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

Iván Fernández-Val — BU

Martin Weidner — UCL

November 2, 2011
Preliminary and incomplete

Abstract

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

1 Introduction

Fixed effects estimators of panel models can be severely biased because of the well-known incidental parameters problem (Neyman and Scott (1948), Heckman (1981), Lancaster (2000), and Greene (2004)). A recent literature, surveyed in Arellano and Hahn (2007) and including Phillips and Moon (1999), Hahn and Kuersteiner (2002), Lancaster (2002), Woutersen (2002), Hahn and Kuersteiner (2011), Hahn and Newey (2004), Carro (2007), and Fernandez-Val (2009), provides a range of solutions, so-called large-T bias corrections, to reduce the incidental parameters problem in long panels. These papers derive the analytical expression of the bias (up to a certain order of the time dimension T), which can be employed to adjust the biased fixed effects estimators. While the existing large-T methods cover a large class of models with individual effects, they do not apply to panel models with individual and time effects. Time effects are important for economic modeling because they allow the researcher to control for aggregate common shocks.

We develop analytical and jackknife bias corrections for nonlinear models with *both* individual and time effects. We consider asymptotics where T grows with the 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 applications that use U.S. state or country level panel data. Under these asymptotics, the incidental parameter problem becomes a finite-sample bias problem in the time dimension and the presence of time effects introduce additional sources of finite-sample bias in the cross sectional dimension. As the existing bias expressions apply to models with only individual effects, we derive the appropriate correction. This correction does not correspond to a sequential application of the existing corrections to each dimension. In addition to model parameters, we provide bias corrections for average partial effects, which are functions of the data, parameters and individual and time effects in nonlinear models. These effects are often the ultimate quantities of interest.

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) derive similar expansions for quasi-likelihood estimators of linear models with interactive individual and time effects. We consider non-linear single index models with additive or interactive individual and time effects. In our case, the nonlinearity of the model introduces nonseparability between the estimators of the model parameters and incidental parameters. Moreover, even in the additive case, we need to deal with an infinite dimensional non-diagonal Hessian matrix for the incidental parameters. We focus on M-estimators with concave objective functions in the index, which cover fixed effects estimators of the most popular limited dependent variable models such as logit, probit, Tobit and Poisson models (Olsen (1978), and Pratt (1981)). Our analysis therefore extends the use of large-T bias adjustments to an important class of models.

Our corrections eliminate the leading term of the bias from the asymptotic expansions. Under asymptotic sequences where N and T grow at the same rate, 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 leave one observation out and split panel jackknife methods of Hahn and Newey (2004) and Dhaene and Jochmans (2010). However, the theory of the previous papers do not cover the models that we consider because they assume either identical distribution or stationarity of the observed variables over the time series dimension, conditional on the unobserved effects. These assumptions are violated in our case due to the presence of the time effects. We therefore extend the validity of the bias corrections to heterogenous processes in multiple dimensions under weak time series dependence conditions. The corrections can be implemented over the objective function, score or estimator. Simulation evidence indicates our approach works well in finite samples and an empirical example illustrates the applicability of our estimator.

The large-T panel literature on models with individual and time effects is sparse. Hahn and Moon (2006) consider bias corrected fixed effects estimators in panel linear AR(1) models with additive individual and time effects. Charbonneau (2011) extends 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. This conditional approach completely eliminates asymptotic bias, but does not permit estimation of average partial effects and has not been developed for models with interactive effects. We consider estimators of model parameters and average partial effects in nonlinear models with predetermined regressors and additive or interactive individual and time effects.

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

2 Model and Estimators

The data consist of $N \times T$ observations $Y = \{Y_{it} : i = 1, ..., N; t = 1, ..., T\}$ and $X = \{X_{it} : i = 1, ..., N; t = 1, ..., T\}$, for an 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:

Assumption 2.1 (Model). *For $X_i^t = \{X_{is} : s = 1, ..., t\}$ and $\gamma^t = (\gamma_s : s = 1, ..., t)$,*

$$Y_{it} \mid X_i^t, \alpha_i, \gamma^t \overset{d}{=} Y_{it} \mid X_{it}, \alpha_i, \gamma_t \sim f_{Y|X}(\cdot \mid X_{it}, \alpha_i, \gamma_t, \beta), \quad i = 1, ..., N; t = 1, ..., T;$$

where $f_{Y|X}$ is a known probability function and β is a finite dimensional parameter vector.

The variables α_i and γ_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 $f_{Y|X}$ 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 illustrative examples throughout the analysis:

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

$$f_{Y|X}(y \mid X_{it}, \alpha_i, \gamma_t, \beta) = F(X_{it}'\beta + g(\alpha_i, \gamma_t))^y [1 - F(X_{it}'\beta + g(\alpha_i, \gamma_t))]^{1-y}, \quad y \in \{0, 1\},$$

where g is a known function.

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

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

where g *is a known function.*

The leading specifications for the function g are the additive effects with $g(\alpha_i, \gamma_t) = \alpha_i + \gamma_t$, and the interactive effects with $g(\alpha_i, \gamma_t) = \alpha_i \gamma_t$.

For estimation, we adopt a fixed effects approach treating 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 β usually includes regression coefficients of interest, while the unobserved effects ϕ_{NT} are treated as nuisance parameters. The true value of the parameters, denoted by β^0 and $\phi_{NT}^0 = (\alpha_1^0, ..., \alpha_N^0, \gamma_1^0, ..., \gamma_T^0)'$, are a solution to the population problem

$$\max_{\beta, \phi_{NT}} L_{NT}^*(\beta, \phi_{NT}), \quad L_{NT}^*(\beta, \phi_{NT}) := (NT)^{-1} \sum_{i,t} \mathbb{E}_t \left[\log f_{Y|X}(Y_{it} \mid X_{it}, \alpha_i, \gamma_t, \beta) \right], \tag{2.1}$$

for every N, T. Here $\mathbb{E}_t$ denotes the expectation with respect to the true conditional distribution $f_{Y|X}(\cdot \mid X_{it}, \alpha_i^0, \gamma_t^0, \beta^0)$. We also use below the notation $\mathbb{E}$ to denote the expectation over both Y_{it} and X_{it}, conditional on the unobserved effects ϕ_{NT}^0 and parameter β^0.[1] The solution to the problem (2.1) is often not unique in ϕ_{NT}. In the description of the estimator below, we propose several normalizations that identify ϕ_{NT}^0.

Other quantities of interest involve averages over the data and nuisance parameters

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

These include average partial effects, which are often the ultimate quantities of interest in nonlinear models. Here are some examples of these effects motivated by the numerical examples of Sections 6 and 7.

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

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

[1] Since the inference is conditional on the realization of the unobserved effects, all the probability statements should be qualified with a.s. We omit this qualifier for notational convenience.

where β_k is the kth element of β, and $X_{it,-k}$ and β_{-k} include all elements of X_{it} and β except for the kth 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} can be approximated by the derivative

$$\Delta(X_{it}, \beta, \phi_{NT}) = \beta_k f(X_{it}'\beta + g(\alpha_i, \gamma_t)), \tag{2.4}$$

where f is the derivative of F.

Example 2 (cont.) *If X_{it} includes Z_{it} and $H(Z_{it})$ with coefficients β_k and β_j, the partial effect of Z_{it} on the conditional expectation of Y_{it} can be approximated by*

$$\Delta(X_{it}, \beta, \phi_{NT}) = [\beta_k + \beta_j h(Z_{it})] \exp(X_{it}'\beta + g(\alpha_i, \gamma_t)), \tag{2.5}$$

where h is the derivative of H.

The parameters are estimated by solving the sample version of problem (2.1), i.e.

$$\max_{\beta, \phi_{NT}} \mathcal{L}^*_{NT}(\beta, \phi_{NT}), \quad \mathcal{L}^*_{NT}(\beta, \phi_{NT}) := (NT)^{-1} \sum_{i,t} \log f_{Y|X}(Y_{it} \mid X_{it}, \alpha_i, \gamma_t, \beta), \tag{2.6}$$

Depending on the specification of the unobserved effects, there might be ambiguity in the solution for ϕ_{NT}. For instance, adding a constant ρ to all α_i, while subtracting it from all γ_t, does not change $\alpha_i + \gamma_t$. This implies that the objective function $\mathcal{L}_{NT}$ has a singular direction with respect to ϕ_{NT} at $v_{NT} = (1_N', -1_T')'$, i.e. $\mathcal{L}^*_{NT}(\beta, \phi_{NT}) = \mathcal{L}^*_{NT}(\beta, \phi_{NT} + \rho v_{NT})$. To eliminate this ambiguity, we normalize ϕ_{NT} to satisfy $v_{NT}'\phi_{NT} = 0$, i.e. $\sum_i \alpha_i = \sum_t \gamma_t$, by modifying the objective function

$$\mathcal{L}_{NT}(\beta, \phi_{NT}) = \mathcal{L}^*_{NT}(\beta, \phi_{NT}) - b(v_{NT}'\phi_{NT})^2/(2NT), \tag{2.7}$$

where $b > 0$ is an arbitrary constant. The maximizer of $\mathcal{L}^*_{NT}(\beta, \phi_{NT})$ automatically satisfies $v_{NT}'\phi_{NT} = 0$, and we choose $v_{NT}'\phi^0_{NT} = 0$ for the true value. A similar issue arises in the interactive model. Here, we use the modification

$$\mathcal{L}_{NT}(\beta, \phi_{NT}) = \mathcal{L}^*_{NT}(\beta, \phi_{NT})(NT)^{-1} - b[v_{NT}'(\phi_{NT} \odot \phi_{NT})]^2/(8NT), \tag{2.8}$$

where $b > 0$ is an arbitrary constant and $\odot$ denotes Hadamard (entry-wise) product. The maximizer of $\mathcal{L}_{NT}(\beta, \phi_{NT})$ automatically satisfies $v_{NT}'(\phi_{NT} \odot \phi_{NT}) = 0$, and we choose $v_{NT}'(\phi^0_{NT} \odot \phi^0_{NT}) = 0$ for the true value, i.e. $\sum_i (\alpha_i^0)^2 = \sum_t (\gamma_t^0)^2$.

