\right)}{\sum_{t=1}^T \mathbb{E}_\phi\left(\partial_{\pi^2}\ell_{it}\right)}}_{=:\sqrt{\frac{N}{T}}\overline{B}^{(1)}} + o_P(1).$$

Here, we use that $\mathbb{E}_\phi\left(\partial_\pi \ell_{it} D_{\beta\pi}\ell_{i\tau}\right) = 0$ for $t > \tau$. Analogously,

$$U^{(1a,4)} = \underbrace{-\sqrt{\frac{T}{N}}\frac{1}{T}\sum_{t=1}^T \frac{\sum_{i=1}^N \mathbb{E}_\phi\left(\partial_\pi \ell_{it} D_{\beta\pi}\ell_{it}\right)}{\sum_{i=1}^N \mathbb{E}_\phi\left(\partial_{\pi^2}\ell_{it}\right)}}_{=:\sqrt{\frac{T}{N}}\overline{D}^{(1)}} + o_P(1).$$

We conclude that $U^{(1a)} = \kappa\overline{B}^{(1)} + \kappa^{-1}\overline{D}^{(1)} + o_P(1)$.

Next, we analyze $U^{(1b)}$. We decompose $\Lambda_{it} = \Lambda_{it}^{(1)} + \Lambda_{it}^{(2)} + \Lambda_{it}^{(3)} + \Lambda_{it}^{(4)}$, where

$$\Lambda_{it}^{(1)} = -\frac{1}{\sqrt{NT}}\sum_{j=1}^N \overline{\mathcal{H}}_{(\alpha\alpha)ij}^{-1}\sum_{\tau=1}^T \partial_\pi \ell_{j\tau}, \qquad \Lambda_{it}^{(2)} = -\frac{1}{\sqrt{NT}}\sum_{j=1}^N \overline{\mathcal{H}}_{(\gamma\alpha)tj}^{-1}\sum_{\tau=1}^T \partial_\pi \ell_{j\tau},$$

$$\Lambda_{it}^{(3)} = -\frac{1}{\sqrt{NT}}\sum_{\tau=1}^T \overline{\mathcal{H}}_{(\alpha\gamma)i\tau}^{-1}\sum_{\tau=1}^T \partial_\pi \ell_{j\tau}, \qquad \Lambda_{it}^{(4)} = -\frac{1}{\sqrt{NT}}\sum_{\tau=1}^T \overline{\mathcal{H}}_{(\gamma\gamma)t\tau}^{-1}\sum_{\tau=1}^T \partial_\pi \ell_{j\tau}.$$

This decomposition of $\Lambda_{it}$ induces the following decomposition of $U^{(1b)}$

$$U^{(1b)} = \sum_{p,q=1}^4 U^{(1b,p,q)}, \qquad U^{(1b,p,q)} = \frac{1}{2\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T \Lambda_{it}^{(p)}\Lambda_{it}^{(q)}\mathbb{E}_\phi(D_{\beta\pi^2}\ell_{it}).$$

Due to the symmetry $U^{(1b,p,q)} = U^{(1b,q,p)}$, this decomposition has 10 distinct terms. Start with $U^{(1b,1,2)}$ noting that

$$U^{(1b,1,2)} = \frac{1}{\sqrt{NT}}\sum_{i=1}^N U_i^{(1b,1,2)},$$

$$U_i^{(1b,1,2)} = \frac{1}{2T}\sum_{t=1}^T \mathbb{E}_\phi(D_{\beta\pi^2}\ell_{it})\frac{1}{N^2}\sum_{j_1,j_2=1}^N \left[NT\overline{\mathcal{H}}_{(\alpha\alpha)ij_1}^{-1}\overline{\mathcal{H}}_{(\gamma\alpha)tj_2}^{-1}\right]\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^T \partial_\pi \ell_{j_1\tau}\right)\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^T \partial_\pi \ell_{j_2\tau}\right).$$

By $\mathbb{E}_\phi(\partial_\pi \ell_{it}) = 0$, $\mathbb{E}_\phi(\partial_\pi \ell_{it}\partial_\pi \ell_{j\tau}) = 0$ for $(i,t) \neq (j,\tau)$, and the properties of the inverse expected Hessian from Lemma D.8, $\mathbb{E}_\phi\left[U_i^{(1b,1,2)}\right] = \mathcal{O}_P(1/N)$, uniformly over $i$, $\mathbb{E}_\phi\left[\left(U_i^{(1b,1,2)}\right)^2\right] = \mathcal{O}_P(1)$, uniformly over $i$, and $\mathbb{E}_\phi\left[U_i^{(1b,1,2)}U_j^{(1b,1,2)}\right] = \mathcal{O}_P(1/N)$, uniformly over $i \neq j$. This implies that $\mathbb{E}_\phi U^{(1b,1,2)} = \mathcal{O}_P(1/N)$ and $\mathbb{E}_\phi\left[\left(U^{(1b,1,2)} - \mathbb{E}_\phi U^{(1b,1,2)}\right)^2\right] = \mathcal{O}_P(1/\sqrt{N})$, and therefore $U^{(1b,1,2)} = o_P(1)$. By similar arguments one obtains $U^{(1b,p,q)} = o_P(1)$ for all combinations of $p,q = 1,2,3,4$, except for $p = q = 1$ and $p = q = 4$.

For $p = q = 1$,

$$U^{(1b,1,1)} = \frac{1}{\sqrt{NT}}\sum_{i=1}^N U_i^{(1b,1,1)},$$

$$U_i^{(1b,1,1)} = \frac{1}{2T}\sum_{t=1}^T \mathbb{E}_\phi(D_{\beta\pi^2}\ell_{it})\frac{1}{N^2}\sum_{j_1,j_2=1}^N \left[NT\overline{\mathcal{H}}_{(\alpha\alpha)ij_1}^{-1}\overline{\mathcal{H}}_{(\alpha\alpha)ij_2}^{-1}\right]\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^T \partial_\pi \ell_{j_1\tau}\right)\left(\frac{1}{\sqrt{T}}\sum_{\tau=1}^T \partial_\pi \ell_{j_2\tau}\right).$$

Analogous to the result for $U^{(1b,1,2)}$, $\mathbb{E}_\phi\left[\left(U^{(1b,1,1)} - \mathbb{E}_\phi U^{(1b,1,1)}\right)^2\right] = \mathcal{O}_P(1/\sqrt{N})$, and therefore $U^{(1b,1,1)} = \mathbb{E}_\phi U^{(1b,1,1)} + o(1)$. Furthermore,

$$\begin{aligned}\mathbb{E}_\phi U^{(1b,1,1)} &= \frac{1}{2\sqrt{NT}} \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} \mathbb{E}_\phi(D_{\beta\pi^2}\ell_{it}) \mathbb{E}_\phi\left[(\partial_\pi \ell_{it})^2\right]}{\left[\sum_{t=1}^{T} \mathbb{E}_\phi\left(\partial_{\pi^2}\ell_{it}\right)\right]^2} + o(1) \\ &= \underbrace{-\sqrt{\frac{N}{T}} \frac{1}{2N} \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} \mathbb{E}_\phi(D_{\beta\pi^2}\ell_{it})}{\sum_{t=1}^{T} \mathbb{E}_\phi\left(\partial_{\pi^2}\ell_{it}\right)}}_{=:\sqrt{\frac{N}{T}}\overline{B}^{(2)}} + o(1).\end{aligned}$$

Analogously,

$$U^{(1b,4,4)} = \mathbb{E}_\phi U^{(1b,4,4)} + o_P(1) = \underbrace{-\sqrt{\frac{T}{N}} \frac{1}{2T} \sum_{t=1}^{T} \frac{\sum_{i=1}^{N} \mathbb{E}_\phi(D_{\beta\pi^2}\ell_{it})}{\sum_{i=1}^{N} \mathbb{E}_\phi\left(\partial_{\pi^2}\ell_{it}\right)}}_{=:\sqrt{\frac{T}{N}}\overline{D}^{(2)}} + o(1).$$

We have thus shown that $U^{(1b)} = \kappa \overline{B}^{(2)} + \kappa^{-1}\overline{D}^{(2)} + o_P(1)$. Since $\overline{B}_\infty = \lim_{N,T\to\infty}[\overline{B}^{(1)} + \overline{B}^{(2)}]$ and $\overline{D}_\infty = \lim_{N,T\to\infty}[\overline{D}^{(1)} + \overline{D}^{(2)}]$ we thus conclude $U^{(1)} = \kappa \overline{B}_\infty + \kappa^{-1}\overline{D}_\infty + o_P(1)$.

# We have shown $U^{(0)} \to_d \mathcal{N}(0, \overline{W}_\infty)$, and $U^{(1)} \to_P \kappa\overline{B}_\infty + \kappa^{-1}\overline{D}_\infty$. Then, part $(ii)$ of Theorem C.1 yields $\sqrt{NT}(\widehat{\beta} - \beta^0) \to_d \overline{W}_\infty^{-1}\mathcal{N}(\kappa\overline{B}_\infty + \kappa^{-1}\overline{D}_\infty,\ \overline{W}_\infty)$. ■

**Proof of Theorem 4.2.** We consider the case of scalar $\Delta_{it}$ to simplify the notation. Decompose

$$r_{NT}(\widehat{\delta} - \delta^0_{NT} - \overline{B}^\delta_\infty/T - \overline{D}^\delta_\infty/N) = r_{NT}(\delta - \delta^0_{NT}) + \frac{r_{NT}}{\sqrt{NT}}\sqrt{NT}(\widehat{\delta} - \delta - \overline{B}^\delta_\infty/T - \overline{D}^\delta_\infty/N).$$

# Part (1): Limit of $\sqrt{NT}(\widehat{\delta} - \delta - \overline{B}^\delta_\infty/T - \overline{D}^\delta_\infty/N)$. An argument analogous to to the proof of Theorem 4.1 using Theorem C.1$(iii)$ yields

$$\sqrt{NT}(\widehat{\delta} - \delta) \to_d \mathcal{N}\left(\kappa\overline{B}^\delta_\infty + \kappa^{-1}\overline{D}^\delta_\infty, \overline{V}^{\delta(1)}_\infty\right),$$

where $\overline{V}^{\delta(1)}_\infty = \overline{\mathbb{E}}\left\{(NT)^{-1}\sum_{i,t}\mathbb{E}_\phi[\Gamma^2_{it}]\right\}$, for the expressions of $\overline{B}^\delta_\infty$, $\overline{D}^\delta_\infty$, and $\Gamma_{it}$ given in the statement of the theorem. Then, by Mann-Wald theorem

$$\sqrt{NT}(\widehat{\delta} - \delta - \overline{B}^\delta_\infty/T - \overline{D}^\delta_\infty/N) \to_d \mathcal{N}\left(0, \overline{V}^{\delta(1)}_\infty\right).$$

# Part (2): Limit of $r_{NT}(\delta - \delta^0_{NT})$. Here we show that $r_{NT}(\delta - \delta^0_{NT}) \to_d \mathcal{N}(0, \overline{V}^{\delta(2)}_\infty)$ for the rates of convergence $r_{NT}$ given in Remark 2, and characterize the asymptotic variance $\overline{V}^{\delta(2)}_\infty$. We determine $r_{NT}$ through $\mathbb{E}[(\delta - \delta^0_{NT})^2] = \mathcal{O}(r^{-2}_{NT})$ and $r^{-2}_{NT} = \mathcal{O}(\mathbb{E}[(\delta - \delta^0_{NT})^2])$, where

$$\mathbb{E}[(\delta - \delta^0_{NT})^2] = \mathbb{E}\left[\left(\frac{1}{NT}\sum_{i,t}\widetilde{\Delta}_{it}\right)^2\right] = \frac{1}{N^2T^2}\sum_{i,j,t,s}\mathbb{E}\left[\widetilde{\Delta}_{it}\widetilde{\Delta}_{js}\right], \tag{C.6}$$

for $\widetilde{\Delta}_{it} = \Delta_{it} - \mathbb{E}(\Delta_{it})$. Then, we characterize $\overline{V}^{\delta(2)}_\infty$ as $\overline{V}^{\delta(2)}_\infty = \overline{\mathbb{E}}\{r^2_{NT}\mathbb{E}[(\delta - \delta^0_{NT})^2]\}$, because $\mathbb{E}[\delta - \delta^0_{NT}] = 0$. The order of $\mathbb{E}[(\delta - \delta^0_{NT})^2]$ is equal to the number of terms of the sums in equation (C.6) that are non zero, which it is determined by the sample properties of $\{(X_{it}, \alpha_i, \gamma_t) : 1 \leq i \leq N, 1 \leq t \leq T)\}$.

Under Assumption 4.2$(i)(a)$,

$$\mathbb{E}[(\delta - \delta^0_{NT})^2] = \frac{1}{N^2T^2}\sum_{i,t,s}\mathbb{E}\left[\widetilde{\Delta}_{it}\widetilde{\Delta}_{is}\right] = \mathcal{O}\left(N^{-1}\right),$$

because $\{\widetilde{\Delta}_{it} : 1 \leq i \leq N; 1 \leq t \leq T\}$ is independent across $i$ and $\alpha$-mixing across $t$.

Under Assumption 4.2$(i)(b)$, if $\{\alpha_i\}_N$ and $\{\gamma_t\}_T$ are independent sequences, and $\alpha_i$ and $\gamma_t$ are independent for all $i, t$, then $\mathbb{E}[\widetilde{\Delta}_{it}\widetilde{\Delta}_{js}] = \mathbb{E}[\widetilde{\Delta}_{it}]\mathbb{E}[\widetilde{\Delta}_{js}] = 0$ if $i \neq j$ and $t \neq s$, so that

$$\mathbb{E}[(\delta - \delta^0_{NT})^2] = \frac{1}{N^2T^2}\left\{\sum_{i,t,s}\mathbb{E}\left[\widetilde{\Delta}_{it}\widetilde{\Delta}_{is}\right] + \sum_{i,j,t}\mathbb{E}\left[\widetilde{\Delta}_{it}\widetilde{\Delta}_{jt}\right] - \sum_{i,t}\mathbb{E}\left[\widetilde{\Delta}^2_{it}\right]\right\} = \mathcal{O}\left(\frac{N+T}{NT}\right),$$

because $\mathbb{E}[\widetilde{\Delta}_{it}\widetilde{\Delta}_{is}] \leq \mathbb{E}[\mathbb{E}_\phi(\widetilde{\Delta}^2_{it})]^{1/2}\mathbb{E}[\mathbb{E}_\phi(\widetilde{\Delta}^2_{is})]^{1/2} < C$ by the Cauchy-Schwarz inequality and Assumption 4.2$(ii)$. We conclude that $r_{NT} = \sqrt{NT/(N+T)}$ and

$$\overline{V}^{\delta(2)} = \overline{\mathbb{E}}\left\{\frac{r^2_{NT}}{N^2T^2}\left(\sum_{i,t,s}\mathbb{E}\left[\widetilde{\Delta}_{it}\widetilde{\Delta}_{is}\right] + \sum_{i\neq j,t}\mathbb{E}\left[\widetilde{\Delta}_{it}\widetilde{\Delta}_{jt}\right]\right)\right\}.$$

Note that in both cases $r_{NT} \to \infty$ and $r_{NT} = \mathcal{O}(\sqrt{NT})$.

# Part (3): limit of $r_{NT}(\widehat{\delta} - \delta^0_{NT} - T^{-1}\overline{B}^\delta_\infty - N^{-1}\overline{D}^\delta_\infty)$. The conclusion of the Theorem follows because $(\delta - \delta^0_{NT})$ and $(\widehat{\delta} - \delta - T^{-1}\overline{B}^\delta_\infty - N^{-1}\overline{D}^\delta_\infty)$ are asymptotically independent and $\overline{V}^\delta_\infty = \overline{V}^{\delta(2)} + \overline{V}^{\delta(1)}\lim_{N,T\to\infty}(r_{NT}/\sqrt{NT})^2$. ■

## C.3 Proofs of Theorems 4.3 and 4.4

We start with a lemma that shows the consistency of the fixed effects estimators of averages of the data and parameters. We will use this result to show the validity of the analytical bias corrections and the consistency of the variance estimators.