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

$$\hat{\phi}_{NT}(\beta) \in \operatorname*{argmax}_{\phi_{NT} \in \mathbb{R}^{\dim \phi_{NT}}} \mathcal{L}_{NT}(\beta, \phi_{NT}) \,,$$

The fixed effects estimators of β^0 and ϕ^0_{NT} are

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^{\dim \beta}}{\operatorname{argmax}} \ \mathcal{L}_{NT}(\beta, \hat{\phi}_{NT}(\beta)) \ , \quad \hat{\phi}_{NT} = \hat{\phi}_{NT}(\hat{\beta}) \ .$$

Estimators of averages over the data and nuisance parameters can be formed by plugging-in the estimators of the model parameters in the sample version of (2.2), i.e.

$$\hat{\delta}_{NT} = \Delta_{NT}(\hat{\beta}, \hat{\phi}_{NT}). \tag{2.9}$$

3 Asymptotic Expansions

In this section, we derive asymptotic expansions for the score of the profile objective function $\mathcal{L}_{NT}(\beta, \hat{\phi}_{NT}(\beta))$ and for the fixed effects estimators of parameters and average effects. We do not employ the panel structure of the model, nor the particular form of the objective function given in Section 2. Instead, we consider the estimation of an unspecified model based on a sample of size NT and a generic objective function $\mathcal{L}_{NT}(\beta, \phi_{NT})$, which depends on the parameter of interest β and the incidental parameter ϕ_{NT}. In the sequel, we 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 ϕ for ϕ_{NT}. To derive the results, we make use of a set of high-level assumptions. These assumptions might appear somewhat abstract, but we will justify them by more primitive conditions in the context of panel models in the next section. There, we will also apply the expansions to obtain the limiting distribution of estimators of parameters and average effects. This section includes the key technical tools used in the paper, but readers only interested in the application to panel data models can skip directly to Section 4.

It is convenient to introduce some notation that will be extensively used in the analysis. Let

$$\mathcal{S}(\beta, \phi) = \partial_{\phi} \mathcal{L}(\beta, \phi), \quad \mathcal{H}(\beta, \phi) = -\partial_{\phi\phi'} \mathcal{L}(\beta, \phi), \tag{3.1}$$

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 (β^0, ϕ^0), e.g. $\mathcal{H} = \mathcal{H}(\beta^0, \phi^0)$. We use a bar to indicate expectations over Y_{it} and X_{it}, e.g. $\partial_\beta \overline{\mathcal{L}} = \mathbb{E}[\partial_\beta \mathcal{L}]$, and a tilde to denote that the variables are in deviation with respect to their expectations, e.g. $\partial_\beta \widetilde{\mathcal{L}} = \partial_\beta \mathcal{L} - \partial_\beta \overline{\mathcal{L}}$.

For $c \geq 0$, we also define the sets $\mathcal{B}(c, \beta^0) = \left\{\beta : \|\beta - \beta^0\|_\infty \leq c\right\}$, and $\mathcal{B}(c, \beta^0, \phi^0) = \left\{(\beta, \phi) : \|\beta - \beta^0\|_\infty < c,\ \|\phi - \phi^0\|_\infty < c\right\}$, which are closed balls of radius c around the true parameters β^0 and (β^0, ϕ^0), respectively, under the infinity norm.

We impose the following high-level conditions.

Assumption 3.1 (**Regularity Conditions for Asymptotic Expansion**).

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

(ii) *For all deterministic series* $\eta \searrow 0$, *we assume that* $\mathcal{L}(\beta, \phi)$ *is four times continuously differentiable with respect to* (β, ϕ) *in* $\mathcal{B}(\eta, \beta^0, \phi^0)$, *and for all integers* $p \geq 0$ *and* $q > 0$ *with* $2 \leq p + q \leq 4$ *and* $k_1, \ldots, k_p \in \{1, \ldots, \dim \beta\}$[2]

$$\sup_{(\beta,\phi)\in\mathcal{B}(\eta,\beta^0,\phi^0)} \max_{g_1\in\{1,\ldots,\dim\phi\}} \sum_{g_2,\ldots,g_q=1}^{\dim\phi} \left|\partial_{\beta_{k_1}\ldots\beta_{k_p},\phi_{g_1}\ldots\phi_{g_q}} \mathcal{L}(\beta,\, \phi)\right| = \mathcal{O}_P\left(\frac{\dim\phi}{NT}\right),$$

and also $\sup_{\beta\in\mathcal{B}(\eta,\beta^0)} \partial_{\beta_{k_1}\beta_{k_2}\beta_{k_3}} \mathcal{L}(\beta,\, \phi) = \mathcal{O}_P\,(1)$.

(iii) *There exists* ε, *with* $0 < \varepsilon < 1/6$, *such that*

$$\|\mathcal{S}\|_\infty = \mathcal{O}_P\left(\frac{(\dim\phi)^{1/2+\varepsilon}}{NT}\right), \qquad \left\|\partial_{\beta\beta'}\widetilde{\mathcal{L}}\right\|_{\max} = o_P(1),$$

$$\left\|\widetilde{\mathcal{H}}\right\|_\infty = \mathcal{O}_P\left(\frac{(\dim\phi)^{1/2+\varepsilon}}{NT}\right), \qquad \left\|\partial_{\beta\phi'}\widetilde{\mathcal{L}}\right\|_{\max} = \mathcal{O}_P\left(\frac{(\dim\phi)^{1/2+\varepsilon}}{NT}\right),$$

and

$$\sum_{g=1}^{\dim\phi} \left\|\partial_{\phi\phi'\phi_g}\widetilde{\mathcal{L}}\right\|_\infty = \mathcal{O}_P\left(\frac{(\dim\phi)^{1/2+\varepsilon}}{NT}\right), \quad \max_{g\in\{1,\ldots,\dim\phi\}} \left\|\partial_{\beta\phi'\phi_g}\widetilde{\mathcal{L}}\right\|_\infty = \mathcal{O}_P\left(\frac{(\dim\phi)^{1/2+\varepsilon}}{NT}\right).$$

Assumption 3.2 (**Order of** $\overline{\mathcal{H}}^{-1}$**).**

$$\left\|\overline{\mathcal{H}}^{-1}\right\|_\infty = \mathcal{O}\left(\frac{NT}{\dim\phi}\right).$$

Assumption 3.3 (**Uniform Consistency of** $\hat{\phi}(\beta)$**).** *For all deterministic series* $\eta \searrow 0$,

$$\sup_{\beta\in\mathcal{B}(\eta,\beta^0)} \left\|\hat{\phi}(\beta) - \phi^0\right\|_\infty = o_P(1).$$

We state Assumptions 3.2 and 3.3 separately because the primitive conditions to verify them in panels models depend on the specification of the individual and time effects. The following theorem is the main result of this section. All the proofs are given in the Appendix.

[2]For $q = 1$ we have no sum, but take the maximum over g_1.

Theorem 3.1 (**Asymptotic Expansions of $\hat\phi(\beta)$ and $d_\beta\mathcal{L}(\beta,\hat\phi(\beta))$**). *Let Assumptions 3.1, 3.2 and 3.3 hold. Then*

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

and

$$\sqrt{NT}\; d_\beta\mathcal{L}(\beta,\hat\phi(\beta))=T-\overline{W}\,\sqrt{NT}(\beta-\beta^0)+R(\beta),$$

where $T=T^{(0)}+T^{(1)}$ and

$$\overline{W}=-\left(\partial_{\beta\beta'}\overline{\mathcal{L}}+[\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}\,[\partial_{\phi\beta'}\overline{\mathcal{L}}]\right),$$
$$T^{(0)}=\sqrt{NT}\left(\partial_\beta\mathcal{L}+[\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}\mathcal{S}\right),$$
$$T^{(1)}=\sqrt{NT}\left([\partial_{\beta\phi'}\widetilde{\mathcal{L}}]\overline{\mathcal{H}}^{-1}\mathcal{S}-[\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}\,\widetilde{\mathcal{H}}\,\overline{\mathcal{H}}^{-1}\,\mathcal{S}\right)$$
$$+\sqrt{NT}\sum_{g=1}^{\dim\phi}\left(\partial_{\beta\phi'\phi_g}\overline{\mathcal{L}}+[\partial_{\beta\phi'}\overline{\mathcal{L}}]\,\overline{\mathcal{H}}^{-1}[\partial_{\phi\phi'\phi_g}\overline{\mathcal{L}}]\right)[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g\overline{\mathcal{H}}^{-1}\mathcal{S}/2.$$

The remainder terms of the expansions satisfy, for all series $\eta\searrow 0$

$$\sup_{\beta\in\mathcal{B}(\eta,\beta^0)}\frac{\sqrt{NT}\;\|r(\beta)\|_\infty}{1+\sqrt{NT}\|\beta-\beta^0\|}=o_P(1)\ ,\qquad \sup_{\beta\in\mathcal{B}(\eta,\beta^0)}\frac{\|R(\beta)\|}{1+\sqrt{NT}\|\beta-\beta^0\|}=o_P(1)\ .$$

Theorem 3.1 provides asymptotic expansions for the incidental parameter estimator $\hat\phi(\beta)$ and the score of the profile objective function $\partial_\beta\mathcal{P}(\beta)$. These are joint expansions in $\beta-\beta^0$ up to linear order, and in the incidental parameter score $\mathcal{S}$ up to quadratic order.[3] The theorem provides bounds on the the remainder terms $r(\beta)$ and $R(\beta)$, which make the expansions applicable to consistent estimators of β^0 that take values within a shrinking neighborhood of β^0 with probability approaching one. Depending on the model, consistency of $\hat\beta$ (and $\hat\phi$) can be established in different ways, as discussed for concrete panel models in the next section. Once we have established consistency of $\hat\beta$, the asymptotic expansion of the profile objective score can be applied to the first order condition for $\hat\beta$, $d_\beta\mathcal{L}(\hat\beta,\hat\phi(\hat\beta))=0$. This gives rise to the following corollary of Theorem 3.1 . Let $\overline{W}_\infty:=\lim_{N,T\to\infty}\overline{W}$.

[3]The terms $T^{(0)}$ and $T^{(1)}$ do not exactly correspond to the zero and first order components of the expansion. We separate the expansion in these terms, however, because it facilitates interpretation. In particular, $T^{(0)}$ is a variance term, while $T^{(1)}$ is a bias term.

Corollary 3.2 (**Stochastic expansion of $\hat{\beta}$**). *Let the assumptions of Theorem 3.1 be satisfied. If $\overline{W}_\infty$ exists, $\overline{W}_\infty > 0$, and $\|\hat{\beta} - \beta^0\| = o_P(1)$,*

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

Using this corollary we can derive the first order asymptotic theory of $\hat{\beta}$ from the limit average Hessian $\overline{W}_\infty$ and the limiting distribution of the approximated score T. In particular, assuming orders in probability correspond to orders in expectation, the first order asymptotic bias of $\hat{\beta}$ is

$$\lim_{N,T\to\infty} \sqrt{NT}\ \mathbb{E}[\hat{\beta} - \beta^0] = \overline{W}_\infty^{-1} \lim_{N,T\to\infty} \mathbb{E}\, T^{(1)}.$$

We illustrate the calculation of this asymptotic bias with two simple panel models. The first example is convenient analytically because the fixed effects estimator and its exact bias have closed form. We can therefore compare the exact bias with its first order asymptotic approximation. We consider more general models in the next section.