**Lemma C.2.** *Let $G(\beta, \phi) := [N(T-j)]^{-1}\sum_{i,t\geq j+1} g(X_{it}, X_{i,t-j}, \beta, \alpha_i + \gamma_t, \alpha_i + \gamma_{t-j})$ for $0 \leq j < T$, and $\mathcal{B}^0_\varepsilon$ be a subset of $\mathbb{R}^{\dim\beta+2}$ that contains an $\varepsilon$-neighborhood of $(\beta, \pi^0_{it}, \pi^0_{i,t-j})$ for all $i, t, j, N, T$, and for some $\varepsilon > 0$. Assume that $(\beta, \pi_1, \pi_2) \mapsto g_{itj}(\beta, \pi_1, \pi_2) := g(X_{it}, X_{i,t-j}, \beta, \pi_1, \pi_2)$ is Lipschitz continuous over $\mathcal{B}^0_\varepsilon$ a.s, i.e. $|g_{itj}(\beta_1, \pi_{11}, \pi_{21}) - g_{itj}(\beta_0, \pi_{10}, \pi_{20})| \leq M_{itj}\|(\beta_1, \pi_{11}, \pi_{21}) - (\beta, \pi_{10}, \pi_{20})\|$ for all $(\beta_1, \pi_{11}, \pi_{21}) \in \mathcal{B}^0_\varepsilon$, $(\beta_1, \pi_{11}, \pi_{21}) \in \mathcal{B}^0_\varepsilon$, and some $M_{itj} = \mathcal{O}_P(1)$ for all $i, t, j, N, T$. Let $(\widehat{\beta}, \widehat{\phi})$ be an estimator of $(\beta, \phi)$ such that $\|\widehat{\beta} - \beta^0\| \to_P 0$ and $\|\widehat{\phi} - \phi^0\|_\infty \to_P 0$. Then,*

$$G(\widehat{\beta}, \widehat{\phi}) \to_P \overline{\mathbb{E}}[G(\beta^0, \phi^0)],$$

*provided that the limit exists.*

**Proof of Lemma C.2.** By the triangle inequality

$$|G(\widehat{\beta}, \widehat{\phi}) - \overline{\mathbb{E}}[G(\beta^0, \phi^0)]| \leq |G(\widehat{\beta}, \widehat{\phi}) - G(\beta^0, \phi^0)| + o_P(1),$$

because $|G(\beta^0, \phi^0) - \overline{\mathbb{E}}[G(\beta^0, \phi^0)]| = o_P(1)$. By the local Lipschitz continuity of $g_{itj}$ and the consistency of $(\widehat{\beta}, \widehat{\phi})$,

$$|G(\widehat{\beta}, \widehat{\phi}) - G(\beta^0, \phi^0)| \leq \frac{1}{N(T-j)} \sum_{i,t \geq j+1} M_{itj} \|(\widehat{\beta}, \widehat{\alpha}_i + \widehat{\gamma}_t, \widehat{\alpha}_i + \widehat{\gamma}_{t-j}) - (\beta^0, \alpha_i^0 + \gamma_t^0, \alpha_i^0 + \gamma_{t-j}^0)\|$$
$$\leq \frac{1}{N(T-j)} \sum_{i,t \geq j+1} M_{itj}(\|\widehat{\beta} - \beta^0\| + 4\|\widehat{\phi} - \phi^0\|_\infty)$$

wpa1. The result then follows because $[N(T-j)]^{-1} \sum_{i,\tau \geq t} M_{it\tau} = \mathcal{O}_P(1)$ and $(\|\widehat{\beta} - \beta^0\| + 4\|\widehat{\phi} - \phi^0\|_\infty) = o_P(1)$ by assumption. ■

**Proof of Theorem 4.3.** We separate the proof in three parts corresponding to the three statements of the theorem.

Part I: Proof of $\widehat{W} \to_P \overline{W}_\infty$. The asymptotic variance and its fixed effects estimators can be expressed as $\overline{W}_\infty = \overline{\mathbb{E}}[W(\beta^0, \phi^0)]$ and $\widehat{W} = W(\widehat{\beta}, \widehat{\phi})$, where $W(\beta, \phi)$ has a first order representation as a continuously differentiable transformation of terms that have the form of $G(\beta, \phi)$ in Lemma C.2. The result then follows by the continuous mapping theorem noting that $\|\widehat{\beta} - \beta^0\| \to_P 0$ and $\|\widehat{\phi} - \phi^0\|_\infty \leq \|\widehat{\phi} - \phi^0\|_q \to_P 0$ by Theorem C.1.

Part II: Proof of $\sqrt{NT}(\widetilde{\beta}^A - \beta^0) \to_d \mathcal{N}(0, \overline{W}_\infty^{-1})$. By the argument given after equation (3.3) in the text, we only need to show that $\widehat{B} \to_P \overline{B}_\infty$ and $\widehat{D} \to_P \overline{D}_\infty$. These asymptotic biases and their fixed effects estimators have a similar structure to $\overline{W}_\infty$ and $\widehat{W}$ in part I, so that the consistency follows by a similar argument using that $L \to \infty$ and $L/T \to 0$ guarantee that the trimmed estimators are consistent for the spectral expectations; see Lemma 6 in Hahn and Kuersteiner (2011).

Part III: Proof of $\sqrt{NT}(\widetilde{\beta}^J - \beta^0) \to_d \mathcal{N}(0, \overline{W}_\infty^{-1})$. For $\mathcal{T}_1 = \{1, \ldots, \lfloor (T+1)/2 \rfloor\}$, $\mathcal{T}_2 = \{\lfloor T/2 \rfloor + 1, \ldots, T\}$, $\mathcal{T}_0 = \mathcal{T}_1 \cup \mathcal{T}_2$, $\mathcal{N}_1 = \{1, \ldots, \lfloor (N+1)/2 \rfloor\}$, $\mathcal{N}_2 = \{\lfloor N/2 \rfloor + 1, \ldots, N\}$, and $\mathcal{N}_0 = \mathcal{N}_1 \cup \mathcal{N}_2$, let $\widehat{\beta}^{(jk)}$ be the fixed effect estimator of $\beta$ in the subpanel defined by $i \in \mathcal{N}_j$ and $t \in \mathcal{T}_k$.[15] In this notation,

$$\widetilde{\beta}^J = 3\widehat{\beta}^{(00)} - \widehat{\beta}^{(10)}/2 - \widehat{\beta}^{(20)}/2 - \widehat{\beta}^{(01)}/2 - \widehat{\beta}^{(02)}/2.$$

We derive the asymptotic distribution of $\sqrt{NT}(\widetilde{\beta}^J - \beta^0)$ from the joint asymptotic distribution of the vector $\widehat{\mathbb{B}} = \sqrt{NT}(\widehat{\beta}^{(00)} - \beta^0, \widehat{\beta}^{(10)} - \beta^0, \widehat{\beta}^{(20)} - \beta^0, \widehat{\beta}^{(01)} - \beta^0, \widehat{\beta}^{(02)} - \beta^0)$ with dimension $5 \times \dim \beta$. By Theorem C.1,

$$\sqrt{NT}(\widehat{\beta}^{(jk)} - \beta^0) = \frac{2^{1(j>0)} 2^{1(k>0)}}{\sqrt{NT}} \sum_{i \in N_j, t \in T_k} [\psi_{it} + b_{it} + d_{it}] + o_P(1),$$

for $\psi_{it} = \overline{W}_\infty^{-1} D_\beta \ell_{it}$, $b_{it} = \overline{W}_\infty^{-1}[U_{it}^{(1a,1)} + U_{it}^{(1b,1,1)}]$, and $d_{it} = \overline{W}_\infty^{-1}[U_{it}^{(1a,4)} + U_{it}^{(1b,4,4)}]$, where the $U_{it}^{(\cdot)}$ is implicitly defined by $U^{(\cdot)} = (NT)^{-1/2} \sum_{i,t} U_{it}^{(\cdot)}$. Here, none of the terms carries a superscript $(jk)$ by Assumption 4.3. The influence function $\psi_{it}$ has zero mean and determines the asymptotic variance $\overline{W}_\infty^{-1}$, whereas $b_{it}$ and $d_{it}$ determine the asymptotic biases $\overline{B}_\infty$ and $\overline{D}_\infty$, but do not affect the asymptotic

[15] Note that this definition of the subpanels covers all the cases regardless of whether $N$ and $T$ are even or odd.

variance. By this representation,

$$\widehat{\mathbb{B}} \to_d \mathcal{N}\left(\kappa \begin{bmatrix} 1 \\ 1 \\ 1 \\ 2 \\ 2 \end{bmatrix} \otimes \overline{B}_\infty + \kappa^{-1} \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \\ 1 \end{bmatrix} \otimes \overline{D}_\infty, \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 1 & 1 \\ 1 & 0 & 2 & 1 & 1 \\ 1 & 1 & 1 & 2 & 0 \\ 1 & 1 & 1 & 0 & 2 \end{bmatrix} \otimes \overline{W}_\infty^{-1}\right),$$

where we use that $\{\psi_{it} : 1 \leq i \leq N, 1 \leq t \leq T\}$ is independent across $i$ and martingale difference across $t$ and Assumption 4.3.

The result follows by writing $\sqrt{NT}(\widetilde{\beta}^J - \beta^0) = (3, -1/2, -1/2, -1/2, -1/2)\widehat{\mathbb{B}}$ and using the properties of the multivariate normal distribution. ■

**Proof of Theorem 4.4.** We separate the proof in three parts corresponding to the three statements of the theorem.

Part I: $\widehat{V}^\delta \to_P \overline{V}^\delta_\infty$. $\overline{V}^\delta_\infty$ and $\widehat{V}^\delta$ have a similar structure to $\overline{W}_\infty$ and $\widehat{W}$ in part I of the proof of Theorem 4.3, so that the consistency follows by an analogous argument.

Part II: $\sqrt{NT}(\widetilde{\delta}^A - \delta^0_{NT}) \to_d \mathcal{N}(0, \overline{V}^\delta_\infty)$. As in the proof of Theorem 4.2, we decompose

$$r_{NT}(\widetilde{\delta}^A - \delta^0_{NT}) = r_{NT}(\delta - \delta^0_{NT}) + \frac{r_{NT}}{\sqrt{NT}}\sqrt{NT}(\widetilde{\delta}^A - \delta).$$

Then, by Mann-Wald theorem,

$$\sqrt{NT}(\widetilde{\delta}^A - \delta) = \sqrt{NT}(\widehat{\delta} - \widehat{B}^\delta/T - \widehat{D}^\delta/N - \delta) \to_d \mathcal{N}(0, \overline{V}^{\delta(1)}_\infty),$$

provided that $\widehat{B}^\delta \to_P \overline{B}^\delta_\infty$ and $\widehat{D}^\delta \to_P \overline{D}^\delta_\infty$, and $r_{NT}(\delta - \delta^0_{NT}) \to_d \mathcal{N}(0, \overline{V}^{\delta(2)}_\infty)$, where $\overline{V}^{\delta(1)}_\infty$ and $\overline{V}^{\delta(2)}_\infty$ are defined as in the proof of Theorem 4.2. The statement thus follows by using a similar argument to part II of the proof of Theorem 4.3 to show the consistency of $\widehat{B}^\delta$ and $\widehat{D}^\delta$, and because $(\delta - \delta^0_{NT})$ and $(\widetilde{\delta}^A - \delta)$ are asymptotically independent, and $\overline{V}^\delta_\infty = \overline{V}^{\delta(2)} + \overline{V}^{\delta(1)} \lim_{N,T\to\infty}(r_{NT}/\sqrt{NT})^2$.

Part III: $\sqrt{NT}(\widetilde{\delta}^J - \delta^0_{NT}) \to_d \mathcal{N}(0, \overline{V}^\delta_\infty)$. As in part II, we decompose

$$r_{NT}(\widetilde{\delta}^J - \delta^0_{NT}) = r_{NT}(\delta - \delta^0_{NT}) + \frac{r_{NT}}{\sqrt{NT}}\sqrt{NT}(\widetilde{\delta}^J - \delta).$$

Then, by an argument similar to part III of the proof of Theorem 4.3,

$$\sqrt{NT}(\widetilde{\delta}^J - \delta) \to_d \mathcal{N}(0, \overline{V}^{\delta(1)}_\infty),$$

and $r_{NT}(\delta - \delta^0_{NT}) \to_d \mathcal{N}(0, \overline{V}^{\delta(2)}_\infty)$, where $\overline{V}^{\delta(1)}_\infty$ and $\overline{V}^{\delta(2)}_\infty$ are defined as in the proof of Theorem 4.2. The statement follows because $(\delta - \delta^0_{NT})$ and $(\widetilde{\delta}^J - \delta)$ are asymptotically independent, and $\overline{V}^\delta_\infty = \overline{V}^{\delta(2)} + \overline{V}^{\delta(1)} \lim_{N,T\to\infty}(r_{NT}/\sqrt{NT})^2$. ■

# D Useful Lemmas

## D.1 Some Properties of Stochastic Processes

Here we collect some known properties of $\alpha$-mixing processes, which are useful for our proofs.

**Lemma D.1.** *Let $\{\xi_t\}$ be an $\alpha$-mixing process with mixing coefficients $a(m)$. Let $\mathbb{E}|\xi_t|^p < \infty$ and $\mathbb{E}|\xi_{t+m}|^q < \infty$ for some $p, q \geq 1$ and $1/p + 1/q < 1$. Then,*

$$|\mathrm{Cov}\,(\xi_t,\, \xi_{t+m})| \;\leq\; 8\; a(m)^{1/r}\; [\mathbb{E}|\xi_t|^p]^{1/p}\, [\mathbb{E}|\xi_{t+m}|^q]^{1/q}\,,$$

*where $r = (1 - 1/p - 1/q)^{-1}$.*

**Proof of Lemma D.1.** See, for example, Proposition 2.5 in Fan and Yao (2003). ∎

The following result is a simple modification of Theorem 1 in Cox and Kim (1995).

**Lemma D.2.** *Let $\{\xi_t\}$ be an $\alpha$-mixing process with mixing coefficients $a(m)$. Let $r \geq 1$ be an integer, and let $\delta > 2r$, $\mu > r/(1-2r/\delta)$, $c > 0$ and $C > 0$. Assume that $\sup_t \mathbb{E}\,|\xi_t|^\delta \leq C$ and that $a(m) \leq c\, m^{-\mu}$ for all $m \in \{1, 2, 3, \ldots\}$. Then there exists a constant $B > 0$ depending on $r$, $\delta$, $\mu$, $c$ and $C$, but not depending on $T$ or any other distributional characteristics of $\xi_t$, such that for any $T > 0$,*

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_t\right)^{2r}\right] \leq B.$$

The following is a central limit theorem for martingale difference sequences.

**Lemma D.3.** *Consider the scalar process $\xi_{it} = \xi_{NT,it}$, $i = 1, \ldots, N$, $t = 1, \ldots, T$. Let $\{(\xi_{i1}, \ldots, \xi_{iT}) : 1 \leq i \leq N\}$ be independent across $i$, and be a martingale difference sequence for each $i$, $N$, $T$. Let $\mathbb{E}|\xi_{it}|^{2+\delta}$ be uniformly bounded across $i, t, N, T$ for some $\delta > 0$. Let $\overline{\sigma} = \overline{\sigma}_{NT} > \Delta > 0$ for all sufficiently large $NT$, and let $\frac{1}{NT}\sum_{i,t}\xi_{it}^2 - \overline{\sigma}^2 \to_P 0$ as $NT \to \infty$.*[16] *Then,*

$$\frac{1}{\overline{\sigma}\,\sqrt{NT}}\sum_{i,t}\xi_{it} \to_d \mathcal{N}(0,1).$$

**Proof of Lemma D.3.** Define $\xi_m = \xi_{M,m} = \xi_{NT,it}$, with $M = NT$ and $m = T(i-1)+t \in \{1, \ldots, M\}$. Then $\{\xi_m,\, m = 1, \ldots, M\}$ is a martingale difference sequence. With this redefinition the statement of the Lemma is equal to Corollary 5.26 in White (2001), which is based on Theorem 2.3 in Mcleish (1974), and which shows that $\frac{1}{\overline{\sigma}\,\sqrt{M}}\sum_{m=1}^{M}\xi_m \to_d \mathcal{N}(0,1)$. ∎

## D.2 Some Bounds for the Norms of Matrices and Tensors

The following lemma provides bounds for the matrix norm $\|.\|_q$ in terms of the matrix norms $\|.\|_1$, $\|.\|_2$, $\|.\|_\infty$, and a bound for $\|.\|_2$ in terms of $\|.\|_q$ and $\|.\|_{q/(q-1)}$. For sake of clarity we use notation $\|.\|_2$ for the spectral norm in this lemma, which everywhere else is denoted by $\|.\|$, without any index. Recall that $\|A\|_\infty = \max_i \sum_j |A_{ij}|$ and $\|A\|_1 = \|A'\|_\infty$.

**Lemma D.4.** *For any matrix $A$ we have*

$$\|A\|_q \leq \|A\|_1^{1/q}\|A\|_\infty^{1-1/q}, \qquad \textit{for } q \geq 1,$$
$$\|A\|_q \leq \|A\|_2^{2/q}\|A\|_\infty^{1-2/q}, \qquad \textit{for } q \geq 2,$$
$$\|A\|_2 \leq \sqrt{\|A\|_q\|A\|_{q/(q-1)}}, \qquad \textit{for } q \geq 1.$$

[16] Here can allow for an arbitrary sequence of $(N,T)$ with $NT \to \infty$.

*Note also that* $\|A\|_{q/(q-1)} = \|A'\|_q$ *for* $q \geq 1$. *Thus, for a symmetric matrix* $A$, *we have* $\|A\|_2 \leq \|A\|_q \leq \|A\|_\infty$ *for any* $q \geq 1$.

**Proof of Lemma D.4.** The statements follow from the fact that $\log \|A\|_q$ is a convex function of $1/q$, which is a consequence of the Riesz-Thorin theorem. For more details and references see e.g. Higham (1992). ■

The following lemma shows that the norm $\|.\|_q$ applied to higher-dimensional tensors with a special structure can be expressed in terms of matrix norms $\|.\|_q$. In our panel application all higher dimensional tensors have such a special structure, since they are obtained as partial derivatives wrt to $\alpha$ and $\gamma$ from the likelihood function.