Example 3 (**Neyman-Scott model with additive effects**). *Consider the model:*

$$Y_{it} = \alpha_i^0 + \gamma_t^0 + \epsilon_{it}, \quad \epsilon_{it} \sim i.i.d.\ \mathcal{N}(0, \beta^0), \quad (i = 1, \ldots, N; t = 1, \ldots, T),$$

where the parameter of interest is $\beta^0 > 0$ and the effects α_i^0's and γ_t^0's are treated as nuisance parameters. This is a version of the classical Neyman-Scott (1948) example with time effects in addition to the individual effects. We normalize the true value of these effects to satisfy $\sum_{i=1}^N \alpha_i^0 = \sum_{t=1}^T \gamma_t^0$ for all N, T. The fixed effects estimator of β^0 is

$$\hat{\beta} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot t} - \bar{Y}_{\cdot\cdot})^2,$$

where $\bar{Y}_{i\cdot} = \sum_{t=1}^T Y_{it}/T$, $\bar{Y}_{\cdot t} = \sum_{i=1}^N Y_{it}/N$, and $\bar{Y}_{\cdot\cdot} = \sum_{i=1}^N \sum_{t=1}^T Y_{it}/(NT)$. The exact finite-sample bias of $\hat{\beta}$ is

$$\sqrt{NT}\ \mathbb{E}[\hat{\beta} - \beta^0] = -\beta^0 \frac{N + T - 1}{\sqrt{NT}}. \tag{3.2}$$

We turn now to the derivation of the first order asymptotic bias. The objective function is, for some constants c and $b > 0$:

$$\mathcal{L}(\beta, \phi) = c - (1/2) \log \beta - \sum_{i,t} (Y_{it} - \alpha_i - \gamma_t)^2 / (2\beta NT) - b \left(\sum_{i=1}^N \alpha_i - \sum_{t=1}^T \gamma_t \right)^2 / (2\beta^0 NT),$$

with corresponding incidental parameters score and Hessian evaluated at the true parameter values

$$\mathcal{S} = \frac{1}{\beta^0} \begin{pmatrix} \bar{\epsilon}_N / N \\ \bar{\epsilon}_T / T \end{pmatrix}, \mathcal{H} = \frac{1}{\beta^0 NT} \begin{pmatrix} T\mathbb{1}_N + b 1_N 1_N' & (1-b) 1_N 1_T' \\ (1-b) 1_T 1_N' & N\mathbb{1}_T + b 1_T 1_T' \end{pmatrix},$$

where $\bar{\epsilon}_N = [\bar{\epsilon}_{i\cdot}]_{i=1,\dots,N}$ *for* $\bar{\epsilon}_{i\cdot} = \sum_{t=1}^T \epsilon_{it}/T$, *and* $\bar{\epsilon}_T = [\bar{\epsilon}_{\cdot t}]_{t=1,\dots,T}$ *for* $\bar{\epsilon}_{\cdot t} = \sum_{i=1}^N \epsilon_{it}/N$. *The matrix* $\mathcal{H}$ *is deterministic so that* $\overline{\mathcal{H}} = \mathcal{H}$, *with inverse*

$$\overline{\mathcal{H}}^{-1} = \beta^0 \begin{pmatrix} N[\mathbb{1}_N - 1_N 1_N'/(T+N)] & 0 \\ 0 & T[\mathbb{1}_T - 1_T 1_T'/(T+N)] \end{pmatrix} + k\beta^0 \mathcal{R}, \quad (3.3)$$

where $k = NT(1-b)/[b(T+N)^2]$, *and* $\mathcal{R} = (1_N', -1_T')'(1_N', -1_T')$. *Note that* $\|\overline{\mathcal{H}}^{-1}\|_\infty = \mathcal{O}(N \vee T)$, *which satisfies Assumption 3.2 if* $\dim \phi / \sqrt{NT} = (N+T)/\sqrt{NT} \to a$, $0 < a < \infty$.

Substituting (3.3) in the expressions for T *and* $\overline{W}$ *in Theorem 3.1,*

$$\mathbb{E}[T] = \frac{1}{2\beta^0} \frac{N+T-1}{\sqrt{NT}}, \quad \overline{W} = -\frac{1}{2(\beta^0)^2},$$

where we use that $\partial_\beta \mathcal{L} = -1/(2\beta^0) + \sum_{i,t} \epsilon_{it}^2/[2(\beta^0)^2 NT]$, $\partial_{\beta\beta}\overline{\mathcal{L}} = -1/2(\beta^0)^2$, $\partial_{\beta\phi'}\mathcal{L}_{NT} = -\mathcal{S}'/\beta^0$, $\partial_{\beta\phi'}\overline{\mathcal{L}}_{NT} = 0$, $\partial_{\beta\phi'}\widetilde{\mathcal{L}}_{NT} = -\mathcal{S}'/\beta^0$, $\partial_{\beta\phi\phi'}\overline{\mathcal{L}} = [\mathcal{H} - \{b/(\beta^0 NT)\}\mathcal{R}]/\beta^0$, *and* $\mathcal{R}\mathcal{H}^{-1}\mathcal{S} = 0$. *Since* $\overline{W}_\infty = \overline{W}$, *the first order asymptotic bias of* $\hat{\beta}$ *is:*

$$\overline{W}_\infty^{-1} \lim_{N,T\to\infty} \mathbb{E}[T] = -\beta^0 \lim_{N,T\to\infty} \frac{N+T-1}{\sqrt{NT}} = -\beta^0 a,$$

which gives a first order approximation to the exact bias in equation (3.2).

For the second example, we consider again the Neyman-Scott model but making the individual and time effects interactive instead of additive. The fixed effects estimator in this case does not have closed form, what makes difficult to derive the exact bias. We find that the first order asymptotic bias is the same as in the additive case. The characterization of the first order bias for this model appears to be new.

Example 4 (Neyman-Scott model with interactive effects). *Consider the model:*

$$Y_{it} = \alpha_i^0 \gamma_t^0 + \epsilon_{it}, \quad \epsilon_{it} \sim i.i.d.\ \mathcal{N}(0, \beta^0), \quad (i = 1, \dots, N; t = 1, \dots, T),$$

where the parameter of interest is $\beta^0 > 0$ *and the effects* α_i^0*'s and* γ_t^0*'s are treated as nuisance parameters. We normalize the true value of the individual and time effects to satisfy* $\|\alpha^0\|^2 = \sum_{i=1}^N (\alpha_i^0)^2 = \sum_{t=1}^T (\gamma_t^0)^2$ *for all* N, T. *The fixed effects estimator of* β^0 *does not have closed form in this case.*

To derive the first order bias, we follow the same steps as in the previous example. The objective function is, for some constants c and $b > 0$:

$$\mathcal{L}(\beta, \phi) = c - (1/2)\log\beta - \sum_{i,t}(Y_{it} - \alpha_i\gamma_t)^2/(2\beta NT) - b(\sum_{i=1}^{N}\alpha_i^2 - \sum_{t=1}^{T}\gamma_t^2)^2/(8\beta^0 NT),$$

with corresponding incidental parameters score and Hessian evaluated at the true parameter values

$$\mathcal{S} = \frac{\|\alpha^0\|}{\beta^0}\begin{pmatrix} \tilde{\epsilon}_N/N \\ \tilde{\epsilon}_T/T \end{pmatrix}, \mathcal{H} = \frac{\|\alpha^0\|^2}{\beta^0 NT}\begin{pmatrix} \mathbb{1}_N + b\ \tilde{\alpha}_0\tilde{\alpha}_0' & (1-b)\tilde{\alpha}_0\tilde{\gamma}_0' \\ (1-b)\tilde{\gamma}_0\tilde{\alpha}_0' & \mathbb{1}_T + b\ \tilde{\gamma}_0\tilde{\gamma}_0' \end{pmatrix}, \qquad (3.4)$$

where $\tilde{\epsilon}_N = [\tilde{\epsilon}_{i\cdot}]_{i=1,\ldots,N}$ *for* $\tilde{\epsilon}_{i\cdot} = \sum_{t=1}^{T}\tilde{\gamma}_t^0\epsilon_{it}/T$, $\tilde{\epsilon}_T = [\tilde{\epsilon}_{\cdot t}]_{t=1,\ldots,T}$ *for* $\tilde{\epsilon}_{\cdot t} = \sum_{i=1}^{N}\tilde{\alpha}_i^0\epsilon_{it}/N$, $\tilde{\gamma}_t^0 = \gamma_t^0/\|\alpha^0\|$, $\tilde{\alpha}_i^0 = \alpha_i^0/\|\alpha^0\|$, $\tilde{\gamma}_0 = (\tilde{\gamma}_1^0, \ldots, \tilde{\gamma}_T^0)$, *and* $\tilde{\alpha}_0 = (\tilde{\alpha}_1^0, \ldots, \tilde{\alpha}_N^0)$. *The matrix* $\mathcal{H}$ *is fixed conditional on the nuisance parameters, so that* $\overline{\mathcal{H}} = \mathcal{H}$, *with inverse*

$$\overline{\mathcal{H}}^{-1} = \beta^0\frac{NT}{\|\alpha^0\|^2}\begin{pmatrix} \mathbb{1}_N - \tilde{\alpha}_0\tilde{\alpha}_0'/2 & 0 \\ 0 & \mathbb{1}_T - \tilde{\gamma}_0\tilde{\gamma}_0'/2 \end{pmatrix} + k\beta^0\frac{NT}{\|\alpha^0\|^2}\tilde{\mathcal{R}},$$

where $k = (1-b)/(4b)$, *and* $\tilde{\mathcal{R}} = (\tilde{\alpha}_0', -\tilde{\gamma}_0')'(\tilde{\alpha}_0', -\tilde{\gamma}_0')$. *Note that* $\|\overline{\mathcal{H}}^{-1}\|_\infty = \mathcal{O}(NT/\|\alpha^0\|^2)$, *which satisfies Assumption 3.2 if* $\|\alpha^0\|^2 = \mathcal{O}(\dim\phi) = \mathcal{O}(N+T)$.

Substituting (3.4) in the expressions for T *and* $\overline{W}$ *in Theorem 3.1,*

$$\mathbb{E}[T] = \frac{1}{2\beta^0}\frac{N+T-1}{\sqrt{NT}}, \quad \overline{W} = -\frac{1}{2(\beta^0)^2},$$

where we use that $\partial_\beta\mathcal{L} = -1/(2\beta^0) + \sum_{i,t}\epsilon_{it}^2/[2(\beta^0)^2 NT]$, $\partial_{\beta\beta}\overline{\mathcal{L}} = -1/2(\beta^0)^2$, $\partial_{\beta\phi'}\mathcal{L}_{NT} = -\mathcal{S}'/\beta^0$, $\partial_{\beta\phi'}\overline{\mathcal{L}}_{NT} = 0$, $\partial_{\beta\phi'}\widetilde{\mathcal{L}}_{NT} = -\mathcal{S}'/\beta^0$, $\partial_{\beta\phi\phi'}\overline{\mathcal{L}} = [\mathcal{H} - \{b\|\alpha^0\|^2/(\beta^0 NT)\}\tilde{\mathcal{R}}]/\beta^0$, *and* $\tilde{\mathcal{R}}\mathcal{H}^{-1}\mathcal{S} = 0$. *Since* $\overline{W}_\infty = \overline{W}$, *the first order asymptotic bias of* $\hat{\beta}$ *is:*

$$\overline{W}_\infty^{-1}\lim_{N,T\to\infty}\mathbb{E}[T] = -\beta^0\lim_{N,T\to\infty}\frac{N+T-1}{\sqrt{NT}} = -\beta^0 a,$$

which is the same as the first order bias for the additive effects model in the previous example.