**Lemma D.5.** *Let* $a$ *be an* $N$*-vector with entries* $a_i$, *let* $b$ *be a* $T$*-vector with entries* $b_t$, *and let* $c$ *be an* $N \times T$ *matrix with entries* $c_{it}$. *Let* $A$ *be an* $\underbrace{N \times N \times \ldots \times N}_{p \text{ times}}$ *tensor with entries*

$$A_{i_1 i_2 \ldots i_p} = \begin{cases} a_{i_1} & \text{if } i_1 = i_2 = \ldots = i_p, \\ 0 & \text{otherwise.} \end{cases}$$

*Let* $B$ *be an* $\underbrace{T \times T \times \ldots \times T}_{r \text{ times}}$ *tensor with entries*

$$B_{t_1 t_2 \ldots t_r} = \begin{cases} b_{t_1} & \text{if } t_1 = t_2 = \ldots = t_r, \\ 0 & \text{otherwise.} \end{cases}$$

*Let* $C$ *be an* $\underbrace{N \times N \times \ldots \times N}_{p \text{ times}} \times \underbrace{T \times T \times \ldots \times T}_{r \text{ times}}$ *tensor with entries*

$$C_{i_1 i_2 \ldots i_p t_1 t_2 \ldots t_r} = \begin{cases} c_{i_1 t_1} & \text{if } i_1 = i_2 = \ldots = i_p \text{ and } t_1 = t_2 = \ldots = t_r, \\ 0 & \text{otherwise.} \end{cases}$$

*Let* $\widetilde{C}$ *be an* $\underbrace{T \times T \times \ldots \times T}_{r \text{ times}} \times \underbrace{N \times N \times \ldots \times N}_{p \text{ times}}$ *tensor with entries*

$$\widetilde{C}_{t_1 t_2 \ldots t_r i_1 i_2 \ldots i_p} = \begin{cases} c_{i_1 t_1} & \text{if } i_1 = i_2 = \ldots = i_p \text{ and } t_1 = t_2 = \ldots = t_r, \\ 0 & \text{otherwise.} \end{cases}$$

*Then,*

$$\begin{aligned} \|A\|_q &= \max_i |a_i|, && \text{for } p \geq 2, \\ \|B\|_q &= \max_t |b_t|, && \text{for } r \geq 2, \\ \|C\|_q &\leq \|c\|_q, && \text{for } p \geq 1,\ r \geq 1, \\ \|\widetilde{C}\|_q &\leq \|c'\|_q, && \text{for } p \geq 1,\ r \geq 1, \end{aligned}$$

*where* $\|.\|_q$ *refers to the* $q$*-norm defined in* (A.1) *with* $q \geq 1$.

**Proof of Lemma D.5.** Since the vector norm $\|.\|_{q/(q-1)}$ is dual to the vector norm $\|.\|_q$ we can rewrite the definition of the tensor norm $\|C\|_q$ as follows

$$\|C\|_q = \max_{\|u^{(1)}\|_{q/(q-1)}=1} \max_{\substack{\|u^{(k)}\|_q = 1 \\ k=2,\dots,p}} \max_{\substack{\|v^{(l)}\|_q = 1 \\ l=1,\dots,r}} \left| \sum_{i_1 i_2 \dots i_p=1}^{N} \sum_{t_1 t_2 \dots t_r=1}^{T} u_{i_1}^{(1)} u_{i_2}^{(2)} \cdots u_{i_p}^{(p)} v_{i_1}^{(1)} v_{t_2}^{(2)} \cdots v_{t_r}^{(r)} C_{i_1 i_2 \dots i_p t_1 t_2 \dots t_r} \right| .$$

The specific structure of $C$ yields

$$\begin{aligned}\|C\|_q &= \max_{\|u^{(1)}\|_{q/(q-1)}=1} \max_{\substack{\|u^{(k)}\|_q = 1 \\ k=2,\dots,p}} \max_{\substack{\|v^{(l)}\|_q = 1 \\ l=1,\dots,r}} \left| \sum_{i=1}^{N} \sum_{t=1}^{T} u_i^{(1)} u_i^{(2)} \cdots u_i^{(p)} v_t^{(1)} v_t^{(2)} \cdots v_t^{(r)} c_{it} \right| \\ &\leq \max_{\|u\|_{q/(q-1)}\leq 1} \max_{\|v\|_q \leq 1} \left| \sum_{i=1}^{N} \sum_{t=1}^{T} u_i v_i c_{it} \right| = \|c\|_q,\end{aligned}$$

where we define $u \in \mathbb{R}^N$ with elements $u_i = u_i^{(1)} u_i^{(2)} \cdots u_i^{(p)}$ and $v \in \mathbb{R}^T$ with elements $v_t = v_t^{(1)} v_t^{(2)} \cdots v_t^{(r)}$, and we use that $\|u^{(k)}\|_q = 1$, for $k = 2, \dots, p$, and $\|v^{(l)}\|_q = 1$, for $l = 2, \dots, r$, implies $|u_i| \leq |u_i^{(1)}|$ and $|v_t| \leq |v_t^{(1)}|$, and therefore $\|u\|_{q/(1-q)} \leq \|u^{(1)}\|_{q/(1-q)} = 1$ and $\|v\|_q \leq \|v^{(1)}\|_q = 1$. The proof of $\|\widetilde{C}\|_q \leq \|c'\|_q$ is analogous.

Let $A^{(p)} = A$, as defined above, for a particular value of $p$. For $p = 2$, $A^{(2)}$ is a diagonal $N \times N$ matrix with diagonal elements $a_i$, so that $\|A^{(2)}\|_q \leq \|A^{(2)}\|_1^{1/q} \|A^{(2)}\|_\infty^{1-1/q} = \max_i |a_i|$. For $p > 2$,

$$\begin{aligned}\left\|A^{(p)}\right\|_q &= \max_{\|u^{(1)}\|_{q/(q-1)}=1} \max_{\substack{\|u^{(k)}\|_q = 1 \\ k=2,\dots,p}} \left| \sum_{i_1 i_2 \dots i_p=1}^{N} u_{i_1}^{(1)} u_{i_2}^{(2)} \cdots u_{i_p}^{(p)} A_{i_1 i_2 \dots i_p} \right| \\ &= \max_{\|u^{(1)}\|_{q/(q-1)}=1} \max_{\substack{\|u^{(k)}\|_q = 1 \\ k=2,\dots,p}} \left| \sum_{i,j=1}^{N} u_i^{(1)} u_i^{(2)} \cdots u_i^{(p-1)} u_j^{(p)} A_{ij}^{(2)} \right| \\ &\leq \max_{\|u\|_{q/(q-1)}\leq 1} \max_{\|v\|_q = 1} \left| \sum_{i=1}^{N} \sum_{t=1}^{T} u_i v_i A_{ij}^{(2)} \right| = \|A^{(2)}\|_q \leq \max_i |a_i|,\end{aligned}$$

where we define $u \in \mathbb{R}^N$ with elements $u_i = u_i^{(1)} u_i^{(2)} \cdots u_i^{(p-1)}$ and $v = u^{(p)}$, and we use that $\|u^{(k)}\|_p = 1$, for $k = 2, \dots, p-1$, implies $|u_i| \leq |u_i^{(1)}|$ and therefore $\|u\|_{q/(q-1)} \leq \|u^{(1)}\|_{q/(q-1)} = 1$. We have thus shown $\left\|A^{(p)}\right\| \leq \max_i |a_i|$. From the definition of $\left\|A^{(p)}\right\|_q$ above, we obtain $\left\|A^{(p)}\right\|_q \geq \max_i |a_i|$ by choosing all $u^{(k)}$ equal to the standard basis vector, whose $i^*$'th component equals one, where $i^* \in \operatorname{argmax}_i |a_i|$. Thus, $\left\|A^{(p)}\right\|_q = \max_i |a_i|$ for $p \geq 2$. The proof for $\|B\|_q = \max_t |b_t|$ is analogous. ■

The following lemma provides an asymptotic bound for the spectral norm of $N \times T$ matrices, whose entries are mean zero, and cross-sectionally independent and weakly time-serially dependent conditional on $\phi$.

**Lemma D.6.** *Let $e$ be an $N \times T$ matrix with entries $e_{it}$. Let $\bar{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}_\phi(e_{it}^2)$, let $\Omega$ be the $T \times T$ matrix with entries $\Omega_{ts} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}_\phi(e_{it}e_{is})$, and let $\eta_{ij} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} [e_{it}e_{jt} - \mathbb{E}_\phi(e_{it}e_{jt})]$. Consider* *asymptotic sequences where $N, T \to \infty$ such that $N/T$ converges to a finite positive constant. Assume that*

*(i) The distribution of $e_{it}$ is independent across $i$, conditional on $\phi$, and satisfies $\mathbb{E}_\phi(e_{it}) = 0$.*

*(ii) $\frac{1}{N}\sum_{i=1}^N \left(\bar{\sigma}_i^2\right)^4 = \mathcal{O}_P(1)$, $\quad \frac{1}{T}\mathrm{Tr}(\Omega^4) = \mathcal{O}_P(1)$, $\quad \frac{1}{N}\sum_{i=1}^N \mathbb{E}_\phi\left(\eta_{ii}^4\right) = \mathcal{O}_P(1)$, $\quad \frac{1}{N^2}\sum_{i,j=1}^N \mathbb{E}_\phi\left(\eta_{ij}^4\right) = \mathcal{O}_P(1)$.*

*Then, $\mathbb{E}_\phi\|e\|^8 = \mathcal{O}_P(N^5)$, and therefore $\|e\| = \mathcal{O}_P(N^{5/8})$.*

**Proof of Lemma D.6.** Let $\|.\|_\mathrm{F}$ be the Frobenius norm of a matrix, i.e. $\|A\|_F = \sqrt{\mathrm{Tr}(AA')}$. For $\bar{\sigma}_i^4 = (\bar{\sigma}_i^2)^2$, $\bar{\sigma}_i^8 = (\bar{\sigma}_i^2)^4$ and $\delta_{jk} = 1(j = k)$,

$$\begin{aligned}
\|e\|^8 = \|ee'ee'\|^2 \leq \|ee'ee'\|_\mathrm{F}^2 &= \sum_{i,j=1}^N \left(\sum_{k=1}^N \sum_{t,\tau=1}^T e_{it}e_{kt}e_{k\tau}e_{j\tau}\right)^2 \\
&= T^2 \sum_{i,j=1}^N \left[\sum_{k=1}^N \left(\eta_{ik} + T^{1/2}\delta_{ik}\bar{\sigma}_i^2\right)\left(\eta_{jk} + T^{1/2}\delta_{jk}\bar{\sigma}_j^2\right)\right]^2 \\
&= T^2 \sum_{i,j=1}^N \left(\sum_{k=1}^N \eta_{ik}\eta_{jk} + 2T^{1/2}\eta_{ij}\bar{\sigma}_i^2 + T\delta_{ij}\bar{\sigma}_i^4\right)^2 \\
&\leq 3T^2 \sum_{i,j=1}^N \left[\left(\sum_{k=1}^N \eta_{ik}\eta_{jk}\right)^2 + 4T\eta_{ij}^2\bar{\sigma}_i^4 + T^2\delta_{ij}\bar{\sigma}_i^8\right] \\
&= 3T^2 \sum_{i,j=1}^N \left(\sum_{k=1}^N \eta_{ik}\eta_{jk}\right)^2 + 12T^3 \sum_{i,j=1}^N \bar{\sigma}_i^4\eta_{ij}^2 + 3T^3 \sum_{i=1}^N \bar{\sigma}_i^8,
\end{aligned}$$

where we used that $(a+b+c)^2 \leq 3(a^2+b^2+c^3)$. By the Cauchy Schwarz inequality,

$$\begin{aligned}
\mathbb{E}_\phi\|e\|^8 &\leq 3T^2\mathbb{E}_\phi\left[\sum_{i,j=1}^N \left(\sum_{k=1}^N \eta_{ik}\eta_{jk}\right)^2\right] + 12T^3\sqrt{\left(N\sum_{i=1}^N \bar{\sigma}_i^8\right)\left(\sum_{i,j=1}^N \mathbb{E}_\phi(\eta_{ij}^4)\right)} + 3T^3\sum_{i=1}^N \bar{\sigma}_i^8 \\
&= 3T^2\mathbb{E}_\phi\left[\sum_{i,j=1}^N \left(\sum_{k=1}^N \eta_{ik}\eta_{jk}\right)^2\right] + \mathcal{O}_P(T^3N^2) + \mathcal{O}_P(T^3N).
\end{aligned}$$

Moreover,

$$\mathbb{E}_\phi\left[\sum_{i,j=1}^N\left(\sum_{k=1}^N \eta_{ik}\eta_{jk}\right)^2\right] = \sum_{i,j,k,l=1}^N \mathbb{E}_\phi(\eta_{ik}\eta_{jk}\eta_{il}\eta_{jl}) = \sum_{i,j,k,l=1}^N \mathbb{E}_\phi(\eta_{ij}\eta_{jk}\eta_{kl}\eta_{li})$$

$$\leq \left|\sum_{\substack{i,j,k,l\\ \text{mutually different}}} \mathbb{E}_\phi\left(\eta_{ij}\eta_{jk}\eta_{kl}\eta_{li}\right)\right| + 4\left|\sum_{i,j,k=1}^N a_{ijk}\mathbb{E}_\phi(\eta_{ii}\eta_{ij}\eta_{jk}\eta_{ki})\right|,$$

$$\leq \left|\sum_{\substack{i,j,k,l\\ \text{mutually different}}} \mathbb{E}_\phi\left(\eta_{ij}\eta_{jk}\eta_{kl}\eta_{li}\right)\right| + 4\left\{\left[\sum_{i,j,k=1}^N \mathbb{E}_\phi(\eta_{ii}^4)\right]\left[\sum_{i,j,k=1}^N \mathbb{E}_\phi(\eta_{ij}^4)\right]^3\right\}^{1/4}$$

$$= \left|\sum_{\substack{i,j,k,l\\ \text{mutually different}}} \mathbb{E}_\phi\left(\eta_{ij}\eta_{jk}\eta_{kl}\eta_{li}\right)\right| + 4N^3\left\{\left[\frac{1}{N}\sum_{i=1}^N \mathbb{E}_\phi(\eta_{ii}^4)\right]\left[\frac{1}{N^2}\sum_{i,j=1}^N \mathbb{E}_\phi(\eta_{ij}^4)\right]^3\right\}^{1/4}$$

$$= \left|\sum_{\substack{i,j,k,l\\ \text{mutually different}}} \mathbb{E}_\phi\left(\eta_{ij}\eta_{jk}\eta_{kl}\eta_{li}\right)\right| + \mathcal{O}_P(N^3).$$

where in the second step we just renamed the indices and used that $\eta_{ij}$ is symmetric in $i,j$; and $a_{ijk} \in [0,1]$ in the second line is a combinatorial pre-factor; and in the third step we applied the Cauchy-Schwarz inequality.

Let $\Omega_i$ be the $T \times T$ matrix with entries $\Omega_{i,ts} = \mathbb{E}_\phi(e_{it}e_{is})$ such that $\Omega = \frac{1}{N}\sum_{i=1}^N \Omega_i$. For $i,j,k,l$ mutually different,

$$\mathbb{E}_\phi\left(\eta_{ij}\eta_{jk}\eta_{kl}\eta_{li}\right) = \frac{1}{T^2}\sum_{t,s,u,v=1}^T \mathbb{E}_\phi(e_{it}e_{jt}e_{js}e_{ks}e_{ku}e_{lu}e_{lv}e_{iv})$$

$$= \frac{1}{T^2}\sum_{t,s,u,v=1}^T \mathbb{E}_\phi(e_{iv}e_{it})\mathbb{E}_\phi(e_{jt}e_{js})\mathbb{E}_\phi(e_{ks}e_{ku})\mathbb{E}_\phi(e_{lu}e_{lv}) = \frac{1}{T^2}\mathrm{Tr}(\Omega_i\Omega_j\Omega_k\Omega_l) \geq 0$$

because $\Omega_i \geq 0$ for all $i$. Thus,

$$\left|\sum_{\substack{i,j,k,l\\ \text{mutually different}}} \mathbb{E}_\phi\left(\eta_{ij}\eta_{jk}\eta_{kl}\eta_{li}\right)\right| = \sum_{\substack{i,j,k,l\\ \text{mutually different}}} \mathbb{E}_\phi\left(\eta_{ij}\eta_{jk}\eta_{kl}\eta_{li}\right) = \frac{1}{T^2}\sum_{\substack{i,j,k,l\\ \text{mut. different}}} \mathrm{Tr}(\Omega_i\Omega_j\Omega_k\Omega_l)$$

$$\leq \frac{1}{T^2}\sum_{i,j,k,l=1}^N \mathrm{Tr}(\Omega_i\Omega_j\Omega_k\Omega_l) = \frac{N^4}{T^2}\mathrm{Tr}(\Omega^4) = \mathcal{O}_P(N^4/T).$$

Combining all the above results gives $\mathbb{E}_\phi\|e\|^8 = \mathcal{O}_P(N^5)$, since $N$ and $T$ are assumed to grow at the same rate. ■

## D.3 Verifying the Basic Regularity Conditions in Panel Models

The following Lemma provides sufficient conditions under which the panel fixed effects estimators in the main text satisfy the high-level regularity conditions in Assumptions B.1($v$) and ($vi$).