Expansion for Average Effects

In nonlinear models, the quantities of interest are often averages of the data and parameters, including average partial effects. We derive asymptotic expansions for the fixed effects estimators of these quantities defined in (2.9). We invoke the following high-level assumption, which is verified under more primitive conditions for panel data models in the next section.

Assumption 3.4 (**Regularity Conditions for Asymptotic Expansion of** $\hat{\delta}$)**.**

(i) For all deterministic series $\eta \searrow 0$*, we assume that* $\Delta(\beta, \phi)$ *is three times continuously differentiable with respect to* (β, ϕ) *in* $\mathcal{B}(\eta, \beta^0, \phi^0)$*, and for all integers* $p \geq 0$*,* $q > 0$ *with* $2 \leq p+q \leq 3$ *and* $k_1, \ldots, k_p \in \{1, \ldots, \dim \beta\}$

$$\sup_{(\beta,\phi)\in\mathcal{B}(\eta,\beta^0,\phi^0)} \max_{g_1\in\{1,\ldots,\dim\phi\}} \sum_{g_2,\ldots,g_q=1}^{\dim\phi} \left|\partial_{\beta_{k_1}\ldots\beta_{k_p}\phi_{g_1}\ldots\phi_{g_q}} \Delta(\beta, \phi)\right| = \mathcal{O}_P\left(\frac{\dim\phi}{NT}\right),$$

and also $\sup_{\beta\in\mathcal{B}(\eta,\beta^0)} \left\|\partial_{\beta\beta'}\Delta(\beta, \phi)\right\|_{\max} = \mathcal{O}_P(1)$.

(ii) There exists ε*, with* $0 < \varepsilon < 1/6$*, such that*

$$\left\|\partial_\beta \widetilde{\Delta}\right\|_{\max} = o_P(1), \quad \left\|\partial_{\phi\phi'}\widetilde{\Delta}\right\|_\infty = \mathcal{O}_P\left(\frac{(\dim\phi)^{1/2+\varepsilon}}{NT}\right).$$

The following result gives the asymptotic expansion for the estimator of the average effects.

Theorem 3.3 (**Asymptotic Expansion of** $\Delta(\beta, \hat{\phi}(\beta))$)**.** *Let Assumptions 3.1, 3.2, 3.3 and 3.4 hold. Then*

$$\Delta(\beta, \hat{\phi}(\beta)) - \Delta = [\partial_\beta \overline{\Delta}]'(\beta - \beta^0) + [\partial_\phi \Delta]'(\hat{\phi}(\beta) - \phi^0) + \tfrac{1}{2}\,\mathcal{S}'\overline{\mathcal{H}}^{-1}[\partial_{\phi\phi'}\overline{\Delta}]\overline{\mathcal{H}}^{-1}\mathcal{S} + r^\Delta(\beta).$$

The remainder term of the expansions satisfies for all series $\eta \searrow 0$

$$\sup_{\beta\in\mathcal{B}(\eta,\beta^0)} \frac{\sqrt{NT}\;\left\|r^\Delta(\beta)\right\|}{1 + \sqrt{NT}\|\beta - \beta^0\|} = o_P(1).$$

As in Theorem 3.1, the remainder term of the asymptotic expansion is bounded in an shrinking neighborhood of β^0. We can therefore apply the expansion directly to characterize the asymptotic distribution of average effects estimators constructed from consistent estimators of β^0. The next corollary formalizes this observation.

Corollary 3.4 (**Stochastic expansion of** $\hat{\delta}$)**.** *Let the assumptions of Theorem 3.1 be satisfied. If* $\overline{W}_\infty$ *exists,* $\overline{W}_\infty > 0$*, and* $\|\hat{\beta} - \beta^0\| = o_P(1)$*,*

$$\begin{aligned}\sqrt{NT}(\hat{\delta} - \delta^0) &= \sqrt{NT}\widetilde{\Delta} + [\partial_\beta\overline{\Delta}]'\overline{W}_\infty^{-1}T + [\partial_\phi\Delta]'\overline{\mathcal{H}}^{-1}(\mathcal{S} - \widetilde{\mathcal{H}}\overline{\mathcal{H}}^{-1}\mathcal{S}) \\ &\quad + [\partial_\phi\Delta]'\overline{\mathcal{H}}^{-1}\left(\sum_{g=1}^{\dim\phi}[\partial_{\phi\phi'\phi_g}\overline{\mathcal{L}}]\overline{\mathcal{H}}^{-1}\mathcal{S}[\overline{\mathcal{H}}^{-1}\mathcal{S}]_g/2 + [\partial_{\beta\phi'}\overline{\mathcal{L}}]'\overline{W}_\infty^{-1}T\right) \\ &\quad + \tfrac{1}{2}\,\mathcal{S}'\overline{\mathcal{H}}^{-1}[\partial_{\phi\phi'}\overline{\Delta}]\overline{\mathcal{H}}^{-1}\mathcal{S} + o_P(1).\end{aligned}$$

4 Application to Panel Models

We provide primitive conditions for the validity of the asymptotic expansions of the previous section in panel models, and characterize the limiting distributions of fixed effects estimators of parameters and average effects in these models. We focus on linear single-index models with additive or interactive individual and time effects. These models are commonly used in applied economics and include the probit and Poisson specifications of Examples 1 and 2. Moreover, under concavity of the objective function in the index, we can establish the consistency of the fixed effects estimator.

In panel linear index models the conditional distributions of Y_{it} depends on the explanatory variables and unobserved effects through a single index function, i.e.

$$\log f_{Y|X}(y \mid X_{it}, \alpha_i, \gamma_t, \beta) = \ell(Y_{it},\, Z_{it}(\beta, \alpha_i, \gamma_t)),$$

where $Z_{it}(\beta, \alpha_i, \gamma_t) = X_{it}'\beta + g(\alpha_i, \gamma_t) \in \mathbb{R}$ is the linear index, and g is a known function. For the derivatives of $z \mapsto \ell_{it}(z) := \ell(Y_{it}, z)$ we use the notation $\partial_{z^q}\ell_{it}(z) = \partial^q \ell_{it}(z)/\partial z^q$, and we drop the argument z when the derivatives are evaluated at the true index $Z_{it} := Z_{it}(\beta^0, \alpha_i^0, \gamma_t^0)$, i.e. $\partial_{z^q}\ell_{it} = \partial_{z^q}\ell_{it}(Z_{it})$. Let $\mathcal{Y}$, $\mathcal{X}$, and $\mathcal{B} \times \mathcal{A} \times \mathcal{G} \subset \mathbb{R}^{\dim \beta + 2}$ be the supports of Y_{it}, X_{it}, and $(\beta^0, \alpha_i^0, \gamma_t^0)$, respectively. Define the support of the index as $\mathcal{Z} = \{z \in \mathbb{R} : z = x'\beta + \alpha + \gamma, x \in \mathcal{X}, \beta \in \mathcal{B}, \alpha \in \mathcal{A}, \gamma \in \mathcal{G}\}$. In all the panel models that we consider $\dim \phi = N + T$.

We make the following assumptions.

Assumption 4.1. (Primitive Conditions for Panel Index Models)

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

(ii) *Sampling:* (Y, X) *is independent across* i *conditional on* ϕ^0*, and alpha mixing across* t*, with mixing coefficient that decrease at an exponential rate, that is* $\sup_i a_i(m) \leq C\epsilon^m$ *for some* ϵ *such that* $0 < \epsilon < 1$ *and some* $C > 0$*, where*

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

for $\mathcal{A}_t^i := \sigma(\alpha_i^0, \gamma_t^0, X_{it}, \gamma_{t-1}^0, X_{i,t-1}, \ldots)$ *and* $\mathcal{B}_t^i := \sigma(\alpha_i^0, \gamma_t^0, X_{it}, \gamma_{t+1}^0, X_{i,t+1}, \ldots)$.

(iii) *Mean zero score:* $\mathbb{E}_t[\partial_z \ell_{it}] = 0$*, for all* i, t, N, T*, where* $\ell_{it} = \ell(Y_{it}, X_{it}'\beta^0 + g(\alpha_i^0, \gamma_t^0))$*, and* ℓ *and* g *are known functions.*

(iv) *Objective function: For all* $y \in \mathcal{Y}$*, the function* $z \mapsto \ell(y, z)$ *is four times continuously differentiable over* $\mathbb{R}$*. For* $q = 0, 1, 2, 3, 4$*, there exists a function*

$M_{it}(z)$ such that $|\partial_{z^q}\ell_{it}(z)| \leq M_{it}(z)$, and we assume that $\max_{i,t}\mathbb{E}[M_{it}(z)^{16}]$ is uniformly bounded over $z \in \mathcal{Z}$ and N, T.

(v) *Index: the function $(\alpha_i, \gamma_t) \mapsto g(\alpha_i, \gamma_t)$ is four times continuously differentiable over $\mathbb{R}^2$ with bounded partial derivatives up to fourth order uniformly over $\mathcal{A} \times \mathcal{G}$. Uniformly over N, T, $\max_{i,t}\mathbb{E}[\|X_{it}\|^{16}] < \infty$.*

(vi) *Average effects: the function $(\beta, \alpha_i, \gamma_t) \mapsto \Delta_{it}(\beta, \alpha_i, \gamma_t) = \Delta(X_{it}, \beta, \alpha_i, \gamma_t)$ is four times continuously differentiable over $\mathcal{B} \times \mathcal{A} \times \mathcal{G}$. The partial derivatives of $\Delta_{it}(\beta, \alpha_i, \gamma_t)$ with respect to the elements of $(\beta, \alpha_i, \gamma_t)$ up to fourth order are bounded by a function $M^{\Delta}_{it}(\beta, \alpha_i, \gamma_t)$, and we assume that $\max_{i,t}\mathbb{E}[M^{\Delta}_{it}(\beta, \alpha_i, \gamma_t)^{16}]$ is uniformly bounded over $\mathcal{B} \times \mathcal{A} \times \mathcal{G}$ and N, T.*

[Note: Continuity of the highest order derivative in (v)–(vii) can be replaced by a Lipschitz condition]

Remark 1. *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). This type of assumption is violated in the models we consider due to the presence of the time effects.*

Remark 2. *Assumption 4.1(iii) covers quasi-likelihood estimators where we only need specify some characteristic of the conditional distribution such as the mean.*

We verify that the previous conditions guarantee that Assumptions 3.1 and 3.4 hold.

Lemma 4.1. *Assumption 4.1 guarantees that Assumption 3.1 holds for the panel objective functions* (2.6), (2.7) *and* (2.8). *It also guarantees that Assumption 3.4 is satisfied by the average effects defined in* (2.2).

To verify Assumptions 3.2 and 3.3, we impose additional assumptions on the objective function ℓ and the specification of the effects g. We treat separately the additive and interactive effects specifications. In both cases we impose concavity assumptions on the objective function with respect to the index. Probit, logit, tobit and Poisson models have strictly concave objective functions in the index.

4.1 Linear single index with additive effects

We first consider models with additive individuals and time effects. Let X_k denote the $N \times T$ matrix with elements $\{X_{it,k} : i = 1, ..., N, t = 1, ..., T\}$ corresponding to all the observations of the k'th regressor of X_{it}.

Assumption 4.2 (Additive effects and concavity).

(i) Additive effects: $g(\alpha_i, \gamma_t) = \alpha_i + \gamma_t$.

(ii) Non-collinearity: The matrix with elements (k_1, k_2)

$$\lim_{N,T\to\infty} (NT)^{-1}\mathbb{E}\mathrm{Tr}(M_{1_N} X_{k_1} M_{1_T} X'_{k_2}), \quad k_1, k_2 \in \{1, ..., \dim X_{it}\},$$

exists and is positive definite.

(iii) Concavity: For all $y \in \mathcal{Y}$ *the function* $z \mapsto \ell(y, z)$ *is strictly concave over* $\mathbb{R}$. *Furthermore, there exist constants* $b_{\min}$ *and* $b_{\max}$ *such that* $0 < b_{\min} \leq -\mathbb{E}\left[\partial_{z^2}\ell_{it}\right] \leq b_{\max}$ *uniformly over* i, t, N, T.

Remark 3. *With additive effects the first part of Assumption 4.1(v) holds trivially because* $\partial_{\alpha_i} g(\alpha_i, \gamma_t) = \partial_{\gamma_t} g(\alpha_i, \gamma_t) = 1$, *and the higher order derivatives are zero.*

We now verify Assumptions 3.2 and 3.3 and derive the asymptotic expansion for the fixed effects estimators in this model. The objective function is given by (2.7), which can also be written as

$$\mathcal{L}(\beta, \phi) = (NT)^{-1} \left\{ \sum_{i,t} \ell(Y_{it},\, X'_{it}\beta + \alpha_i + \gamma_t) - \frac{b}{2} \left(\sum_i \alpha_i - \sum_t \gamma_t \right)^2 \right\}.$$

Recall that the second term is a penalty that deals with the degenerate direction of the parameter space corresponding to adding a constant to all α_i and subtracting the same constant from all γ_t. The incidental parameter score evaluated at the true parameters is

$$\mathcal{S} = (NT)^{-1} \begin{pmatrix} \left[\sum_{t=1}^{T} \partial_z \ell_{it}\right]_{i=1,\ldots,N} \\ \left[\sum_{i=1}^{N} \partial_z \ell_{it}\right]_{t=1,\ldots,T} \end{pmatrix},$$

and the expected incidental parameter Hessian is

$$\overline{\mathcal{H}} = (NT)^{-1} \left\{ \begin{pmatrix} \overline{\mathcal{D}}_\alpha & \{-\mathbb{E}[\partial_{z^2}\ell_{it}]\}_{i,t} \\ \{-\mathbb{E}[\partial_{z^2}\ell_{it}]\}_{t,i} & \overline{\mathcal{D}}_\gamma \end{pmatrix} + b\, vv' \right\},$$

where $\overline{\mathcal{D}}_\alpha$ and $\overline{\mathcal{D}}_\gamma$ are diagonal $N \times N$ and $T \times T$ matrices, namely

$$\overline{\mathcal{D}}_\alpha = \mathrm{diag}\left(\left\{ -\sum_{t=1}^{T} \mathbb{E}[\partial_{z^2}\ell_{it}] \right\}_{i=1,\ldots,N} \right), \quad \overline{\mathcal{D}}_\gamma = \mathrm{diag}\left(\left\{ -\sum_{i=1}^{N} \mathbb{E}[\partial_{z^2}\ell_{it}] \right\}_{t=1,\ldots,T} \right),$$

and $v = (1'_N, 1'_T)'$.

In panel models with only individual effects, it is straightforward to determine the order of magnitude of $\overline{\mathcal{H}}^{-1}$ in Assumption 3.2, because $\overline{\mathcal{H}}$ contains only the diagonal matrix $\overline{\mathcal{D}}_\alpha$. Here, $\overline{\mathcal{H}}$ is no longer diagonal, but it has a special structure. The diagonal terms are of orders T^{-1} and N^{-1}, whereas the off-diagonal terms are of order $(NT)^{-1}$. Moreover, $\left\|\overline{\mathcal{H}} - \overline{\mathcal{D}}\right\|_{\max} = \mathcal{O}((NT)^{-1})$ by Assumption 4.2, where $\overline{\mathcal{D}} = (NT)^{-1}\text{diag}(\overline{\mathcal{D}}_\alpha, \overline{\mathcal{D}}_\gamma)$. 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 $\mathcal{O}(N+T)$. The following lemma shows that indeed the diagonal terms of $\overline{\mathcal{H}}$ dominate in the determination the order of the inverse.

Lemma 4.2. *Under Assumptions 4.1 and 4.2,*

$$\left\|\overline{\mathcal{H}}^{-1} - \overline{\mathcal{D}}^{-1}\right\|_{\max} = \mathcal{O}(1)$$

This result establishes that $\overline{\mathcal{H}}^{-1}$ can be uniformly approximated by the diagonal matrix $\overline{\mathcal{D}}^{-1}$, which is given by the inverse of the diagonal terms of $\overline{\mathcal{H}}$ without the penalty. The non-zero elements of $\overline{\mathcal{D}}^{-1}$ are of order N and T, respectively, i.e. the order of the difference established by the lemma is relatively small. This technical intermediate result is key to determine the order of $\overline{\mathcal{H}}^{-1}$ in Assumption 3.2, and to simplify the terms in the asymptotic expansion of $\hat{\beta}$.

Remark 4. *The choice of penalty in the objective function* (4.1) *is important to obtain the result. Different penalties, corresponding to other normalizations (e.g. a penalty proportional to* α_1^2*, corresponding to the normalization* $\alpha_1^0 = 0$*), would fail to deliver Lemma 4.2, although these choices would not affect the estimators* $\hat{\beta}$ *or* $\hat{\Delta}$*.*

The following lemma verifies Assumptions 3.2 and 3.3, proves the consistency of $\hat{\beta}$, and shows that $\overline{W}_\infty$ exists and is positive definite. The concavity assumption helps show the consistency of $\hat{\beta}$ and $\hat{\phi}$, and together with non-collinearity guarantees the results for $\overline{W}_\infty$.

Lemma 4.3. *Under Assumptions 4.1 and 4.2, Assumption 3.2 and 3.3 hold,* $\hat{\beta} = \beta^0 + o_P(1)$*, and* $\overline{W}_\infty$ *exists and is positive definite.*

We have now all the ingredients to apply the asymptotic expansions of Section 3 in the panel model. In the rest of this section, we use these expansions to characterize the asymptotic distribution of $\hat{\beta}$ and $\hat{\Delta}$.

We start the analysis introducing some additional notation. Let

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

where the blocks $\overline{\mathcal{H}}^{-1}_{\alpha\alpha}$, $\overline{\mathcal{H}}^{-1}_{\alpha\gamma}$, $\overline{\mathcal{H}}^{-1}_{\gamma\alpha}$, and $\overline{\mathcal{H}}^{-1}_{\gamma\gamma}$ are $N \times N$, $N \times T$, $T \times N$ and $T \times T$ matrices, respectively. Let

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

This matrix can be interpreted as the population projection of X_{it} on the space spanned by the incidental parameters under a metric given by $\partial_{z^2}\ell_{it}$, namely, for $k = 1, \ldots, \dim X_{it}$,

$$\mathcal{X}_{it,k} = \alpha^*_{i,k} + \gamma^*_{t,k}, \;\; (\alpha^*_k, \, \gamma^*_k) = \underset{\alpha_{i,k}, \gamma_{t,k}}{\operatorname{argmin}} \sum_{i,t} \mathbb{E} \left[-\partial_{z^2} \ell_{it} \left(X_{it,k} - \alpha_{i,k} - \gamma_{t,k} \right)^2 \right]. \quad (4.1)$$

Finally, let

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

Substituting these definitions in the expressions of Theorem 3.1 yields

$$\overline{W} = -\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E} \left[\partial_{z^2} \ell_{it} \left(X_{it} - \mathcal{X}_{it} \right) \left(X_{it} - \mathcal{X}_{it} \right)' \right],$$

$$T^{(0)} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \partial_z \ell_{it} \left(X_{it} - \mathcal{X}_{it} \right),$$

$$T^{(1)} = -\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \Lambda_{it} \, \partial_{z^2} \ell_{it} \left(X_{it} - \mathcal{X}_{it} \right) + \frac{1}{2} \Lambda^2_{it} \, \mathbb{E} \left[\partial_{z^3} \ell_{it} \left(X_{it} - \mathcal{X}_{it} \right) \right] \right\},$$

The following result uses these expressions to chacterize the asymptotic distribution of $\hat{\beta}$ and $\hat{\Delta}$.