**Lemma D.7.** *Let $\mathcal{L}(\beta,\phi) = \frac{1}{\sqrt{NT}}\left[\sum_{i,t}\ell_{it}(\beta,\pi_{it}) - \frac{b}{2}(v'\phi)^2\right]$, where $\pi_{it} = \alpha_i + \gamma_t$, $\alpha = (\alpha_1,\ldots,\alpha_N)'$, $\gamma = (\gamma_1,\ldots,\gamma_T)$, $\phi = (\alpha',\gamma')'$, and $v = (1_N', 1_T')'$. Assume that $\ell_{it}(.,.)$ is four times continuously differentiable in an appropriate neighborhood of the true parameter values $(\beta^0,\phi^0)$. Consider limits as $N,T\to\infty$ with $N/T\to\kappa^2>0$. Let $4<q\le 8$ and $0\le\epsilon<1/8-1/(2q)$. Let $r_\beta = r_{\beta,NT}>0$, $r_\phi = r_{\phi,NT}>0$, with $r_\beta = o\left[(NT)^{-1/(2q)-\epsilon}\right]$ and $r_\phi = o\left[(NT)^{-\epsilon}\right]$. Assume that*

*(i) For $k,l,m\in\{1,2,\ldots,\dim\beta\}$,*

$$\frac{1}{\sqrt{NT}}\sum_{i,t}\partial_{\beta_k}\ell_{it} = \mathcal{O}_P(1),\quad \frac{1}{NT}\sum_{i,t}\partial_{\beta_k\beta_l}\ell_{it} = \mathcal{O}_P(1),\quad \frac{1}{NT}\sum_{i,t}\left\{\partial_{\beta_k\beta_l}\ell_{it} - \mathbb{E}_\phi\left[\partial_{\beta_k\beta_l}\ell_{it}\right]\right\} = o_P(1),$$

$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\ \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\frac{1}{NT}\sum_{i,t}\partial_{\beta_k\beta_l\beta_m}\ell_{it}(\beta,\pi_{it}) = \mathcal{O}_P(1).$$

*(ii) Let $k,l\in\{1,2,\ldots,\dim\beta\}$. For $\xi_{it}(\beta,\phi) = \partial_{\beta_k\pi}\ell_{it}(\beta,\pi_{it})$ or $\xi_{it}(\beta,\phi) = \partial_{\beta_k\beta_l\pi}\ell_{it}(\beta,\pi_{it})$,*

$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\ \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\frac{1}{T}\sum_t\left|\frac{1}{N}\sum_i\xi_{it}(\beta,\phi)\right|^q = \mathcal{O}_P(1),$$

$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\ \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\frac{1}{N}\sum_i\left|\frac{1}{T}\sum_t\xi_{it}(\beta,\phi)\right|^q = \mathcal{O}_P(1).$$

*(iii) Let $k,l\in\{1,2,\ldots,\dim\beta\}$. For $\xi_{it}(\beta,\phi) = \partial_{\pi^r}\ell_{it}(\beta,\pi_{it})$, with $r\in\{3,4\}$, or $\xi_{it}(\beta,\phi) = \partial_{\beta_k\pi^r}\ell_{it}(\beta,\pi_{it})$, with $r\in\{2,3\}$, or $\xi_{it}(\beta,\phi) = \partial_{\beta_k\beta_l\pi^2}\ell_{it}(\beta,\pi_{it})$,*

$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\ \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\max_i\frac{1}{T}\sum_t|\xi_{it}(\beta,\phi)| = \mathcal{O}_P\left(N^{2\epsilon}\right),$$

$$\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)}\ \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)}\max_t\frac{1}{N}\sum_i|\xi_{it}(\beta,\phi)| = \mathcal{O}_P\left(N^{2\epsilon}\right).$$

*(iv) Moreover,*

$$\frac{1}{T}\sum_t\left|\frac{1}{\sqrt{N}}\sum_i\partial_\pi\ell_{it}\right|^q = \mathcal{O}_P(1),\quad \frac{1}{N}\sum_i\left|\frac{1}{\sqrt{T}}\sum_t\partial_\pi\ell_{it}\right|^q = \mathcal{O}_P(1),$$

$$\frac{1}{T}\sum_t\left|\frac{1}{\sqrt{N}}\sum_i\partial_{\beta_k\pi}\ell_{it} - \mathbb{E}_\phi\left[\partial_{\beta_k\pi}\ell_{it}\right]\right|^2 = \mathcal{O}_P(1),$$

$$\frac{1}{N}\sum_i\left|\frac{1}{\sqrt{T}}\sum_t\partial_{\beta_k\pi}\ell_{it} - \mathbb{E}_\phi\left[\partial_{\beta_k\pi}\ell_{it}\right]\right|^2 = \mathcal{O}_P(1).$$

*(v) The sequence $\{(\ell_{i1},\ldots,\ell_{iT}) : 1\le i\le N\}$ is independent across $i$ conditional on $\phi$.*

*(vi) Let $k\in\{1,2,\ldots,\dim\beta\}$. For $\xi_{it} = \partial_{\pi^r}\ell_{it} - \mathbb{E}_\phi\left[\partial_{\pi^r}\ell_{it}\right]$, with $r\in\{2,3\}$, or $\xi_{it} = \partial_{\beta_k\pi^2}\ell_{it} - \mathbb{E}_\phi\left[\partial_{\beta_k\pi^2}\ell_{it}\right]$, and some $\tilde{\nu}>0$,*

$$\max_i\mathbb{E}_\phi\left[\xi_{it}^{8+\tilde{\nu}}\right]\le C,\quad \max_i\max_t\sum_s\mathbb{E}_\phi\left[\xi_{it}\xi_{is}\right]\le C,\quad \max_i\mathbb{E}_\phi\left\{\left[\frac{1}{\sqrt{T}}\sum_t\xi_{it}\right]^8\right\}\le C,$$

$$\max_t\mathbb{E}_\phi\left\{\left[\frac{1}{\sqrt{N}}\sum_i\xi_{it}\right]^8\right\}\le C,\quad \max_{i,j}\mathbb{E}_\phi\left\{\left[\frac{1}{\sqrt{T}}\sum_t\left[\xi_{it}\xi_{jt} - \mathbb{E}_\phi\left(\xi_{it}\xi_{jt}\right)\right]\right]^4\right\}\le C,$$

*uniformly in $N,T$, where $C>0$ is a constant.*

*(vii)* $\left\|\overline{\mathcal{H}}^{-1}\right\|_q = \mathcal{O}_P(1)$.

*Then, Assumptions B.1(v) and (vi) are satisfied with the same parameters $q$, $\epsilon$, $r_\beta = r_{\beta,NT}$ and $r_\phi = r_{\phi,NT}$ used here.*

**Proof of Lemma D.7.** The penalty term $(v'\phi)^2$ is quadratic in $\phi$ and does not depend on $\beta$. This term thus only enters $\partial_\phi \mathcal{L}(\beta,\phi)$ and $\partial_{\phi\phi'}\mathcal{L}(\beta,\phi)$, but it does not effect any other partial derivative of $\mathcal{L}(\beta,\phi)$. Furthermore, the contribution of the penalty drops out of $\mathcal{S} = \partial_\phi \mathcal{L}(\beta^0,\phi^0)$, because we impose the normalization $v'\phi^0 = 0$. It also drops out of $\tilde{\mathcal{H}}$, because it contributes the same to $\mathcal{H}$ and $\overline{\mathcal{H}}$. We can therefore ignore the penalty term for the purpose of proving the lemma (but it is necessary to satisfy the assumption $\left\|\overline{\mathcal{H}}^{-1}\right\|_q = \mathcal{O}_P(1)$).

# Assumption $(i)$ implies that $\|\partial_\beta \mathcal{L}\| = \mathcal{O}_P(1)$, $\|\partial_{\beta\beta'}\mathcal{L}\| = \mathcal{O}_P(\sqrt{NT})$, $\left\|\partial_{\beta\beta'}\widetilde{\mathcal{L}}\right\| = o_P(\sqrt{NT})$, and $\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\beta\beta}\mathcal{L}(\beta,\,\phi)\| = \mathcal{O}_P\left(\sqrt{NT}\right)$. Note that it does not matter which norms we use here because $\dim\beta$ is fixed.

# By Assumption $(ii)$, $\|\partial_{\beta\phi'}\mathcal{L}\|_q = \mathcal{O}_P\left((NT)^{1/(2q)}\right)$ and $\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\beta\phi}\mathcal{L}(\beta,\,\phi)\|_q = \mathcal{O}_P\left((NT)^{1/(2q)}\right)$. For example, $\partial_{\beta_k\alpha_i}\mathcal{L} = \frac{1}{\sqrt{NT}}\sum_t \partial_{\beta_k\pi}\ell_{it}$ and therefore

$$\|\partial_{\beta_k\alpha}\mathcal{L}\|_q = \left(\sum_i \left|\frac{1}{\sqrt{NT}}\sum_t \partial_{\beta_k\pi}\ell_{it}\right|^q\right)^{1/q} = \mathcal{O}_P\left(N^{1/q}\right) = \mathcal{O}_P\left((NT)^{1/(2q)}\right).$$

Analogously, $\|\partial_{\beta_k\gamma}\mathcal{L}\|_q = \mathcal{O}_P\left((NT)^{1/(2q)}\right)$, and therefore $\|\partial_{\beta_k\phi}\mathcal{L}\|_q \le \|\partial_{\beta_k\alpha}\mathcal{L}\|_q + \|\partial_{\beta_k\gamma}\mathcal{L}\|_q = \mathcal{O}_P\left((NT)^{1/(2q)}\right)$. This also implies that $\|\partial_{\beta\phi'}\mathcal{L}\|_q = \mathcal{O}_P\left((NT)^{1/(2q)}\right)$ because $\dim\beta$ is fixed.

# By Assumption $(iii)$, $\|\partial_{\phi\phi\phi}\mathcal{L}\|_q = \mathcal{O}_P\left((NT)^\epsilon\right)$, $\|\partial_{\beta\phi\phi}\mathcal{L}\|_q = \mathcal{O}_P((NT)^\epsilon)$, $\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\beta\phi\phi}\mathcal{L}(\beta,\,\phi)\|_q = \mathcal{O}_P\left((NT)^\epsilon\right)$, $\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\beta\phi\phi\phi}\mathcal{L}(\beta,\,\phi)\|_q = \mathcal{O}_P\left((NT)^\epsilon\right)$, and $\sup_{\beta\in\mathcal{B}(r_\beta,\beta^0)} \sup_{\phi\in\mathcal{B}_q(r_\phi,\phi^0)} \|\partial_{\phi\phi\phi\phi}\mathcal{L}(\beta,\phi)\|_q = \mathcal{O}_P\left((NT)^\epsilon\right)$. For example,

$$\begin{aligned}
\|\partial_{\phi\phi\phi}\mathcal{L}\|_q &\le \|\partial_{\alpha\alpha\alpha}\mathcal{L}\|_q + \|\partial_{\alpha\alpha\gamma}\mathcal{L}\|_q + \|\partial_{\alpha\gamma\alpha}\mathcal{L}\|_q + \|\partial_{\alpha\gamma\gamma}\mathcal{L}\|_q \\
&\qquad + \|\partial_{\gamma\alpha\alpha}\mathcal{L}\|_q + \|\partial_{\gamma\alpha\gamma}\mathcal{L}\|_q + \|\partial_{\gamma\gamma\alpha}\mathcal{L}\|_q + \|\partial_{\gamma\gamma\gamma}\mathcal{L}\|_q \\
&\le \|\partial_{\pi\alpha\alpha}\mathcal{L}\|_q + \|\partial_{\pi\gamma\gamma}\mathcal{L}\|_q + 3\,\|\partial_{\pi\alpha\gamma}\mathcal{L}\|_q + 3\,\|\partial_{\pi\gamma\alpha}\mathcal{L}\|_q \\
&\le \|\partial_{\pi\alpha\alpha}\mathcal{L}\|_\infty + \|\partial_{\pi\gamma\gamma}\mathcal{L}\|_\infty + 3\,\|\partial_{\pi\alpha\gamma}\mathcal{L}\|_\infty^{1-1/q}\,\|\partial_{\pi\gamma\alpha}\mathcal{L}\|_\infty^{1/q} + 3\,\|\partial_{\pi\alpha\gamma}\mathcal{L}\|_\infty^{1/q}\,\|\partial_{\pi\gamma\alpha}\mathcal{L}\|_\infty^{1-1/q} \\
&= \frac{1}{\sqrt{NT}}\left[\max_i\left|\sum_t \partial_{\pi^3}\ell_{it}\right| + \max_t\left|\sum_i \partial_{\pi^3}\ell_{it}\right| + 3\left(\max_i\sum_t|\partial_{\pi^3}\ell_{it}|\right)^{1-1/q}\left(\max_t\sum_t|\partial_{\pi^3}\ell_{it}|\right)^{1/q}\right. \\
&\qquad \left. + 3\left(\max_i\sum_t|\partial_{\pi^3}\ell_{it}|\right)^{1/q}\left(\max_t\sum_t|\partial_{\pi^3}\ell_{it}|\right)^{1-1/q}\right] \\
&\le \frac{1}{\sqrt{NT}}\left[\max_i\sum_t|\partial_{\pi^3}\ell_{it}| + \max_t\sum_i|\partial_{\pi^3}\ell_{it}| + 3\left(\max_i\sum_t|\partial_{\pi^3}\ell_{it}|\right)^{1-1/q}\left(\max_t\sum_t|\partial_{\pi^3}\ell_{it}|\right)^{1/q}\right. \\
&\qquad \left. + 3\left(\max_i\sum_t|\partial_{\pi^3}\ell_{it}|\right)^{1/q}\left(\max_t\sum_t|\partial_{\pi^3}\ell_{it}|\right)^{1-1/q}\right] = \mathcal{O}_P(N^{2\epsilon}) = \mathcal{O}_P((NT)^\epsilon).
\end{aligned}$$

Here, we use Lemma D.5 to bound the norms of the 3-tensors in terms of the norms of matrices, e.g. $\|\partial_{\alpha\alpha\gamma}\mathcal{L}\|_q \leq \|\partial_{\pi\alpha\gamma}\mathcal{L}\|_q$, because $\partial_{\alpha_i\alpha_j\gamma_t}\mathcal{L} = 0$ if $i \neq j$ and $\partial_{\alpha_i\alpha_i\gamma_t}\mathcal{L} = (NT)^{-1/2}\partial_{\pi\alpha_i\gamma_t}$.[17] Then, we use Lemma D.4 to bound $q$-norms in terms of $\infty$-norms, and then explicitly expressed those $\infty$-norm in terms of the elements of the matrices. Finally, we use that $|\sum_i \partial_{\pi^3}\ell_{it}| \leq \sum_i |\partial_{\pi^3}\ell_{it}|$ and $|\sum_t \partial_{\pi^3}\ell_{it}| \leq \sum_t |\partial_{\pi^3}\ell_{it}|$, and apply Assumption $(iii)$.

# By Assumption $(iv)$, $\|\mathcal{S}\|_q = \mathcal{O}_P\left((NT)^{-1/4+1/(2q)}\right)$ and $\left\|\partial_{\beta\phi'}\widetilde{\mathcal{L}}\right\| = \mathcal{O}_P(1)$. For example,

$$\|\mathcal{S}\|_q = \frac{1}{\sqrt{NT}}\left(\sum_i \left|\sum_t \partial_\pi \ell_{it}\right|^q + \sum_t \left|\sum_i \partial_\pi \ell_{it}\right|^q\right)^{1/q} = \mathcal{O}_P\left(N^{-1/2+1/q}\right) = \mathcal{O}_P\left((NT)^{-1/4+1/(2q)}\right).$$

# By Assumption $(v)$ and $(vi)$, $\|\widetilde{\mathcal{H}}\| = \mathcal{O}_P\left((NT)^{-3/16}\right) = o_P\left((NT)^{-1/8}\right)$ and $\left\|\partial_{\beta\phi\phi}\widetilde{\mathcal{L}}\right\| = \mathcal{O}_P\left((NT)^{-3/16}\right) = o_P\left((NT)^{-1/8}\right)$. We now show it $\|\widetilde{\mathcal{H}}\|$. The proof for $\left\|\partial_{\beta\phi\phi}\widetilde{\mathcal{L}}\right\|$ is analogous.

By the triangle inequality,

$$\|\widetilde{\mathcal{H}}\| = \|\partial_{\phi\phi'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\phi\phi'}\mathcal{L}\right]\| \leq \|\partial_{\alpha\alpha'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\alpha\alpha'}\mathcal{L}\right]\| + \|\partial_{\gamma\gamma'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\gamma\gamma'}\mathcal{L}\right]\| + 2\left\|\partial_{\alpha\gamma'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\alpha\gamma'}\mathcal{L}\right]\right\|.$$

Let $\xi_{it} = \partial_{\pi^2}\ell_{it} - \mathbb{E}_\phi\left[\partial_{\pi^2}\ell_{it}\right]$. Since $\partial_{\alpha\alpha'}\mathcal{L}$ is a diagonal matrix with diagonal entries $\frac{1}{\sqrt{NT}}\sum_t \xi_{it}$, $\|\partial_{\alpha\alpha'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\alpha\alpha'}\mathcal{L}\right]\| = \max_i \frac{1}{\sqrt{NT}}\sum_t \xi_{it}$, and therefore

$$\begin{aligned}\mathbb{E}_\phi\|\partial_{\alpha\alpha'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\alpha\alpha'}\mathcal{L}\right]\|^8 &= \mathbb{E}_\phi\left[\max_i\left(\frac{1}{\sqrt{NT}}\sum_t \xi_{it}\right)^8\right] \\ &\leq \mathbb{E}_\phi\left[\sum_i\left(\frac{1}{\sqrt{NT}}\sum_t \xi_{it}\right)^8\right] \leq CN\left(\frac{1}{\sqrt{N}}\right)^8 = \mathcal{O}_P(N^{-3}).\end{aligned}$$

Thus, $\|\partial_{\alpha\alpha'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\alpha\alpha'}\mathcal{L}\right]\| = \mathcal{O}_P(N^{-3/8})$. Analogously, $\|\partial_{\gamma\gamma'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\gamma\gamma'}\mathcal{L}\right]\| = \mathcal{O}_P(N^{-3/8})$.