Theorem 4.4 (**Asymptotic Distribution of $\hat{\beta}$ and $\hat{\Delta}$ with Additive Effects**). *Let Assumptions 2.1, 4.1, and 4.2 be satisfied, and furthermore assume that the following limits exist*

$$\overline{B}_{\infty} = - \lim_{N,T \to \infty} \frac{1}{N} \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} \sum_{\tau=t}^{T} \mathbb{E} \left[\left(\partial_z \ell_{it} \partial_{z^2} \ell_{i\tau} + 1(\tau = t) \partial_{z^3} \ell_{it}/2 \right) \left(X_{i\tau} - \mathcal{X}_{i\tau} \right) \right]}{\sum_{t=1}^{T} \mathbb{E} \left(\partial_{z^2} \ell_{it} \right)},$$

$$\overline{D}_{\infty} = - \lim_{N,T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \frac{\sum_{i=1}^{N} \mathbb{E} \left[\left(\partial_z \ell_{it} \partial_{z^2} \ell_{it} + \partial_{z^3} \ell_{it}/2 \right) \left(X_{it} - \mathcal{X}_{it} \right) \right]}{\sum_{i=1}^{N} \mathbb{E} \left(\partial_{z^2} \ell_{it} \right)}.$$

Then, $T_{NT}^{(0)} \to_d \mathcal{N}(0,\, \overline{W}_\infty)$, $T_{NT}^{(1)} \to_P \kappa \overline{B}_\infty + \kappa^{-1}\overline{D}_\infty$, *and therefore*

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

Also,

$$\sqrt{NT}(\hat{\delta} - \delta^0) \to_d \mathcal{N}(\kappa\overline{B}_\infty^\delta + \kappa^{-1}\overline{D}_\infty^\delta,\, \overline{V}_\infty^\delta)$$

where [TBA: expressions of $\overline{B}_\infty^\delta$, $\overline{D}_\infty^\delta$, *and* $\overline{V}_\infty^\delta$*]*

Remark 5. *The expressions of the bias terms use that*

$$\Lambda_{it} = \frac{\sum_{\tau=1}^T \partial_z \ell_{i\tau}}{\sum_{\tau=1}^T \mathbb{E}\left(\partial_{z^2}\ell_{i\tau}\right)} + \frac{\sum_{j=1}^N \partial_z \ell_{jt}}{\sum_{j=1}^N \mathbb{E}\left(\partial_{z^2}\ell_{jt}\right)} + \text{terms of smaller order,}$$

by Lemma 4.2. Moreover, under Assumption 2.1 all the score functions evaluated at the true parameter values follow martingale differences and we apply Bartlett identities to simplify further the expressions. Without Assumption 2.1 the expressions for the biases are

$$\overline{B}_\infty = -\lim_{N,T\to\infty} \frac{1}{N}\sum_{i=1}^N \frac{\sum_{t=1}^T \sum_{\tau=t}^T \mathbb{E}\left[\partial_z \ell_{it} \partial_{z^2}\ell_{i\tau}\left(X_{i\tau} - \mathcal{X}_{it}\right)\right]}{\sum_{t=1}^T \mathbb{E}\left(\partial_{z^2}\ell_{it}\right)}$$
$$+ \frac{1}{2}\lim_{N,T\to\infty} \frac{1}{N}\sum_{i=1}^N \frac{\sum_{t=1}^T \mathbb{E}[(\partial_z \ell_{it})^2] \sum_{t=1}^T \mathbb{E}[\partial_{z^3}\ell_{it}\left(X_{it} - \mathcal{X}_{it}\right)]}{\left[\sum_{t=1}^T \mathbb{E}\left(\partial_{z^2}\ell_{it}\right)\right]^2},$$
$$\overline{D}_\infty = -\lim_{N,T\to\infty} \frac{1}{T}\sum_{t=1}^T \frac{\sum_{i=1}^N \mathbb{E}\left[\partial_z \ell_{it} \partial_{z^2}\ell_{it}\left(X_{it} - \mathcal{X}_{it}\right)\right]}{\sum_{i=1}^N \mathbb{E}\left(\partial_{z^2}\ell_{it}\right)}$$
$$+ \frac{1}{2}\lim_{N,T\to\infty} \frac{1}{T}\sum_{t=1}^T \frac{\sum_{i=1}^N \mathbb{E}[(\partial_z \ell_{it})^2] \sum_{i=1}^N \mathbb{E}[\partial_{z^3}\ell_{it}\left(X_{it} - \mathcal{X}_{it}\right)]}{\left[\sum_{i=1}^N \mathbb{E}\left(\partial_{z^2}\ell_{it}\right)\right]^2}.$$

Remark 6. *The structure of* $\overline{B}_\infty$ *is similar to the expression of the bias when the model has only individual effects, but the expression of the projected regressor* $\mathcal{X}_{it}$ *is different. The structure of* $\overline{D}_\infty$ *is symmetric to that of* $\overline{B}_\infty$*, with the role of time and cross-sectional dimensions interchanged and using cross-sectional independence.*

[TBA: explanation of the components of the bias]

We apply the expressions of the bias of Theorem 4.4 to the Poisson model of Example 2. We find that the fixed effects estimator of the model parameters do not have first order asymptotic bias if all the regressors are strictly exogenous. This finding extends the Hausman, Hall and Griliches (1984) result of zero bias in models with individual effects to models with individual and time effects. Chabornneau (2011) has recently found an alternative conditional fixed effects estimator that conditions on sufficient statistics of the individual and time effects. Unlike in the model with only individual effects, this conditional estimator does not seem to be equivalent to the fixed effects estimator that concentrates out the effects.

Example 2 (cont.) In the Poisson model, $\ell(y,z) = zy - \exp z - \log y!$, so that $\partial_z \ell(y,z) = y - \exp z$, and $\partial_{z^2}\ell(y,z) = \partial_{z^3}\ell(y,z) = -\exp z$. Substituting in the expressions of the bias of Theorem 4.4,

$$\overline{B}_\infty = -\lim_{N,T\to\infty} \frac{1}{N}\sum_{i=1}^{N} \frac{\sum_{t=1}^{T}\sum_{\tau=t+1}^{T} \mathbb{E}\left[(Y_{it} - \exp Z_{it})\exp Z_{it}\,(X_{i\tau} - \mathcal{X}_{i\tau})\right]}{\sum_{t=1}^{T}\mathbb{E}\left(\exp Z_{it}\right)},$$

and $\overline{D}_\infty = 0$. Here, we use that $\mathbb{E}_t[\partial_z \ell_{it} \partial_z^2 \ell_{it}] = 0$, and $\mathbb{E}[\partial_z^3 \ell_{it}(X_{it} - \mathcal{X}_{it})] = 0$ by the first order conditions of program (4.1) since $\partial_{z^3}\ell_{it} = \partial_{z^2}\ell_{it}$. If in addition all the components of X_{it} are strictly exogenous, the score $(Y_{it} - \exp Z_{it})\exp Z_{it}$ is uncorrelated to past, present and future values of $X_{i\tau}$ and therefore $\overline{B}_\infty = 0$.

4.2 Linear single index model with interactive effects

[TBA]

5 Bias Corrections

The results of the previous section show that the asymptotic distribution of the fixed effects estimator of the model parameters and average effects have a bias of the same order as the asymptotic variance under sequences where T grows at the same rate as N. This is the large-T version of the incidental parameters problem that invalidates any inference based on the asymptotic distribution. In this section we consider analytical and jackknife bias corrections. We focus the discussion on a generic parameter θ with true value θ^0 and fixed effects estimator $\hat{\theta}$. For example, $\theta = \beta$ for the model parameters or $\theta = \delta$ for an average partial effect.

To understand how the bias corrections work, it is useful to start from the following expansion for the expected value of the fixed effect estimator:

$$\mathbb{E}[\hat{\theta}] = \theta_{NT} = \theta^0 + \overline{B}_\infty/T + \overline{D}_\infty/N + o(T^{-1} \vee N^{-1}). \tag{5.1}$$

Note also that by the properties of the maximum likelihood estimator

$$\sqrt{NT}(\hat{\theta} - \theta_{NT}) \to_d N(0, \overline{V}_\infty).$$

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

$$\tilde{\theta}_A = \hat{\theta} - \hat{B}/T - \hat{D}/N. \tag{5.2}$$

If $\sqrt{NT}(\hat{B} - \overline{B}_\infty)/T \to_P 0$, $\sqrt{NT}(\hat{D} - \overline{D}_\infty)/N \to_P 0$, and $N/T \to \kappa^2$, then

$$\begin{aligned}\sqrt{NT}(\tilde{\theta}_A - \theta^0) &= \sqrt{NT}(\hat{\theta} - \theta_{NT}) + \sqrt{NT}(\hat{B} - \overline{B}_\infty)/T \\ &\quad + \sqrt{NT}(\hat{D} - \overline{D}_\infty)/N + \sqrt{NT}\ o(T^{-1} \vee N^{-1}) \to_d N(0, \overline{V}_\infty).\end{aligned}$$

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

We also consider two different jackknife bias correction methods that do not do require explicit estimation of the bias, but are computationally more intensive. The first method is based on applying the leave one observation out panel jackknife of Hahn and Newey (2004) to the individual dimension, and the split panel jackknife of Dhaene and Jochmans (2010) to the time dimension. The second method is based on applying the split panel jackknife to both dimensions, and allows for cross sectional dependencies.

To describe the first jackknife correction, let $\bar{\theta}_{N-1,T}$ be the average of the N jackknife estimators that leave out one individual, $\bar{\theta}_{N,T/2}$ be the average of the 2 split jackknife estimators that leave out the first and second halves of the time periods, and $\bar{\theta}_{N-1,T/2}$ be the average of the $2N$ split jackknife estimators that leave out one individual and one half of the time periods. The bias corrected estimator is

$$\tilde{\theta}_{J1} = 2N\hat{\theta} - 2(N-1)\bar{\theta}_{N-1,T} - N\bar{\theta}_{N,T/2} + (N-1)\bar{\theta}_{N-1,T/2}.$$

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

$$\begin{aligned}\mathbb{E}[\tilde{\theta}_{J1}] &= [2N - 2(N-1) - N + (N-1)]\theta^0 \\ &\quad + [2N/T - 2(N-1)/T - 2N/T + 2(N-1)/T]\overline{B}_\infty + [2 - 2 - 1 + 1]\overline{D}_\infty \\ &\quad + [2N - 2(N-1) - N + (N-1)]o(T^{-1} \vee N^{-1}) = \theta^0 + o(T^{-1} \vee N^{-1}).\end{aligned}$$

where we use the expansion (5.1) under suitable assumptions on the remainder term.

To describe the second jackknife correction, let $\bar{\theta}_{N/2,T/2}$ be the average of the four split jackknife estimators that leave out half of the cross-sectional units and the first or second halve of the time periods. In choosing the cross sectional division of the panel, we might want to take into account clustering structures to preserve and account for cross sectional dependencies. The bias corrected estimator is

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

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

$$\mathbb{E}[\tilde{\theta}_{J2}] = (2-1)\theta^0 + [2/T - 1/(T/2)]\overline{B}_\infty + [2/N - 1/(N/2)]\overline{D}_\infty + o(T^{-1} \vee N^{-1})$$
$$= \theta^0 + o(T^{-1} \vee N^{-1}).$$

where we use the expansion (5.1) under suitable assumptions on the remainder term.

The following result shows that the analytical and jackknife bias corrections center the asymptotic distribution of the fixed effects estimator without increasing asymptotic variance.

Theorem 5.1. *Under the conditions of Theorem 4.4, for* $C \in \{A, J1, J2\}$

$$\sqrt{NT}(\tilde{\theta}_C - \theta^0) \to_d \mathcal{N}(0, \overline{V}_\infty).$$

6 Monte Carlo Experiments

This section reports evidence on the finite sample behavior of the bias correction methods in binary choice and count data models. We compare the performance of uncorrected and bias-corrected estimators in terms of bias, dispersion and mean squared error. We focus on Jackknife bias corrections because they are easy to implement using standard software routines. All the results are based on 500 replications.