Let $\xi$ be the $N \times T$ matrix with entries $\xi_{it}$. We now show that $\xi$ satisfies all the regularity condition of Lemma D.6 with $e_{it} = \xi_{it}$. Independence across $i$ is assumed. Furthermore, $\bar{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^T \mathbb{E}_\phi(\xi_{it}^2) \leq C^{1/4}$ so that $\frac{1}{N}\sum_{i=1}^N \left(\bar{\sigma}_i^2\right)^4 = \mathcal{O}_P(1)$. For $\Omega_{ts} = \frac{1}{N}\sum_{i=1}^N \mathbb{E}_\phi(\xi_{it}\xi_{is})$,

$$\frac{1}{T}\mathrm{Tr}(\Omega^4) \leq \|\Omega\|^4 \leq \|\Omega\|_\infty^4 = \left(\max_t \sum_s \mathbb{E}_\phi\left[\xi_{it}\xi_{is}\right]\right)^4 \leq C = \mathcal{O}_P(1).$$

For $\eta_{ij} = \frac{1}{\sqrt{T}}\sum_{t=1}^T\left[\xi_{it}\xi_{jt} - \mathbb{E}_\phi(\xi_{it}\xi_{jt})\right]$ we assume $\mathbb{E}_\phi\eta_{ij}^4 \leq C$, which implies $\frac{1}{N}\sum_{i=1}^N \mathbb{E}_\phi\left(\eta_{ii}^4\right) = \mathcal{O}_P(1)$ and $\frac{1}{N^2}\sum_{i,j=1}^N \mathbb{E}_\phi\left(\eta_{ij}^4\right) = \mathcal{O}_P(1)$. Then, Lemma D.6 gives $\|\xi\| = \mathcal{O}_P(N^{5/8})$. Note that $\xi = \frac{1}{\sqrt{NT}}\partial_{\alpha\gamma'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\alpha\gamma'}\mathcal{L}\right]$ and therefore $\|\partial_{\alpha\gamma'}\mathcal{L} - \mathbb{E}_\phi\left[\partial_{\alpha\gamma'}\mathcal{L}\right]\| = \mathcal{O}_P(N^{-3/8})$. We conclude that $\|\widetilde{\mathcal{H}}\| = \mathcal{O}_P(N^{-3/8}) = \mathcal{O}_P\left((NT)^{-3/16}\right)$.

# Moreover, for $\xi_{it} = \partial_{\pi^2}\ell_{it} - \mathbb{E}_\phi\left[\partial_{\pi^2}\ell_{it}\right]$

$$\begin{aligned}\mathbb{E}_\phi\|\widetilde{\mathcal{H}}\|_\infty^{8+\tilde{\nu}} &= \mathbb{E}_\phi\left(\frac{1}{\sqrt{NT}}\max_i \sum_t |\xi_{it}|\right)^{8+\tilde{\nu}} = \mathbb{E}_\phi \max_i\left(\frac{1}{\sqrt{NT}}\sum_t |\xi_{it}|\right)^{8+\tilde{\nu}} \\ &\leq \mathbb{E}_\phi\sum_i\left(\frac{1}{\sqrt{NT}}\sum_t |\xi_{it}|\right)^{8+\tilde{\nu}} \leq \mathbb{E}_\phi\sum_i\left(\frac{T}{\sqrt{NT}}\right)^{8+\tilde{\nu}}\left(\frac{1}{T}\sum_t |\xi_{it}|^{8+\tilde{\nu}}\right) = \mathcal{O}_P(N),\end{aligned}$$

[17]With a slight abuse of notation we write $\partial_{\pi\alpha\gamma}\mathcal{L}$ for the $N \times T$ matrix with entries $(NT)^{-1/2}\partial_{\pi^3}\ell_{it} = (NT)^{-1/2}\partial_{\pi^3}\ell_{it}$, and analogously for $\partial_{\pi\alpha\alpha}\mathcal{L}$, $\partial_{\pi\gamma\gamma}\mathcal{L}$, and $\partial_{\pi\gamma\alpha}\mathcal{L}$.

and therefore $\|\widetilde{\mathcal{H}}\|_\infty = o_P(N^{1/8})$. Thus, by Lemma D.4

$$\|\widetilde{\mathcal{H}}\|_q \leq \|\widetilde{\mathcal{H}}\|_2^{2/q} \|\widetilde{\mathcal{H}}\|_\infty^{1-2/q} = o_P\left(N^{1/8[-6/q+(1-2/q)]}\right) = o_P\left(N^{-1/q+1/8}\right) = o_P(1),$$

where we use that $q \leq 8$.

\# Finally we show that $\left\|\sum_{g,h=1}^{\dim\phi} \partial_{\phi\phi_g\phi_h} \widetilde{\mathcal{L}}\, [\overline{\mathcal{H}}^{-1}\mathcal{S}]_g [\overline{\mathcal{H}}^{-1}\mathcal{S}]_h\right\| = o_P\left((NT)^{-1/4}\right)$. First,

$$\begin{aligned}
&\left\|\sum_{g,h=1}^{\dim\phi} \partial_{\phi\phi_g\phi_h} \widetilde{\mathcal{L}}\, [\overline{\mathcal{H}}^{-1}\mathcal{S}]_g [\overline{\mathcal{H}}^{-1}\mathcal{S}]_h\right\| \\
&\qquad \leq \left\|\sum_{g,h=1}^{\dim\phi} \partial_{\alpha\phi_g\phi_h} \widetilde{\mathcal{L}}\, [\overline{\mathcal{H}}^{-1}\mathcal{S}]_g [\overline{\mathcal{H}}^{-1}\mathcal{S}]_h\right\| + \left\|\sum_{g,h=1}^{\dim\phi} \partial_{\gamma\phi_g\phi_h} \widetilde{\mathcal{L}}\, [\overline{\mathcal{H}}^{-1}\mathcal{S}]_g [\overline{\mathcal{H}}^{-1}\mathcal{S}]_h\right\|.
\end{aligned}$$

Let $(v,w)' := \overline{\mathcal{H}}^{-1}\mathcal{S}$, where $v$ is a $N$-vector and $w$ is a $T$-vector. We assume $\left\|\overline{\mathcal{H}}^{-1}\right\|_q = \mathcal{O}_P(1)$. By Lemma B.5 this also implies $\left\|\overline{\mathcal{H}}^{-1}\right\| = \mathcal{O}_P(1)$ and $\|\mathcal{S}\| = \mathcal{O}_P(1)$. Thus, $\|v\| \leq \left\|\overline{\mathcal{H}}^{-1}\right\| \|\mathcal{S}\| = \mathcal{O}_P(1)$, $\|w\| \leq \left\|\overline{\mathcal{H}}^{-1}\right\| \|\mathcal{S}\| = \mathcal{O}_P(1)$, $\|v\|_\infty \leq \|v\|_q \leq \left\|\overline{\mathcal{H}}^{-1}\right\|_q \|\mathcal{S}\|_q = \mathcal{O}_P\left((NT)^{-1/4+1/(2q)}\right)$, $\|w\|_\infty \leq \|w\|_q \leq \left\|\overline{\mathcal{H}}^{-1}\right\|_q \|\mathcal{S}\|_q = \mathcal{O}_P\left((NT)^{-1/4+1/(2q)}\right)$. Furthermore, by an analogous argument to the above proof for $\|\widetilde{\mathcal{H}}\|$, Assumption $(v)$ and $(vi)$ imply that $\left\|\partial_{\pi\alpha\alpha'}\widetilde{\mathcal{L}}\right\| = \mathcal{O}_P(N^{-3/8})$, $\left\|\partial_{\pi\alpha\gamma'}\widetilde{\mathcal{L}}\right\| = \mathcal{O}_P(N^{-3/8})$, $\left\|\partial_{\pi\gamma\gamma'}\widetilde{\mathcal{L}}\right\| = \mathcal{O}_P(N^{-3/8})$. Then,

$$\begin{aligned}
\sum_{g,h=1}^{\dim\phi} \partial_{\alpha_i\phi_g\phi_h} \widetilde{\mathcal{L}}\, [\overline{\mathcal{H}}^{-1}\mathcal{S}]_g [\overline{\mathcal{H}}^{-1}\mathcal{S}]_h &= \sum_{j,k=1}^{N} (\partial_{\alpha_i\alpha_j\alpha_k}\widetilde{\mathcal{L}}) v_j v_k + 2\sum_{j=1}^{N}\sum_{t=1}^{T} (\partial_{\alpha_i\alpha_j\gamma_t}\widetilde{\mathcal{L}}) v_j w_t + \sum_{t,s=1}^{T} (\partial_{\alpha_i\gamma_t\gamma_s}\widetilde{\mathcal{L}}) w_t w_s \\
&= \sum_{j=1}^{N} (\partial_{\pi^2\alpha_i}\widetilde{\mathcal{L}}) v_i^2 + 2\sum_{t=1}^{T} (\partial_{\pi\alpha_i\gamma_t}\widetilde{\mathcal{L}}) v_i w_t + \sum_{t=1}^{T} (\partial_{\pi\alpha_i\gamma_t}\widetilde{\mathcal{L}}) w_t^2,
\end{aligned}$$

and therefore

$$\begin{aligned}
\left\|\sum_{g,h=1}^{\dim\phi} \partial_{\alpha\phi_g\phi_h} \widetilde{\mathcal{L}}\, [\overline{\mathcal{H}}^{-1}\mathcal{S}]_g [\overline{\mathcal{H}}^{-1}\mathcal{S}]_h\right\| &\leq \left\|\partial_{\pi\alpha\alpha'}\widetilde{\mathcal{L}}\right\| \|v\|\|v\|_\infty + 2\left\|\partial_{\pi\alpha\gamma'}\widetilde{\mathcal{L}}\right\| \|w\|\|v\|_\infty + \left\|\partial_{\pi\alpha\gamma'}\widetilde{\mathcal{L}}\right\| \|w\|\|w\|_\infty \\
&= \mathcal{O}_P(N^{-3/8})\mathcal{O}_P\left((NT)^{-1/4+1/(2q)}\right) = \mathcal{O}_P\left((NT)^{-1/4-3/16+1/(2q)}\right) = o_P\left((NT)^{-1/4}\right),
\end{aligned}$$

where we use that $q > 4$. Analogously, $\left\|\sum_{g,h=1}^{\dim\phi} \partial_{\gamma\phi_g\phi_h} \widetilde{\mathcal{L}}\, [\overline{\mathcal{H}}^{-1}\mathcal{S}]_g [\overline{\mathcal{H}}^{-1}\mathcal{S}]_h\right\| = o_P\left((NT)^{-1/4}\right)$ and thus also $\left\|\sum_{g,h=1}^{\dim\phi} \partial_{\phi\phi_g\phi_h} \widetilde{\mathcal{L}}\, [\overline{\mathcal{H}}^{-1}\mathcal{S}]_g [\overline{\mathcal{H}}^{-1}\mathcal{S}]_h\right\| = o_P\left((NT)^{-1/4}\right)$.[18] ■

[18] Given the structure of this last part of the proof of Lemma D.7 one might wonder why, instead of $\left\|\sum_{g,h=1}^{\dim\phi} \partial_{\phi\phi_g\phi_h} \widetilde{\mathcal{L}}\, [\overline{\mathcal{H}}^{-1}\mathcal{S}]_g [\overline{\mathcal{H}}^{-1}\mathcal{S}]_h\right\| = o_P\left((NT)^{-1/4}\right)$, we did not directly impose $\sum_g \left\|\partial_{\phi_g\phi\phi'}\widetilde{\mathcal{L}}\right\| = o_P\left((NT)^{-1/(2q)}\right)$ as a high-level condition in Assumption B.1$(vi)$. While this alternative high-level assumption would indeed be more elegant and sufficient to derive our results, it would not be satisfied for panel models, because it involves bounding $\sum_i \left\|\partial_{\alpha_i\gamma\gamma'}\widetilde{\mathcal{L}}\right\|$ and $\sum_t \left\|\partial_{\gamma_t\alpha\alpha'}\widetilde{\mathcal{L}}\right\|$, which was avoided in the proof of Lemma D.7.

## D.4 Properties of the Inverse Expected Incidental Parameter Hessian

The expected incidental parameter Hessian evaluated at the true parameter values is

$$\overline{\mathcal{H}} = \mathbb{E}_\phi[\partial_{\phi\phi'}\mathcal{L}] = \begin{pmatrix} \overline{\mathcal{H}}^*_{(\alpha\alpha)} & \overline{\mathcal{H}}^*_{(\alpha\gamma)} \\ [\overline{\mathcal{H}}^*_{(\alpha\gamma)}]' & \overline{\mathcal{H}}^*_{(\gamma\gamma)} \end{pmatrix} + \frac{b}{\sqrt{NT}}\, vv',$$

where $v = v_{NT} = (1'_N, -1'_T)'$, $\overline{\mathcal{H}}^*_{(\alpha\alpha)} = \operatorname{diag}(\frac{1}{\sqrt{NT}}\sum_t \mathbb{E}_\phi[-\partial_{\pi^2}\ell_{it}])$, $\overline{\mathcal{H}}^*_{(\alpha\gamma)it} = \frac{1}{\sqrt{NT}}\mathbb{E}_\phi[-\partial_{\pi^2}\ell_{it}]$, and $\overline{\mathcal{H}}^*_{(\gamma\gamma)} = \operatorname{diag}(\frac{1}{\sqrt{NT}}\sum_i \mathbb{E}_\phi[-\partial_{\pi^2}\ell_{it}])$.

In panel models with only individual effects, it is straightforward to determine the order of magnitude of $\overline{\mathcal{H}}^{-1}$ in Assumption B.1$(iv)$, because $\overline{\mathcal{H}}$ contains only the diagonal matrix $\overline{\mathcal{H}}^*_{(\alpha\alpha)}$. In our case, $\overline{\mathcal{H}}$ is no longer diagonal, but it has a special structure. The diagonal terms are of order 1, whereas the off-diagonal terms are of order $(NT)^{-1/2}$. Moreover, $\left\|\overline{\mathcal{H}} - \operatorname{diag}(\overline{\mathcal{H}}^*_{(\alpha\alpha)}, \overline{\mathcal{H}}^*_{(\gamma\gamma)})\right\|_{\max} = \mathcal{O}_P((NT)^{-1/2})$. These observations, however, are not sufficient to establish the order of $\overline{\mathcal{H}}^{-1}$ because the number of non-zero off-diagonal terms is of much larger order than the number of diagonal terms; compare $\mathcal{O}(NT)$ to $\mathcal{O}(N+T)$. Note also that the expected Hessian without penalty term $\overline{\mathcal{H}}^*$ has the same structure as $\overline{\mathcal{H}}$ itself, but is not even invertible, i.e. the observation on the relative size of diagonal vs. off-diagonal terms is certainly not sufficient to make statements about the structure of $\overline{\mathcal{H}}^{-1}$. The result of the following lemma is therefore not obvious. It shows that the diagonal terms of $\overline{\mathcal{H}}$ also dominate in determining the order of $\overline{\mathcal{H}}^{-1}$.

**Lemma D.8.** *Under Assumptions 4.1,*

$$\left\|\overline{\mathcal{H}}^{-1} - \operatorname{diag}\left(\overline{\mathcal{H}}^*_{(\alpha\alpha)}, \overline{\mathcal{H}}^*_{(\gamma\gamma)}\right)^{-1}\right\|_{\max} = \mathcal{O}_P\left((NT)^{-1/2}\right).$$

This result establishes that $\overline{\mathcal{H}}^{-1}$ can be uniformly approximated by a diagonal matrix, which is given by the inverse of the diagonal terms of $\overline{\mathcal{H}}$ without the penalty. The diagonal elements of $\operatorname{diag}(\overline{\mathcal{H}}^*_{(\alpha\alpha)}, \overline{\mathcal{H}}^*_{(\gamma\gamma)})^{-1}$ are of order 1, i.e. the order of the difference established by the lemma is relatively small.

Note that the choice of penalty in the objective function is important to obtain Lemma D.8. Different penalties, corresponding to other normalizations (e.g. a penalty proportional to $\alpha_1^2$, corresponding to the normalization $\alpha_1 = 0$), would fail to deliver Lemma D.8. However, these alternative choices do not affect the estimators $\widehat{\beta}$ and $\widehat{\delta}$, i.e. which normalization is used to compute $\widehat{\beta}$ and $\widehat{\delta}$ in practice is irrelevant (up to numerical precision errors).