6.1 Example 1: binary choice models

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

6.1.1 Static probit model

The data generating process is

$$Y_{it} = \mathbf{1}\left\{X_{it}\beta^0 + \alpha_i^0 + \gamma_t^0 > \varepsilon_{it}\right\}, \quad (i = 1, ..., N;\ t = 1, ..., T),$$

where $\alpha_i^0 \sim \mathcal{N}(0, 1/16)$, $\gamma_t^0 \sim \mathcal{N}(0, 1/16)$, $\varepsilon_{it} \sim \mathcal{N}(0, 1)$, and $\beta^0 = 1$. We consider two alternative designs for X_{it}: correlated and uncorrelated with the individual and time effects. In the first design, $X_{it} = X_{i,t-1}/2 + \alpha_i^0 + \gamma_t^0 + \upsilon_{it}$, $\upsilon_{it} \sim \mathcal{N}(0, 1/2)$, 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 $X_{i0} \sim \mathcal{N}(0,1)$. The unconditional variance of X_{it} is one in both designs. The variables α_i^0, γ_t^0, ε_{it}, υ_{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 1 reports bias, standard deviation and root mean squared error for uncorrected and bias corrected estimators of the probit coefficient β^0, and the average partial effect of X_{it}. We compute the average effect using the derivative approximation of equation (2.4) for $g(\alpha_i, \gamma_t) = \alpha_i + \gamma_t$. Throughout the table, MLE-FETE corresponds to the probit maximum likelihood estimator with individual and time fixed effects, Jackknife 1 is the bias corrected estimator that uses leave-one-out jackknife in the individual dimension and split jackknife in the time dimension; and Jackknife 2 is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension. The cross-sectional division in Jackknife 2 follows the order of the observations. All the results in the tables are reported in percentage of the true parameter value.

The results of the table show that the bias is of the same order of magnitude as the standard deviation for the uncorrected estimator of the probit coefficient. This result holds for both designs and all the sample sizes considered. The bias corrections, specially Jackknife 1, remove a large proportion of the bias without increasing dispersion, and produce substantial reductions in rmse. For example, Jackknife 1 reduces rmse by about 30 % in the correlated design. As in Hahn and Newey (2004) and Fernandez-Val (2009), despite the large bias in the probit coefficients, we find very little bias in the uncorrected estimates of the average marginal effect.

6.1.2 Dynamic probit model

The data generating process is

$$\begin{aligned} Y_{it} &= \mathbf{1}\left\{Y_{i,t-1}\beta_Y^0 + Z_{it}\beta_Z^0 + \alpha_i^0 + \gamma_t^0 > \varepsilon_{it}\right\}, \quad (i = 1,...,N;\ t = 1,...,T), \\ Y_{i0} &= \mathbf{1}\left\{Z_{i0}\beta_Z^0 + \alpha_i^0 + \gamma_0^0 > \varepsilon_{i0}\right\}, \end{aligned}$$

where $\alpha_i^0 \sim \mathcal{N}(0, 1/16)$, $\gamma_t^0 \sim \mathcal{N}(0, 1/16)$, $\varepsilon_{it} \sim \mathcal{N}(0,1)$, $\beta_Y^0 = 0.5$, and $\beta_Z^0 = 1$. We consider two alternative designs for Z_{it}: correlated an uncorrelated with the individual effects. In the first design, $Z_{it} = Z_{i,t-1}/2 + \alpha_i^0 + \gamma_t^0 + \upsilon_{it}$, $\upsilon_{it} \sim \mathcal{N}(0, 1/2)$, 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 α_i^0, γ_t^0, ε_{it}, υ_{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 2 reports bias, standard deviation and root mean squared error for uncorrected and bias corrected estimators of the probit coefficient β_Y^0 and the average partial effect of $Y_{i,t-1}$. We compute the partial effect of $Y_{i,t-1}$ using the expression in equation (2.3) for $X_{it,k} = Y_{i,t-1}$ and $g(\alpha_i, \gamma_t) = \alpha_i + \gamma_t$. This effect is commonly reported as a measure of state dependence for dynamic binary processes. Table 3 reports the same statistics for the estimators of the probit coefficient β_Z^0 and the average partial effect of Z_{it}. We compute the partial effect using the derivative approximation of equation (2.4) for $X_{it,k} = Z_{it}$ and $g(\alpha_i, \gamma_t) = \alpha_i + \gamma_t$. Throughout the tables, MLE-FETE corresponds to the probit maximum likelihood estimator with individual and time fixed effects, Jackknife 1 is the bias corrected estimator that uses leave-one-out jackknife in the individual dimension and split jackknife in the time dimension; and Jackknife 2 is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension. The cross-sectional division in Jackknife 2 follows the order of the observations. All the results in the tables are reported in percentage of the true parameter value.

The results in table 2 show important biases toward zero for *both* the probit coefficient and the average effect of $Y_{i,t-1}$ in both designs. This bias can indeed be substantially larger than the corresponding standard deviation for short panels. Jackknife 1 reduces bias with little increase of dispersion, reducing the rmse between 23 and 38 % for the coefficient and between 24 and 43 % for the average marginal effect. The results for the effects of Z_{it} in table 3 are similar to the static probit model. There is significant bias in the estimator of the coefficient, which is removed by the corrections, whereas there is little bias in the estimator of the average marginal effects. Jackknife 2 increases dispersion and is not very effective in short panels.

6.2 Example 2: count data 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 estimate 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.

6.2.1 Static Poisson model

The data generating process is

$$Y_{it} \mid X, \phi^0 \sim \mathcal{P}(\exp[Z_{it}\beta_1^0 + Z_{it}^2\beta_2^0 + \alpha_i^0 + \gamma_t^0]), \quad (i = 1, ..., N;\ t = 1, ..., T),$$

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 two different numbers of industries N: 17 and 34. In the second case, we double the cross-sectional size by merging two independent realizations of the panel.

Table 4 reports bias, standard deviation and root mean squared error for uncorrected and bias corrected estimators of the coefficients β_1^0 and β_2^0, and the average partial effect of Z_{it}. We compute the average effect using the expression (2.5) for $H(Z_{it}) = Z_{it}^2$ and $g(\alpha_i, \gamma_t) = \alpha_i + \gamma_t$. 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 probit, MLE-FETE corresponds to the Poisson maximum likelihood estimator with individual and time fixed effects, Jackknife 1 is the bias corrected estimator that uses leave-one-out jackknife in the individual dimension and split jackknife in the time dimension; and Jackknife 2 is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension. The cross-sectional division in Jackknife 2 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 theoretical finding of Section 4.1 for the Poisson model. Thus, the bias of MLE-FETE for the coefficients and average marginal effect is negligible relative to the standard deviation. The bias corrections increase 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.

6.2.2 Dynamic Poisson model

The data generating process is

$$Y_{it} \mid X, \phi^0 \sim \mathcal{P}(\exp[\beta_Y \log(1 + Y_{i,t-1}) + Z_{it}\beta_1^0 + Z_{it}^2\beta_2^0 + \alpha_i^0 + \gamma_t^0]),$$

$(i = 1, ..., N; t = 1, ..., T)$. 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) model with individual and time effects. We then draw panel data sets with $T = 21$ years and two different numbers of industries N: 17 and 34. In the second case, we double the cross-sectional size by merging two 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 5 reports bias, standard deviation and root mean squared error for uncorrected and bias corrected estimators of the coefficient β_Y^0 and the average partial effect of $Y_{i,t-1}$. We compute the partial effect of $Y_{i,t-1}$ using the derivative approximation in expression (2.5) for $Z_{it} = Y_{i,t-1}$, $H(Z_{it}) = \log(1 + Z_{it})$, dropping the linear term, and $g(\alpha_i, \gamma_t) = \alpha_i + \gamma_t$. This effect is commonly reported as a measure of state dependence for dynamic processes. Table 6 reports the same statistics for the estimators of the coefficients β_1^0 and β_2^0, and the average partial effect of Z_{it}. We compute the partial effect using the derivative approximation in expression (2.5) for $H(Z_{it}) = Z_{it}^2$ and $g(\alpha_i, \gamma_t) = \alpha_i + \gamma_t$. Throughout the tables, we compare the same estimators as for the static model. Again, all the results in the tables are reported in percentage of the true parameter value.

The results in table 5 show biases of the same order of magnitude as the standard deviation for the fixed effects estimators of the coefficient and average effect of $Y_{i,t-1}$. Jackknife 1 reduces bias with little increase of dispersion, reducing the rmse for both sample sizes. The results for the coefficient of Z_{it} in table 6 are similar to the static model. There is no significant bias and the corrections increase dispersion and rmse, less so for the larger sample size. The results for the average effect of Z_{it} are imprecise, because of the large variability of the partial effects.

7 Empirical Example

To illustrate the results we revisit the empirical application of Aghion, Bloom, Blundell, Griffith and Howitt (2005) that estimates a count data model to analyze the relationship between innovation and competition. They use 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

$$Y_{it} \sim \mathcal{P}(\exp[\beta_1 Z_{it} + \beta_2 Z_{it}^2 + \alpha_i + \gamma_t]), \quad (i = 1, ..., 17; t = 1973, ..., 1994);$$

and extend the analysis to a dynamic Poisson model with industry and year effects where

$$Y_{it} \sim \mathcal{P}(\exp[\beta_Y \log(1+Y_{i,t-1})+\beta_1 Z_{it}+\beta_2 Z_{it}^2+\alpha_i+\gamma_t]), \ (i = 1, ..., 17; t = 1974, ..., 1994).$$

In the dynamic model we use the year 1973 as the initial condition for Y_{it}.

Table 7 reports the results of the analysis. Columns (3) and (4) for the static model replicate the empirical results of Table I (p. 708) in ABBGH, adding estimates for the average partial effects. The bias corrected estimates in columns (5) and (6), while significantly different from the uncorrected estimates in column (3), agree with the inverted-U pattern in the relationship between innovation and competition. The difference between uncorrected and bias corrected estimates might be due to lack of strict exogeneity of the competition variable. 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 average partial effect of lag patents relative to the Jackknife bias corrections. Controlling for state dependence does not change the inverted-U pattern, but flattens the relationship between innovation and competition.

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 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. (2011). Multiple fixed effects in nonlinear panel data models. *Unpublished manuscript*.

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

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

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

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

Hahn, J. and Kuersteiner, G. (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.

Hausman, J., Hall, B., and Griliches, Z. (1984). Econometric models for count data with an application to the patents-r & d relationship. *Econometrica*, 52(4):909–938.

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

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

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

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

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

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

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

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

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

Table 1: Finite sample properties of static probit estimators (N = 52)

	Coefficient			Average Marginal Effect		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
Design 1: correlated individual and time effects						
			T = 14			
MLE-FETE	13	12	17	1	8	8
Jackknife 1	-3	12	12	1	9	9
Jackknife 2	-11	12	16	0	9	9
			T = 26			
MLE-FETE	8	8	11	0	6	6
Jackknife 1	-1	7	7	0	6	6
Jackknife 2	-4	7	8	0	6	6
			T = 52			
MLE-FETE	5	5	7	0	4	4
Jackknife 1	0	5	5	0	4	4
Jackknife 2	-1	5	5	0	4	4
Design 2: uncorrelated individual and time effects						
			T = 14			
MLE-FETE	12	9	15	0	5	5
Jackknife 1	-3	10	10	0	7	7
Jackknife 2	-10	10	14	-2	7	7
			T = 26			
MLE-FETE	7	6	10	0	4	4
Jackknife 1	-1	6	6	0	4	4
Jackknife 2	-3	6	7	-1	4	4
			T = 52			
MLE-FETE	5	4	6	0	2	2
Jackknife 1	0	4	4	0	2	2
Jackknife 2	-1	4	4	0	2	2

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 ε_{it} ~ i.i.d. N(0,1), α_i ~ i.i.d. N(0,1/16), γ_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 α_i y γ_t . Average marginal effect is β E[$\phi(\beta X_{it} + \alpha_i + \gamma_t)$], where $\phi()$ is the PDF of the standard normal distribution. MLE-FETE is the probit maximum likelihood estimator with individual and time fixed effects; Jackknife 1 is the bias corrected estimator that uses leave-one-out panel jackknife in the individual dimension and split panel jackknife in the time dimension; Jackknife 2 is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension.

Table 2: Finite sample properties of dynamic probit estimators (N = 52)

	Coefficient $Y_{i,t-1}$			Average Marginal Effect $Y_{i,t-1}$		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
Design 1: correlated individual and time effects						
			T = 14			
MLE-FETE	-44	30	53	-52	26	58
Jackknife 1	8	32	33	-5	32	33
Jackknife 2	16	38	42	-4	33	33
			T = 26			
MLE-FETE	-23	21	30	-29	19	35
Jackknife 1	1	21	21	-1	22	22
Jackknife 2	3	22	22	-1	23	23
			T = 52			
MLE-FETE	-9	14	17	-14	14	20
Jackknife 1	1	14	14	1	15	15
Jackknife 2	1	14	14	1	15	15
Design 2: uncorrelated individual and time effects						
			T = 14			
MLE-FETE	-35	28	44	-44	24	50
Jackknife 1	11	28	30	1	29	29
Jackknife 2	16	29	33	1	29	29
			T = 26			
MLE-FETE	-19	18	26	-26	17	31
Jackknife 1	1	18	18	-1	19	19
Jackknife 2	2	18	18	-1	19	19
			T = 52			
MLE-FETE	-7	13	15	-12	13	17
Jackknife 1	1	13	13	1	13	13
Jackknife 2	1	13	13	1	13	13

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 $Y_{i0} = 1(\beta_Z Z_{i0} + \alpha_i + \gamma_0 > \varepsilon_{i0})$, ε_{it} ~ i.i.d. N(0,1), α_i ~ i.i.d. N(0,1/16), γ_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 α_i y γ_t. Average marginal effect is β $E[\Phi(\beta_Y + \beta_Z Z_{it} + \alpha_i + \gamma_t) - \Phi(\beta_z Z_{it} + \alpha_i + \gamma_t)]$, where Φ() is the CDF of the standard normal distribution. MLE-FETE is the probit maximum likelihood estimator with individual and time fixed effects; Jackknife 1 is the bias corrected estimator that uses leave-one-out panel jackknife in the individual dimension and split panel jackknife in the time dimension; Jackknife 2 is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension.

Table 3: Finite sample properties of dynamic probit estimators (N = 52)

	Coefficient Z_{it}			Average Marginal Effect Z_{it}		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
Design 1: correlated individual and time effects						
			T = 14			
MLE-FETE	20	13	23	4	10	10
Jackknife 1	-2	14	14	4	11	12
Jackknife 2	-22	83	86	1	12	12
			T = 26			
MLE-FETE	10	8	13	2	7	7
Jackknife 1	-1	8	8	1	7	7
Jackknife 2	-6	8	10	0	7	7
			T = 52			
MLE-FETE	6	5	8	1	5	5
Jackknife 1	0	5	5	1	5	5
Jackknife 2	-2	5	5	0	5	5
Design 2: uncorrelated individual and time effects						
			T = 14			
MLE-FETE	17	11	20	3	6	7
Jackknife 1	-3	12	12	2	8	8
Jackknife 2	-15	14	20	0	8	8
			T = 26			
MLE-FETE	10	7	12	2	4	4
Jackknife 1	-1	7	7	0	5	5
Jackknife 2	-5	7	9	0	5	5
			T = 52			
MLE-FETE	6	4	7	1	3	3
Jackknife 1	-1	4	4	0	3	3
Jackknife 2	-2	4	5	0	3	3

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 $Y_{i0} = 1(\beta_Z Z_{i0} + \alpha_i + \gamma_0 > \varepsilon_{i0})$, $\varepsilon_{it} \sim$ i.i.d. $N(0,1)$, $\alpha_i \sim$ i.i.d. $N(0,1/16)$, $\gamma_t \sim$ i.i.d. $N(0, 1/16)$, $\beta_Y = 0.5$, and $\beta_Z = 1$. In design 1, $Z_{it} = Z_{i,t-1} / 2 + \alpha_i + \gamma_t + v_{it}$, $v_{it} \sim$ i.i.d. $N(0, 1/2)$, and $Z_{i0} \sim N(0,1)$. In design 2, $Z_{it} = Z_{i,t-1} / 2 + v_{it}$, $v_{it} \sim$ i.i.d. $N(0, 3/4)$, and $Z_{i0} \sim N(0,1)$, independent of α_i y γ_t .Average marginal effect is β $E[\varphi(\beta_Y Y_{i,t-1} + \beta_Z Z_{it} + \alpha_i + \gamma_t)]$, where $\varphi()$ is the PDF of the standard normal distribution. MLE-FETE is the probit maximum likelihood estimator with individual and time fixed effects; Jackknife 1 is the bias corrected estimator that uses leave-one-out panel jackknife in the individual dimension and split panel jackknife in the time dimension; Jackknife 2 is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension.

Table 4: Finite sample properties of static Poisson estimators

	Coefficient Z_{it}			Coefficient Z_{it}^2			Average Marginal Effect		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
				N = 17, T = 22, unbanced					
MLE	-59	14	60	-58	14	60	222	113	248
MLE-TE	-62	14	64	-62	14	64	-9	139	139
MLE-FETE	-2	17	17	-2	17	17	-15	226	226
Jackknife 1	-2	23	23	-3	23	23	-37	332	334
Jackknife 2	-4	27	27	-4	27	27	-9	339	339
				N = 34, T = 22, unbalanced					
MLE	-58	10	59	-57	10	58	226	81	240
MLE-TE	-61	10	62	-61	10	62	-3	97	97
MLE-FETE	0	12	12	0	13	13	-6	158	158
Jackknife 1	-1	14	14	-1	14	14	-15	213	213
Jackknife 2	-1	14	14	-1	14	14	-14	206	206

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 marginal 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; Jackknife 1 is the bias corrected estimator that uses leave-one-out panel jackknife in the individual dimension and split panel jackknife in the time dimension; Jackknife 2 is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension.

Table 5: Finite sample properties of dynamic Poisson estimators

	Coefficient $Y_{i,t-1}$			Average Marginal Effect $Y_{i,t-1}$		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
			N = 17,T = 21, unbalanced			
MLE	135	3	135	158	2	158
MLE-TE	142	3	142	163	3	163
MLE-FETE	-17	15	23	-17	15	22
Jackknife 1	2	20	20	2	20	20
Jackknife 2	6	21	22	7	20	22
			N = 34, T = 21, unbalanced			
MLE	135	2	135	158	2	158
MLE-TE	141	2	141	162	2	162
MLE-FETE	-16	11	19	-16	10	19
Jackknife 1	2	13	14	2	13	13
Jackknife 2	4	13	14	4	13	14

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 marginal effect is $\beta_Y\ E[\exp\{(\beta_Y \log(1 + Y_{i,t-1}) + \beta_1 Z_{it} + \beta_2 Z_{it}^2 + \alpha_i + \gamma_t\}]$. 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; Jackknife 1 is the bias corrected estimator that uses leave-one-out panel jackknife in the individual dimension and split panel jackknife in the time dimension; Jackknife 2 is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension.

Table 6: Finite sample properties of dynamic Poisson estimators

	Coefficient Z_{it}			Coefficient Z_{it}^2			Average Marginal Effect Z_{it}		
	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE	Bias	Std. Dev.	RMSE
				N = 17, T = 21, unbalanced					
MLE	-76	27	81	-76	27	80	760	351	837
MLE-TE	-65	28	71	-65	29	71	541	356	647
MLE-FETE	9	40	41	9	41	42	-3	1151	1150
Jackknife 1	3	54	54	4	54	54	43	1659	1658
Jackknife 2	1	63	63	1	63	63	-44	1682	1681
				N = 34, T = 21, unbalanced					
MLE	-75	19	77	-74	19	77	777	252	817
MLE-TE	-65	19	67	-64	19	67	534	248	589
MLE-FETE	6	28	28	6	28	29	-68	734	736
Jackknife 1	3	32	32	3	33	33	-25	1029	1028
Jackknife 2	2	31	31	2	32	32	-30	1012	1011

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 marginal 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; Jackknife 1 is the bias corrected estimator that uses leave-one-out panel jackknife in the individual dimension and split panel jackknife in the time dimension; Jackknife 2 is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension.

Table 7: Poisson model for patents

Dependent variable: citation-weighted patents	(1)	(2)	(3)	(4)	(5)	(6)
Static model						
Competition	165.12	393.03	152.81	387.46	675.67	405.36
	(54.77)	(63.13)	(55.74)	(67.74)		
	-20.00	*-37.76*	*-6.43*	*-5.98*	*-16.47*	*-16.23*
Competition squared	-88.55	-210.00	-80.99	-204.55	-357.99	-217.70
	(29.08)	(33.55)	(29.61)	(36.17)		
Dynamic model						
Lag-patents	1.05	0.45	1.07	0.46	0.75	0.62
	(0.02)	(0.05)	(0.03)	(0.05)		
	0.86	*0.35*	*0.87*	*0.36*	*0.58*	*0.51*
Competition	62.95	205.65	95.70	199.68	324.48	-19.00
	(62.68)	(71.42)	(65.08)	(76.66)		
	-12.78	*-24.72*	*-9.03*	*-1.68*	*-24.56*	*-31.22*
Competition squared	-34.15	-110.21	-51.09	-105.24	-173.47	4.48
	(33.21)	(37.95)	(34.48)	(40.87)		
Year effects			Yes	Yes	Yes	Yes
Industry effects		Yes		Yes	Yes	Yes
Bias correction					Jackknife 1	Jackknife 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 (3) and (4) replicate the results in ABBGH. Jackknife 1 is the bias corrected estimator that uses leave-one-out panel jackknife in the individual dimension and split panel jackknife in the time dimension; Jackknife 2 is the bias corrected estimator that uses split panel jackknife in both the individual and time dimension. Standard errors in parentheses and average partial effects in italics.