### D.4.1 Proof of Lemma D.8

The following Lemmas are useful to prove Lemma D.8. Let $\mathcal{L}^*(\beta, \phi) = (NT)^{-1/2}\sum_{i,t}\ell_{it}(\beta, \alpha_i + \gamma_t)$.

**Lemma D.9.** *If the statement of Lemma D.8 holds for some constant $b > 0$, then it holds for any constant $b > 0$.*

**Proof of Lemma D.9.** Write $\overline{\mathcal{H}} = \overline{\mathcal{H}}^* + \frac{b}{\sqrt{NT}}vv'$, where $\overline{\mathcal{H}}^* = \mathbb{E}_\phi\left[-\frac{\partial^2}{\partial\phi\partial\phi'}\mathcal{L}^*\right]$. Since $\overline{\mathcal{H}}^* v = 0$,

$$\overline{\mathcal{H}}^{-1} = \left(\overline{\mathcal{H}}^*\right)^\dagger + \left(\frac{b}{\sqrt{NT}}vv'\right)^\dagger = \left(\overline{\mathcal{H}}^*\right)^\dagger + \frac{\sqrt{NT}}{b\|vv'\|^2}vv' = \left(\overline{\mathcal{H}}^*\right)^\dagger + \frac{\sqrt{NT}}{b(N+T)^2}vv',$$

where † refers to the Moore-Penrose pseudo-inverse. Thus, if $\overline{\mathcal{H}}_1$ is the expected Hessian for $b = b_1 > 0$ and $\overline{\mathcal{H}}_2$ is the expected Hessian for $b = b_2 > 0$, $\left\| \overline{\mathcal{H}}_1^{-1} - \overline{\mathcal{H}}_2^{-1} \right\|_{\max} = \left\| \left( \frac{1}{b_1} - \frac{1}{b_2} \right) \frac{\sqrt{NT}}{(N+T)^2} vv' \right\|_{\max} = \mathcal{O}\left( (NT)^{-1/2} \right)$. ■

**Lemma D.10.** *Let Assumption 4.1 hold and let* $0 < b \leq b_{\min} \left( 1 + \frac{\max(N,T)}{\min(N,T)} \frac{b_{\max}}{b_{\min}} \right)^{-1}$. *Then,*

$$\left\| \overline{\mathcal{H}}_{(\alpha\alpha)}^{-1} \overline{\mathcal{H}}_{(\alpha\gamma)} \right\|_{\infty} < 1 - \frac{b}{b_{\max}}, \quad \textit{and} \quad \left\| \overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} \overline{\mathcal{H}}_{(\gamma\alpha)} \right\|_{\infty} < 1 - \frac{b}{b_{\max}}.$$

**Proof of Lemma D.10.** Let $h_{it} = \mathbb{E}_{\phi}(-\partial_{\pi^2} \ell_{it})$, and define

$$\tilde{h}_{it} = h_{it} - b - \frac{1}{b^{-1} + \sum_j \left( \sum_\tau h_{j\tau} \right)^{-1}} \sum_j \frac{h_{jt} - b}{\sum_\tau h_{j\tau}}.$$

By definition, $\overline{\mathcal{H}}_{(\alpha\alpha)} = \overline{\mathcal{H}}^*_{(\alpha\alpha)} + b 1_N 1_N' / \sqrt{NT}$ and $\overline{\mathcal{H}}_{(\alpha\gamma)} = \overline{\mathcal{H}}^*_{(\alpha\gamma)} - b 1_N 1_T' / \sqrt{NT}$. The matrix $\overline{\mathcal{H}}^*_{(\alpha\alpha)}$ is diagonal with elements $\sum_t h_{it} / \sqrt{NT}$. The matrix $\overline{\mathcal{H}}^*_{(\alpha\gamma)}$ has elements $h_{it} / \sqrt{NT}$. The Woodbury identity states that

$$\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1} = \overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1} - \overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1} 1_N \left( \sqrt{NT}\, b^{-1} + 1_N' \overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1} 1_N \right)^{-1} 1_N' \overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1}.$$

Then, $\overline{\mathcal{H}}_{(\alpha\alpha)}^{-1} \overline{\mathcal{H}}_{(\alpha\gamma)} = \overline{\mathcal{H}}_{(\alpha\alpha)}^{*-1} \tilde{H} / \sqrt{NT}$, where $\tilde{H}$ is the $N \times T$ matrix with elements $\tilde{h}_{it}$. Therefore

$$\left\| \overline{\mathcal{H}}_{(\alpha\alpha)}^{-1} \overline{\mathcal{H}}_{(\alpha\gamma)} \right\|_{\infty} = \max_i \frac{\sum_t \left| \tilde{h}_{it} \right|}{\sum_t h_{it}}.$$

Assumption 4.1($iv$) guarantees that $b_{\max} \geq h_{it} \geq b_{\min}$, which implies $h_{jt} - b \geq b_{\min} - b > 0$, and

$$\tilde{h}_{it} > h_{it} - b - \frac{1}{b^{-1}} \sum_j \frac{h_{jt} - b}{\sum_\tau h_{j\tau}} \geq b_{\min} - b \left( 1 + \frac{N}{T} \frac{b_{\max}}{b_{\min}} \right) \geq 0.$$

We conclude that

$$\begin{aligned} \left\| \overline{\mathcal{H}}_{(\alpha\alpha)}^{-1} \overline{\mathcal{H}}_{(\alpha\gamma)} \right\|_{\infty} &= \max_i \frac{\sum_t \tilde{h}_{it}}{\sum_t h_{it}} = 1 - \min_i \frac{1}{\sum_t h_{it}} \sum_t \left( b + \frac{1}{b^{-1} + \sum_j \left( \sum_\tau h_{j\tau} \right)^{-1}} \sum_j \frac{h_{jt} - b}{\sum_\tau h_{j\tau}} \right) \\ &< 1 - \frac{b}{b_{\max}}. \end{aligned}$$

Analogously, $\left\| \overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} \overline{\mathcal{H}}_{(\gamma\alpha)} \right\|_{\infty} < 1 - \frac{b}{b_{\max}}$. ■

**Proof of Lemma D.8.** We choose $b < b_{\min} \left( 1 + \max(\kappa^2, \kappa^{-2}) \frac{b_{\max}}{b_{\min}} \right)^{-1}$. Then, $b \leq b_{\min} \left( 1 + \frac{\max(N,T)}{\min(N,T)} \frac{b_{\max}}{b_{\min}} \right)^{-1}$ for large enough $N$ and $T$, so that Lemma D.10 becomes applicable. The choice of $b$ has no effect on the general validity of the lemma for all $b > 0$ by Lemma D.9.

By the inversion formula for partitioned matrices,

$$\overline{\mathcal{H}}^{-1} = \begin{pmatrix} A & -A\, \overline{\mathcal{H}}_{(\alpha\gamma)}\, \overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} \\ -\overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\, \overline{\mathcal{H}}_{(\gamma\alpha)}\, A & \overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} + \overline{\mathcal{H}}_{(\gamma\gamma)}^{-1}\, \overline{\mathcal{H}}_{(\gamma\alpha)}\, A\, \overline{\mathcal{H}}_{(\alpha\gamma)}\, \overline{\mathcal{H}}_{(\gamma\gamma)}^{-1} \end{pmatrix},$$

with $A := (\overline{\mathcal{H}}_{(\alpha\alpha)} - \overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)})^{-1}$. The Woodbury identity states that

$$\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)} = \overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)} - \underbrace{\overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)} 1_N \left(\sqrt{NT}/b + 1_N' \overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)} 1_N\right)^{-1} 1_N' \overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)}}_{=:C_{(\alpha\alpha)}},$$

$$\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)} = \overline{\mathcal{H}}^{*-1}_{(\gamma\gamma)} - \underbrace{\overline{\mathcal{H}}^{*-1}_{(\gamma\gamma)} 1_T \left(\sqrt{NT}/b + 1_T' \overline{\mathcal{H}}^{*-1}_{(\gamma\gamma)} 1_T\right)^{-1} 1_T' \overline{\mathcal{H}}^{*-1}_{(\gamma\gamma)}}_{=:C_{(\gamma\gamma)}}.$$

By Assumption 4.1($v$), $\|\overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)}\|_\infty = \mathcal{O}_P(1)$, $\|\overline{\mathcal{H}}^{*-1}_{(\gamma\gamma)}\|_\infty = \mathcal{O}_P(1)$, $\|\overline{\mathcal{H}}^{*}_{(\alpha\gamma)}\|_{\max} = \mathcal{O}_P(1/\sqrt{NT})$. Therefore[19]

$$\|C_{(\alpha\alpha)}\|_{\max} \leq \|\overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)}\|^2_\infty \, \|1_N 1_N'\|_{\max} \left(\sqrt{NT}/b + 1_N'\overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)} 1_N\right)^{-1} = \mathcal{O}_P(1/\sqrt{NT}),$$
$$\|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\|_\infty \leq \|\overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)}\|_\infty + N\|C_{(\alpha\alpha)}\|_{\max} = \mathcal{O}_P(1).$$

Analogously, $\|C_{(\gamma\gamma)}\|_{\max} = \mathcal{O}_P(1/\sqrt{NT})$ and $\|\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\|_\infty = \mathcal{O}_P(1)$. Furthermore, $\|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_{\max} \leq \|\overline{\mathcal{H}}^{*}_{(\alpha\gamma)}\|_{\max} + b/\sqrt{NT} = \mathcal{O}_P(1/\sqrt{NT})$. Define

$$B := \left(\mathbb{1}_N - \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}\right)^{-1} - \mathbb{1}_N = \sum_{n=1}^{\infty}\left(\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}\right)^n.$$

Then, $A = \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)} + \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}B = \overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)} - C_{(\alpha\alpha)} + \overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}B$. By Lemma D.10, $\|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}\|_\infty \leq \|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\|_\infty\|\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}\|_\infty < \left(1 - \frac{b}{b_{\max}}\right)^2 < 1$, and

$$\|B\|_{\max} \leq \sum_{n=0}^{\infty}\left(\|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\overline{\mathcal{H}}_{(\alpha\gamma)}\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\overline{\mathcal{H}}_{(\gamma\alpha)}\|_\infty\right)^n \|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\|_\infty\|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_\infty\|\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\|_\infty\|\overline{\mathcal{H}}_{(\gamma\alpha)}\|_{\max}$$
$$\leq \left[\sum_{n=0}^{\infty}\left(1 - \frac{b}{b_{\max}}\right)^{2n}\right] T\,\|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\|_\infty\|\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\|_\infty\|\overline{\mathcal{H}}_{(\gamma\alpha)}\|^2_{\max} = \mathcal{O}_P(1/\sqrt{NT}).$$

By the triangle inequality,

$$\|A\|_\infty \leq \|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\|_\infty + N\|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\|_\infty\|B\|_{\max} = \mathcal{O}_P(1).$$

Thus, for the different blocks of

$$\overline{\mathcal{H}}^{-1} - \begin{pmatrix} \overline{\mathcal{H}}^{*}_{(\alpha\alpha)} & 0 \\ 0 & \overline{\mathcal{H}}^{*}_{(\gamma\gamma)} \end{pmatrix}^{-1} = \begin{pmatrix} A - \overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)} & -A\,\overline{\mathcal{H}}_{(\alpha\gamma)}\,\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)} \\ -\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\,\overline{\mathcal{H}}_{(\gamma\alpha)}\,A & \overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\,\overline{\mathcal{H}}_{(\gamma\alpha)}\,A\,\overline{\mathcal{H}}_{(\alpha\gamma)}\,\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)} - C_{(\gamma\gamma)} \end{pmatrix},$$

we find

$$\begin{aligned}
\left\|A - \overline{\mathcal{H}}^{*-1}_{(\alpha\alpha)}\right\|_{\max} &= \left\|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}B - C_{(\alpha\alpha)}\right\|_{\max} \\
&\leq \|\overline{\mathcal{H}}^{-1}_{(\alpha\alpha)}\|_\infty\|B\|_{\max} - \|C_{(\alpha\alpha)}\|_{\max} = \mathcal{O}_P(1/\sqrt{NT}), \\
\left\|-A\,\overline{\mathcal{H}}_{(\alpha\gamma)}\,\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\right\|_{\max} &\leq \|A\|_\infty\|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_{\max}\|\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\|_\infty = \mathcal{O}_P(1/\sqrt{NT}), \\
\left\|\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\,\overline{\mathcal{H}}_{(\gamma\alpha)}\,A\,\overline{\mathcal{H}}_{(\alpha\gamma)}\,\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)} - C_{(\gamma\gamma)}\right\|_{\max} &\leq \|\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\|^2_\infty\|\overline{\mathcal{H}}_{(\gamma\alpha)}\|_\infty\|A\|_\infty\|\overline{\mathcal{H}}_{(\alpha\gamma)}\|_{\max} + \|C_{(\gamma\gamma)}\|_{\max} \\
&\leq N\|\overline{\mathcal{H}}^{-1}_{(\gamma\gamma)}\|^2_\infty\|A\|_\infty\|\overline{\mathcal{H}}_{(\alpha\gamma)}\|^2_{\max} + \|C_{(\gamma\gamma)}\|_{\max} = \mathcal{O}_P(1/\sqrt{NT}).
\end{aligned}$$

[19] Here and in the following me make use of the inequalities $\|AB\|_{\max} < \|A\|_\infty\|B\|_{\max}$, $\|AB\|_{\max} < \|A\|_{\max}\|B'\|_\infty$, $\|A\|_\infty \leq n\|A_{\max}\|$, which hold for any $m \times n$ matrix $A$ and $n \times p$ matrix $B$.

The bound $\mathcal{O}_P(1/\sqrt{NT})$ for the max-norm of each block of the matrix yields the same bound for the max-norm of the matrix itself. ■

### D.5 A Useful Algebraic Result

Let $\widetilde{\mathbb{P}}$ be the linear operator defined in equation (C.3), and and let $\mathbb{P}$ be the related projection operator defined in (C.2). Lemma D.11 shows how in the context of panel data models some expressions that appear in the general expansion of Appendix B can be conveniently expressed using the operator $\widetilde{\mathbb{P}}$. This lemma is used extensively in the proof of part *(ii)* of Theorem C.1.

**Lemma D.11.** *Let $A$, $B$ and $C$ be $N \times T$ matrices, and let the expected incidental parameter Hessian $\overline{\mathcal{H}}$ be invertible. Define the $N+T$ vectors $\mathcal{A}$ and $\mathcal{B}$ and the $(N+T)\times(N+T)$ matrix $\mathcal{C}$ as follows*[20]

$$\mathcal{A} = \frac{1}{NT}\begin{pmatrix} A1_T \\ A'1_N \end{pmatrix}, \qquad \mathcal{B} = \frac{1}{NT}\begin{pmatrix} B1_T \\ B'1_N \end{pmatrix}, \qquad \mathcal{C} = \frac{1}{NT}\begin{pmatrix} \operatorname{diag}(C1_T) & C \\ C' & \operatorname{diag}(C'1_N) \end{pmatrix}.$$

*Then,*

$$\textit{(i)} \qquad \mathcal{A}'\,\overline{\mathcal{H}}^{-1}\,\mathcal{B} = \frac{1}{\sqrt{NT}}\sum_{i,t}(\widetilde{\mathbb{P}}A)_{it}B_{it} = \frac{1}{\sqrt{NT}}\sum_{i,t}(\widetilde{\mathbb{P}}B)_{it}A_{it},$$

$$\textit{(ii)} \qquad \mathcal{A}'\,\overline{\mathcal{H}}^{-1}\,\mathcal{B} = \frac{1}{\sqrt{NT}}\sum_{i,t}\mathbb{E}_\phi(-\partial_{\pi^2}\ell_{it})(\widetilde{\mathbb{P}}A)_{it}(\widetilde{\mathbb{P}}B)_{it},$$

$$\textit{(iii)} \qquad \mathcal{A}'\,\overline{\mathcal{H}}^{-1}\,\mathcal{C}\,\overline{\mathcal{H}}^{-1}\,\mathcal{B} = \sum_{i,t}(\widetilde{\mathbb{P}}A)_{it}C_{it}(\widetilde{\mathbb{P}}B)_{it}.$$

**Proof.** Let $\tilde{\alpha}^*_i + \tilde{\gamma}^*_t = (\mathbb{P}\tilde{A})_{it} = (\widetilde{\mathbb{P}}A)_{it}$, with $\tilde{A}$ as defined in equation (C.3). The first order condition of the minimization problem in the definition of $(\mathbb{P}\tilde{A})_{it}$ can be written as $\frac{1}{\sqrt{NT}}\overline{\mathcal{H}}^*\binom{\tilde{\alpha}^*}{\tilde{\gamma}^*} = \mathcal{A}$. One solution to this equation is $\binom{\tilde{\alpha}^*}{\tilde{\gamma}^*} = \sqrt{NT}\,\overline{\mathcal{H}}^{-1}\mathcal{A}$ (this is the solution that imposes the normalization $\sum_i \tilde{\alpha}^*_i = \sum_t \tilde{\gamma}^*$, but this is of no importance in the following). Thus,

$$\mathcal{A}'\,\overline{\mathcal{H}}^{-1}\,\mathcal{B} = \begin{pmatrix}\tilde{\alpha}^* \\ \tilde{\gamma}^*\end{pmatrix}'\mathcal{B} = \frac{1}{\sqrt{NT}}\left[\sum_{i,t}\tilde{\alpha}^*_i B_{it} + \sum_{i,t}\tilde{\gamma}^*_t B_{it}\right] = \frac{1}{\sqrt{NT}}\sum_{i,t}(\widetilde{\mathbb{P}}A)_{it}B_{it}.$$

This gives the first equality of Statement $(i)$. The second equality of Statement $(i)$ follows by symmetry. Statement $(ii)$ is a special case of of Statement $(iii)$ with $\mathcal{C} = \frac{1}{\sqrt{NT}}\overline{\mathcal{H}}^*$, so we only need to prove Statement $(iii)$.

Let $\alpha^*_i + \gamma^*_t = (\mathbb{P}\tilde{B})_{it} = (\widetilde{\mathbb{P}}B)_{it}$, where $\tilde{B}_{it} = \frac{B_{it}}{\mathbb{E}_\phi(-\partial_{\pi^2}\ell_{it})}$. By an argument analogous to the one given above, we can choose $\binom{\alpha^*}{\gamma^*} = \sqrt{NT}\,\overline{\mathcal{H}}^{-1}\mathcal{B}$ as one solution to the minimization problem. Then,

$$\mathcal{A}'\,\overline{\mathcal{H}}^{-1}\,\mathcal{C}\,\overline{\mathcal{H}}^{-1}\,\mathcal{B} = \sum_{i,t}[\tilde{\alpha}^*_i C_{it}\alpha^*_i + \tilde{\alpha}^*_i C_{it}\gamma^*_t + \tilde{\gamma}^*_t C_{it}\alpha^*_i + \tilde{\gamma}^*_t C_{it}\gamma^*_t] = \sum_{i,t}(\widetilde{\mathbb{P}}A)_{it}C_{it}(\widetilde{\mathbb{P}}B)_{it}.$$

■

[20] Note that $A1_T$ is simply the $N$-vectors with entries $\sum_t A_{it}$ and $A'1_N$ is simply the $T$-vector with entries $\sum_i A_{it}$, and analogously for $B$ and $C$.

**Table 3: Finite sample properties of static probit estimators (N = 52)**

| | Coefficient | | | | | Average Effect | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Bias | Std. Dev. | RMSE | SE/SD | p; .95 | Bias | Std. Dev. | RMSE | SE/SD | p; .95 |
| *Design 1: correlated individual and time effects* | | | | | | | | | | |
| | | | | T = 14 | | | | | | |
| MLE-FETE | 13 | 12 | 17 | 0.88 | 0.76 | 1 | 8 | 8 | 0.93 | 0.93 |
| Analytical | 0 | 10 | 10 | 1.05 | 0.96 | -1 | 8 | 8 | 0.95 | 0.95 |
| Jackknife | -7 | 11 | 13 | 0.89 | 0.85 | 0 | 9 | 9 | 0.80 | 0.88 |
| | | | | T = 26 | | | | | | |
| MLE-FETE | 8 | 8 | 11 | 0.93 | 0.81 | 0 | 6 | 6 | 0.94 | 0.95 |
| Analytical | 0 | 7 | 7 | 1.03 | 0.95 | 0 | 6 | 6 | 0.95 | 0.95 |
| Jackknife | -3 | 7 | 8 | 0.97 | 0.91 | 0 | 6 | 6 | 0.89 | 0.92 |
| | | | | T = 52 | | | | | | |
| MLE-FETE | 5 | 5 | 7 | 0.98 | 0.83 | 0 | 4 | 4 | 0.99 | 0.94 |
| Analytical | 0 | 5 | 5 | 1.05 | 0.96 | 0 | 4 | 4 | 0.99 | 0.94 |
| Jackknife | -1 | 5 | 5 | 0.99 | 0.95 | 0 | 4 | 4 | 0.94 | 0.93 |
| *Design 2: uncorrelated individual and time effects* | | | | | | | | | | |
| | | | | T = 14 | | | | | | |
| MLE-FETE | 12 | 9 | 15 | 0.93 | 0.74 | 0 | 5 | 5 | 1.06 | 0.97 |
| Analytical | -1 | 8 | 8 | 1.11 | 0.97 | -1 | 5 | 5 | 1.08 | 0.97 |
| Jackknife | -7 | 9 | 11 | 0.94 | 0.84 | -1 | 6 | 7 | 0.83 | 0.90 |
| | | | | T = 26 | | | | | | |
| MLE-FETE | 7 | 6 | 10 | 0.93 | 0.75 | 0 | 4 | 4 | 0.98 | 0.95 |
| Analytical | 0 | 6 | 6 | 1.03 | 0.96 | 0 | 4 | 4 | 0.99 | 0.95 |
| Jackknife | -2 | 6 | 6 | 1.00 | 0.92 | 0 | 4 | 4 | 0.90 | 0.93 |
| | | | | T = 52 | | | | | | |
| MLE-FETE | 5 | 4 | 6 | 1.00 | 0.79 | 0 | 2 | 2 | 1.07 | 0.96 |
| Analytical | 0 | 4 | 4 | 1.07 | 0.97 | 0 | 2 | 2 | 1.07 | 0.96 |
| Jackknife | 0 | 4 | 4 | 1.04 | 0.96 | 0 | 2 | 2 | 1.00 | 0.94 |

Notes: All the entries are in percentage of the true parameter value. 500 repetitions. Data generated from the probit model: $Y_{it} = 1(\beta X_{it} + \alpha_i + \gamma_t > \varepsilon_{it})$, with $\varepsilon_{it}$ ~ i.i.d. N(0,1), $\alpha_i$ ~ i.i.d. N(0,1/16), $\gamma_t$ ~ i.i.d. N(0, 1/16) and $\beta = 1$. In design 1, $X_{it} = X_{i,t-1} / 2 + \alpha_i + \gamma_t + v_{it}$, $v_{it}$ ~ i.i.d. N(0, 1/2), and $X_{i0}$ ~ N(0,1). In design 2, $X_{it} = X_{i,t-1} / 2 + v_{it}$, $v_{it}$ ~ i.i.d. N(0, 3/4), and $X_{i0}$ ~ N(0,1), independent of $\alpha_i$ y $\gamma_t$ . Average effect is $\beta$ E[$\varphi(\beta X_{it} + \alpha_i + \gamma_t)$], where $\varphi()$ is the PDF of the standard normal distribution. MLE-FETE is the probit maximum likelihood estimator with individual and time fixed effects; Analytical is the bias corrected estimator that uses an analytical correction; and Jackknife is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension.

**Table 4: Finite sample properties of dynamic probit estimators (N = 52)**

| | Coefficient $Y_{i,t-1}$ | | | | | Average Effect $Y_{i,t-1}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Bias | Std. Dev. | RMSE | SE/SD | p; .95 | Bias | Std. Dev. | RMSE | SE/SD | p; .95 |
| *Design 1: correlated individual and time effects* | | | | | | | | | | |
| | | | | | T = 14 | | | | | |
| MLE-FETE | -44 | 30 | 53 | 0.96 | 0.67 | -52 | 26 | 58 | 0.96 | 0.43 |
| Analytical (L=1) | -5 | 26 | 26 | 1.10 | 0.96 | -6 | 27 | 28 | 0.90 | 0.91 |
| Analytical (L=2) | -4 | 28 | 28 | 1.03 | 0.95 | -4 | 29 | 30 | 0.85 | 0.90 |
| Jackknife | 12 | 33 | 35 | 0.88 | 0.89 | -4 | 33 | 33 | 0.76 | 0.85 |
| | | | | | T = 26 | | | | | |
| MLE-FETE | -23 | 21 | 30 | 0.98 | 0.79 | -29 | 19 | 35 | 0.98 | 0.65 |
| Analytical (L=1) | -4 | 19 | 19 | 1.05 | 0.96 | -3 | 20 | 20 | 0.96 | 0.94 |
| Analytical (L=2) | -1 | 20 | 20 | 1.02 | 0.96 | -1 | 21 | 21 | 0.92 | 0.93 |
| Jackknife | 2 | 22 | 22 | 0.93 | 0.94 | -1 | 23 | 23 | 0.85 | 0.91 |
| | | | | | T = 52 | | | | | |
| MLE-FETE | -9 | 14 | 17 | 0.99 | 0.90 | -14 | 14 | 20 | 0.98 | 0.82 |
| Analytical (L=1) | -1 | 13 | 13 | 1.04 | 0.95 | -1 | 14 | 14 | 0.97 | 0.94 |
| Analytical (L=2) | 0 | 14 | 14 | 1.02 | 0.95 | 1 | 15 | 15 | 0.96 | 0.94 |
| Jackknife | 1 | 14 | 14 | 0.98 | 0.94 | 0 | 15 | 15 | 0.91 | 0.92 |
| *Design 2: uncorrelated individual and time effects* | | | | | | | | | | |
| | | | | | T = 14 | | | | | |
| MLE-FETE | -38 | 28 | 47 | 0.94 | 0.66 | -46 | 24 | 52 | 0.95 | 0.45 |
| Analytical (L=1) | -5 | 24 | 25 | 1.07 | 0.97 | -6 | 25 | 26 | 0.91 | 0.92 |
| Analytical (L=2) | -4 | 26 | 26 | 1.01 | 0.95 | -4 | 26 | 27 | 0.86 | 0.89 |
| Jackknife | 9 | 31 | 32 | 0.85 | 0.89 | -3 | 31 | 31 | 0.75 | 0.84 |
| | | | | | T = 26 | | | | | |
| MLE-FETE | -19 | 19 | 27 | 0.97 | 0.80 | -26 | 18 | 31 | 0.96 | 0.67 |
| Analytical (L=1) | -4 | 17 | 18 | 1.05 | 0.95 | -4 | 18 | 18 | 0.95 | 0.93 |
| Analytical (L=2) | -2 | 18 | 18 | 1.02 | 0.95 | -2 | 19 | 19 | 0.92 | 0.93 |
| Jackknife | 1 | 19 | 19 | 0.94 | 0.94 | -1 | 20 | 20 | 0.84 | 0.90 |
| | | | | | T = 52 | | | | | |
| MLE-FETE | -8 | 13 | 15 | 0.98 | 0.90 | -12 | 12 | 17 | 0.98 | 0.82 |
| Analytical (L=1) | -1 | 12 | 12 | 1.03 | 0.95 | -1 | 12 | 13 | 0.98 | 0.94 |
| Analytical (L=2) | 0 | 12 | 12 | 1.01 | 0.94 | 0 | 13 | 13 | 0.96 | 0.94 |
| Jackknife | 0 | 13 | 13 | 0.98 | 0.95 | 0 | 13 | 13 | 0.93 | 0.92 |

Notes: All the entries are in percentage of the true parameter value. 500 repetitions. Data generated from the probit model: $Y_{it} = 1(\beta_Y Y_{i,t-1} + \beta_Z Z_{it} + \alpha_i + \gamma_t > \varepsilon_{it})$, with $Y_{i0} = 1(\beta_Z Z_{i0} + \alpha_i + \gamma_0 > \varepsilon_{i0})$, $\varepsilon_{it}$ ~ i.i.d. N(0,1), $\alpha_i$ ~ i.i.d. N(0,1/16), $\gamma_t$ ~ i.i.d. N(0, 1/16), $\beta_Y = 0.5$, and $\beta_Z = 1$. In design 1, $Z_{it} = Z_{i,t-1} / 2 + \alpha_i + \gamma_t + \nu_{it}$, $\nu_{it}$ ~ i.i.d. N(0, 1/2), and $Z_{i0}$ ~ N(0,1). In design 2, $Z_{it} = Z_{i,t-1} / 2 + \nu_{it}$, $\nu_{it}$ ~ i.i.d. N(0, 3/4), and $Z_{i0}$ ~ N(0,1), independent of $\alpha_i$ y $\gamma_t$. Average effect is $E[\Phi(\beta_Y + \beta_Z Z_{it} + \alpha_i + \gamma_t) - \Phi(\beta_z Z_{it} + \alpha_i + \gamma_t)]$, where $\Phi()$ is the CDF of the standard normal distribution. MLE-FETE is the probit maximum likelihood estimator with individual and time fixed effects; Analytical (L = l) is the bias corrected estimator that uses an analytical correction with l lags to estimate the spectral expectations; and Jackknife is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension.

**Table 5: Finite sample properties of dynamic probit estimators (N = 52)**

| | Coefficient $Z_{it}$ | | | | | Average Effect $Z_{it}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Bias | Std. Dev. | RMSE | SE/SD | p; .95 | Bias | Std. Dev. | RMSE | SE/SD | p; .95 |
| *Design 1: correlated individual and time effects* | | | | | | | | | | |
| | | | | | T = 14 | | | | | |
| MLE-FETE | 20 | 13 | 23 | 0.86 | 0.59 | 4 | 10 | 10 | 0.86 | 0.90 |
| Analytical (L=1) | 2 | 11 | 11 | 1.06 | 0.97 | 1 | 9 | 9 | 0.88 | 0.93 |
| Analytical (L=2) | 2 | 11 | 11 | 1.05 | 0.97 | 1 | 10 | 10 | 0.87 | 0.93 |
| Jackknife | -9 | 14 | 16 | 0.81 | 0.81 | 3 | 11 | 12 | 0.74 | 0.86 |
| | | | | | T = 26 | | | | | |
| MLE-FETE | 10 | 8 | 13 | 0.94 | 0.74 | 2 | 7 | 7 | 0.92 | 0.92 |
| Analytical (L=1) | 0 | 7 | 7 | 1.06 | 0.96 | 1 | 7 | 7 | 0.93 | 0.94 |
| Analytical (L=2) | 0 | 7 | 7 | 1.06 | 0.96 | 1 | 7 | 7 | 0.93 | 0.94 |
| Jackknife | -3 | 8 | 8 | 0.97 | 0.91 | 1 | 7 | 7 | 0.86 | 0.91 |
| | | | | | T = 52 | | | | | |
| MLE-FETE | 6 | 5 | 8 | 0.94 | 0.75 | 1 | 5 | 5 | 0.94 | 0.92 |
| Analytical (L=1) | 0 | 5 | 5 | 1.01 | 0.96 | 0 | 5 | 5 | 0.94 | 0.92 |
| Analytical (L=2) | 0 | 5 | 5 | 1.01 | 0.96 | 0 | 5 | 5 | 0.94 | 0.92 |
| Jackknife | -1 | 5 | 5 | 0.99 | 0.94 | 0 | 5 | 5 | 0.94 | 0.93 |
| *Design 2: uncorrelated individual and time effects* | | | | | | | | | | |
| | | | | | T = 14 | | | | | |
| MLE-FETE | 17 | 10 | 20 | 0.92 | 0.58 | 3 | 6 | 6 | 1.07 | 0.93 |
| Analytical (L=1) | 1 | 8 | 8 | 1.13 | 0.97 | 0 | 6 | 6 | 1.08 | 0.97 |
| Analytical (L=2) | 1 | 8 | 8 | 1.12 | 0.97 | 0 | 6 | 6 | 1.08 | 0.96 |
| Jackknife | -8 | 11 | 14 | 0.84 | 0.82 | 2 | 7 | 8 | 0.84 | 0.90 |
| | | | | | T = 26 | | | | | |
| MLE-FETE | 10 | 7 | 12 | 0.92 | 0.68 | 2 | 4 | 5 | 1.03 | 0.94 |
| Analytical (L=1) | 1 | 6 | 6 | 1.03 | 0.96 | 0 | 4 | 4 | 1.03 | 0.96 |
| Analytical (L=2) | 0 | 6 | 6 | 1.04 | 0.96 | 0 | 4 | 4 | 1.03 | 0.96 |
| Jackknife | -3 | 6 | 7 | 0.98 | 0.90 | 0 | 5 | 5 | 0.94 | 0.93 |
| | | | | | T = 52 | | | | | |
| MLE-FETE | 6 | 5 | 7 | 0.92 | 0.74 | 1 | 3 | 3 | 1.01 | 0.93 |
| Analytical (L=1) | 0 | 4 | 4 | 0.99 | 0.95 | 0 | 3 | 3 | 1.01 | 0.94 |
| Analytical (L=2) | 0 | 4 | 4 | 0.99 | 0.95 | 0 | 3 | 3 | 1.01 | 0.95 |
| Jackknife | -1 | 4 | 5 | 0.95 | 0.94 | 0 | 3 | 3 | 0.95 | 0.94 |

Notes: All the entries are in percentage of the true parameter value. 500 repetitions. Data generated from the probit model: $Y_{it} = 1(\beta_Y Y_{i,t-1} + \beta_Z Z_{it} + \alpha_i + \gamma_t > \varepsilon_{it})$, with $Y_{i0} = 1(\beta_Z Z_{i0} + \alpha_i + \gamma_0 > \varepsilon_{i0})$, $\varepsilon_{it}$ ~ i.i.d. N(0,1), $\alpha_i$ ~ i.i.d. N(0,1/16), $\gamma_t$ ~ i.i.d. N(0, 1/16), $\beta_Y = 0.5$, and $\beta_Z = 1$. In design 1, $Z_{it} = Z_{i,t-1} / 2 + \alpha_i + \gamma_t + v_{it}$, $v_{it}$ ~ i.i.d. N(0, 1/2), and $Z_{i0}$ ~ N(0,1). In design 2, $Z_{it} = Z_{i,t-1} / 2 + v_{it}$, $v_{it}$ ~ i.i.d. N(0, 3/4), and $Z_{i0}$ ~ N(0,1), independent of $\alpha_i$ y $\gamma_t$. Average effect is $\beta_Z$ E[$\varphi(\beta_Y Y_{i,t-1} + \beta_Z Z_{it} + \alpha_i + \gamma_t)$], where φ() is the PDF of the standard normal distribution. MLE-FETE is the probit maximum likelihood estimator with individual and time fixed effects; Analytical (L = l) is the bias corrected estimator that uses an analytical correction with l lags to estimate the spectral expectations; and Jackknife is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension.

**Table 6: Finite sample properties of static Poisson estimators**

| | Coefficient $Z_{it}$ | | | | | Coefficient $Z_{it}^2$ | | | | | Average Effect $Z_{it}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Bias | Std. Dev. | RMSE | SE/SD | p; .95 | Bias | Std. Dev. | RMSE | SE/SD | p; .95 | Bias | Std. Dev. | RMSE | SE/SD | p; .95 |
| | | | | | | N = 17, T = 22, unbalanced | | | | | | | | | |
| MLE | -59 | 14 | 60 | 1.04 | 0.01 | -58 | 14 | 60 | 1.03 | 0.01 | 222 | 113 | 248 | 1.15 | 0.60 |
| MLE-TE | -62 | 14 | 64 | 1.01 | 0.01 | -62 | 14 | 64 | 1.01 | 0.01 | -9 | 139 | 139 | 1.04 | 0.94 |
| MLE-FETE | -2 | 17 | 17 | 1.02 | 0.96 | -2 | 17 | 17 | 1.02 | 0.96 | -15 | 226 | 226 | 1.49 | 1.00 |
| Analytical (L=1) | -1 | 17 | 17 | 1.02 | 0.96 | -1 | 17 | 17 | 1.02 | 0.96 | -9 | 225 | 225 | 1.50 | 1.00 |
| Analytical (L=2) | -1 | 17 | 17 | 1.02 | 0.96 | -1 | 17 | 17 | 1.02 | 0.96 | -6 | 225 | 225 | 1.50 | 1.00 |
| Jackknife | -3 | 25 | 25 | 0.69 | 0.83 | -3 | 25 | 25 | 0.70 | 0.83 | -15 | 333 | 333 | 1.01 | 0.95 |
| | | | | | | N = 34, T = 22, unbalanced | | | | | | | | | |
| MLE | -58 | 10 | 59 | 1.03 | 0.00 | -57 | 10 | 58 | 1.03 | 0.00 | 226 | 81 | 240 | 0.98 | 0.20 |
| MLE-TE | -61 | 10 | 62 | 1.00 | 0.00 | -61 | 10 | 62 | 1.00 | 0.00 | -3 | 97 | 97 | 0.95 | 0.94 |
| MLE-FETE | 0 | 12 | 12 | 0.99 | 0.96 | 0 | 13 | 13 | 0.99 | 0.96 | -6 | 158 | 158 | 1.12 | 0.98 |
| Analytical (L=1) | 0 | 12 | 12 | 0.99 | 0.96 | 0 | 13 | 13 | 0.99 | 0.96 | 0 | 159 | 158 | 1.11 | 0.98 |
| Analytical (L=2) | 1 | 13 | 13 | 0.99 | 0.96 | 1 | 13 | 13 | 0.99 | 0.96 | 3 | 159 | 159 | 1.11 | 0.98 |
| Jackknife | -1 | 14 | 14 | 0.90 | 0.93 | -1 | 14 | 14 | 0.90 | 0.93 | -15 | 208 | 208 | 0.85 | 0.90 |
| | | | | | | N = 51, T = 22, unbalanced | | | | | | | | | |
| MLE | -58 | 8 | 58 | 1.00 | 0.00 | -57 | 8 | 57 | 1.00 | 0.00 | 228 | 66 | 238 | 0.96 | 0.06 |
| MLE-TE | -61 | 8 | 61 | 1.00 | 0.00 | -61 | 8 | 61 | 1.00 | 0.00 | -1 | 77 | 77 | 0.95 | 0.94 |
| MLE-FETE | 0 | 10 | 10 | 0.97 | 0.94 | 0 | 11 | 11 | 0.97 | 0.94 | -4 | 128 | 128 | 1.04 | 0.96 |
| Analytical (L=1) | 0 | 10 | 10 | 0.97 | 0.94 | 0 | 11 | 11 | 0.97 | 0.94 | 2 | 129 | 128 | 1.04 | 0.96 |
| Analytical (L=2) | 1 | 10 | 11 | 0.96 | 0.94 | 1 | 11 | 11 | 0.96 | 0.94 | 5 | 129 | 129 | 1.04 | 0.96 |
| Jackknife | 0 | 11 | 11 | 0.90 | 0.93 | 0 | 11 | 11 | 0.90 | 0.94 | -12 | 169 | 170 | 0.79 | 0.88 |

Notes: All the entries are in percentage of the true parameter value. 500 repetitions. The data generating process is: $Y_{it} \sim \text{Poisson}(\exp\{\beta_1 X_{it} + \beta_2 X_{it}^2 + \alpha_i + \gamma_t\})$ with all the variables and coefficients calibrated to the dataset of ABBGH. Average effect is $E[(\beta_1 + 2\beta_2 X_{it})\exp(\beta_1 X_{it} + \beta_2 X_{it}^2 + \alpha_i + \gamma_t)]$. MLE is the Poisson maximum likelihood estimator without individual and time fixed effects; MLE-TE is the Poisson maximum likelihood estimator with time fixed effects; MLE-FETE is the Poisson maximum likelihood estimator with individual and time fixed effects;Analytical (L = l) is the bias corrected estimator that uses an analytical correction with l lags to estimate the spectral expectations; and Jackknife is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension.

**Table 7: Finite sample properties of dynamic Poisson estimators**

| | Coefficient $Y_{i,t-1}$ | | | | | Average Effect $Y_{i,t-1}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Bias | Std. Dev. | RMSE | SE/SD | p; .95 | Bias | Std. Dev. | RMSE | SE/SD | p; .95 |
| | | | | | N = 17,T = 21, unbalanced | | | | | |
| MLE | 135 | 3 | 135 | 1.82 | 0.00 | 158 | 2 | 158 | 3.75 | 0.00 |
| MLE-TE | 142 | 3 | 142 | 1.95 | 0.00 | 163 | 3 | 163 | 4.17 | 0.00 |
| MLE-FETE | -17 | 15 | 23 | 0.96 | 0.78 | -17 | 15 | 22 | 1.38 | 0.89 |
| Analytical (L=1) | -7 | 15 | 17 | 0.98 | 0.91 | -8 | 14 | 16 | 1.41 | 0.97 |
| Analytical (L=2) | -5 | 15 | 16 | 0.96 | 0.92 | -5 | 15 | 16 | 1.38 | 0.98 |
| Jackknife | 4 | 20 | 21 | 0.73 | 0.85 | 4 | 20 | 20 | 1.03 | 0.95 |
| | | | | | N = 34, T = 21, unbalanced | | | | | |
| MLE | 135 | 2 | 135 | 1.76 | 0.00 | 158 | 2 | 158 | 2.82 | 0.00 |
| MLE-TE | 141 | 2 | 141 | 1.77 | 0.00 | 162 | 2 | 162 | 2.69 | 0.00 |
| MLE-FETE | -16 | 11 | 19 | 0.93 | 0.65 | -16 | 10 | 19 | 1.05 | 0.71 |
| Analytical (L=1) | -7 | 11 | 13 | 0.95 | 0.89 | -7 | 10 | 12 | 1.08 | 0.92 |
| Analytical (L=2) | -4 | 11 | 12 | 0.93 | 0.91 | -4 | 10 | 11 | 1.05 | 0.94 |
| Jackknife | 3 | 13 | 14 | 0.77 | 0.85 | 3 | 13 | 13 | 0.86 | 0.89 |
| | | | | | N = 51, T = 21, unbalanced | | | | | |
| MLE | 135 | 2 | 135 | 1.81 | 0.00 | 158 | 1 | 158 | 2.58 | 0.00 |
| MLE-TE | 141 | 2 | 141 | 1.79 | 0.00 | 162 | 2 | 162 | 2.41 | 0.00 |
| MLE-FETE | -15 | 8 | 17 | 0.97 | 0.55 | -15 | 8 | 17 | 1.03 | 0.55 |
| Analytical (L=1) | -6 | 8 | 10 | 0.99 | 0.90 | -6 | 8 | 10 | 1.05 | 0.91 |
| Analytical (L=2) | -3 | 8 | 9 | 0.97 | 0.93 | -4 | 8 | 9 | 1.03 | 0.93 |
| Jackknife | 3 | 11 | 11 | 0.77 | 0.87 | 3 | 10 | 11 | 0.80 | 0.88 |

Notes: All the entries are in percentage of the true parameter value. 500 repetitions. The data generating process is: $Y_{it} \sim \text{Poisson}(\exp\{\beta_Y \log(1 + Y_{i,t-1}) + \beta_1 Z_{it} + \beta_2 Z_{it}^2 + \alpha_i + \gamma_t\})$, where all the exogenous variables, initial condition and coefficients are calibrated to the application of ABBGH. Average effect is $\beta_Y E[\exp\{((\beta_Y - 1)\log(1 + Y_{i,t-1}) + \beta_1 Z_{it} + \beta_2 Z_{it}^2 + \alpha_i + \gamma_t\}]$. MLE is the Poisson maximum likelihood estimator without individual and time fixed effects; MLE-TE is the Poisson maximum likelihood estimator with time fixed effects; MLE-FETE is the Poisson maximum likelihood estimator with individual and time fixed effects; Analytical (L = l) is the bias corrected estimator that uses an analytical correction with l lags to estimate the spectral expectations; and Jackknife is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension.

**Table 8: Finite sample properties of dynamic Poisson estimators**

| | Coefficient $Z_{it}$ | | | | | Coefficient $Z_{it}^2$ | | | | | Average Effect $Z_{it}$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Bias | Std. Dev. | RMSE | SE/SD | p; .95 | Bias | Std. Dev. | RMSE | SE/SD | p; .95 | Bias | Std. Dev. | RMSE | SE/SD | p; .95 |
| | | | | | | N = 17, T = 21, unbalanced | | | | | | | | | |
| MLE | -76 | 27 | 81 | 1.13 | 0.29 | -76 | 27 | 80 | 1.13 | 0.30 | 760 | 351 | 837 | 1.65 | 0.89 |
| MLE-TE | -65 | 28 | 71 | 1.12 | 0.44 | -65 | 29 | 71 | 1.12 | 0.45 | 541 | 356 | 647 | 1.75 | 0.99 |
| MLE-FETE | 9 | 40 | 41 | 0.95 | 0.92 | 9 | 41 | 42 | 0.95 | 0.92 | -3 | 1151 | 1150 | 1.08 | 0.99 |
| Analytical (L=1) | 4 | 40 | 40 | 0.97 | 0.94 | 4 | 40 | 40 | 0.97 | 0.94 | 11 | 1117 | 1116 | 1.11 | 0.99 |
| Analytical (L=2) | 3 | 39 | 39 | 0.97 | 0.94 | 3 | 40 | 40 | 0.97 | 0.94 | 15 | 1110 | 1109 | 1.12 | 0.99 |
| Jackknife | 3 | 57 | 57 | 0.68 | 0.82 | 3 | 57 | 57 | 0.68 | 0.81 | 24 | 1653 | 1651 | 0.75 | 0.86 |
| | | | | | | N = 34, T = 21, unbalanced | | | | | | | | | |
| MLE | -75 | 19 | 77 | 1.18 | 0.04 | -74 | 19 | 77 | 1.18 | 0.05 | 777 | 252 | 817 | 1.47 | 0.42 |
| MLE-TE | -65 | 19 | 67 | 1.18 | 0.15 | -64 | 19 | 67 | 1.18 | 0.15 | 534 | 248 | 589 | 1.65 | 0.88 |
| MLE-FETE | 6 | 28 | 28 | 0.97 | 0.94 | 6 | 28 | 29 | 0.97 | 0.94 | -68 | 734 | 736 | 1.03 | 0.94 |
| Analytical (L=1) | 2 | 27 | 27 | 0.99 | 0.95 | 2 | 28 | 28 | 0.99 | 0.95 | -51 | 713 | 714 | 1.06 | 0.95 |
| Analytical (L=2) | 0 | 27 | 27 | 0.99 | 0.95 | 0 | 27 | 27 | 1.00 | 0.95 | -47 | 706 | 707 | 1.07 | 0.95 |
| Jackknife | 2 | 31 | 31 | 0.87 | 0.92 | 2 | 31 | 31 | 0.87 | 0.92 | -38 | 1012 | 1012 | 0.74 | 0.85 |
| | | | | | | N = 51, T = 21, unbalanced | | | | | | | | | |
| MLE | -74 | 15 | 76 | 1.17 | 0.00 | -73 | 15 | 75 | 1.17 | 0.00 | 768 | 201 | 794 | 1.48 | 0.18 |
| MLE-TE | -63 | 16 | 65 | 1.15 | 0.05 | -63 | 16 | 65 | 1.15 | 0.05 | 535 | 197 | 570 | 1.68 | 0.74 |
| MLE-FETE | 8 | 22 | 23 | 1.01 | 0.93 | 8 | 22 | 24 | 1.01 | 0.93 | -27 | 606 | 606 | 0.99 | 0.95 |
| Analytical (L=1) | 4 | 21 | 22 | 1.02 | 0.95 | 4 | 22 | 22 | 1.02 | 0.95 | -11 | 588 | 587 | 1.02 | 0.96 |
| Analytical (L=2) | 2 | 21 | 21 | 1.03 | 0.95 | 2 | 22 | 22 | 1.03 | 0.95 | -5 | 581 | 580 | 1.03 | 0.96 |
| Jackknife | 3 | 25 | 25 | 0.89 | 0.91 | 4 | 25 | 25 | 0.89 | 0.91 | 8 | 838 | 837 | 0.71 | 0.83 |

Notes: All the entries are in percentage of the true parameter value. 500 repetitions. The data generating process is: $Y_{it} \sim \text{Poisson}(\exp\{\beta_Y \log(1 + Y_{i,t-1}) + \beta_1 Z_{it} + \beta_2 Z_{it}^2 + \alpha_i + \gamma_t\})$, where all the exogenous variables, initial condition and coefficients are calibrated to the application of ABBGH. Average effect is $E[(\beta_1 + 2\beta_2 Z_{it}) \exp\{\beta_Y \log(1 + Y_{i,t-1}) + \beta_1 Z_{it} + \beta_2 Z_{it}^2 + \alpha_i + \gamma_t\}]$. MLE is the Poisson maximum likelihood estimator without individual and time fixed effects; MLE-TE is the Poisson maximum likelihood estimator with time fixed effects; MLE-FETE is the Poisson maximum likelihood estimator with individual and time fixed effects; Analytical (L = l) is the bias corrected estimator that uses an analytical correction with l lags to estimate the spectral expectations; and Jackknife is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension.

**Table 9: Poisson model for patents**

| Dependent variable: citation-weighted patents | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| *Static model* | | | | | | |
| Competition | 165.12 | 152.81 | 387.46 | 389.99 | 401.88 | 401.51 |
| | (54.77) | (55.74) | (67.74) | | | |
| | *-20.00* | *-6.43* | *-5.98* | *-5.49* | *-6.25* | *-4.74* |
| | *(7.74)* | *(8.61)* | *(19.68)* | | | |
| Competition squared | -88.55 | -80.99 | -204.55 | -205.84 | -212.15 | -214.03 |
| | (29.08) | (29.61) | (36.17) | | | |
| *Dynamic model* | | | | | | |
| Lag-patents | 1.05 | 1.07 | 0.46 | 0.48 | 0.50 | 0.70 |
| | (0.02) | (0.03) | (0.05) | | | |
| | *0.86* | *0.87* | *0.36* | *0.38* | *0.39* | *0.56* |
| | *(0.02)* | *(0.03)* | *(0.07)* | | | |
| Competition | 62.95 | 95.70 | 199.68 | 184.70 | 184.64 | 255.44 |
| | (62.68) | (65.08) | (76.66) | | | |
| | *-12.78* | *-9.03* | *-1.68* | *-0.15* | *-0.43* | *-18.45* |
| | *(7.54)* | *(8.18)* | *(15.53)* | | | |
| Competition squared | -34.15 | -51.09 | -105.24 | -97.23 | -97.22 | -136.97 |
| | (33.21) | (34.48) | (40.87) | | | |
| Year effects | | Yes | Yes | Yes | Yes | Yes |
| Industry effects | | | Yes | Yes | Yes | Yes |
| Bias correction | | | | A | A | J |
| (number of lags) | | | | 1 | 2 | |

Notes: Data set obtained from ABBGH. Competition is measured by (1-Lerner index) in the industry-year. All columns are estimated using an unbalanced panel of seventeen industries over the period 1973 to 1994. First year available used as initial condition in dynamic model. The estimates of the coefficients for the static model in columns (2) and (3) replicate the results in ABBGH. A is the bias corrected estimator that uses an analytical correction with a number lags to estimate the spectral expectations specified at the bottom cell. J is the jackknife bias corrected estimator that uses split panel jackknife in both the individual and time dimensions. Standard errors in parentheses and average partial effects in italics.