

Fixed effect estimation of large T panel data models

Iván Fernández-Val
Martin Weidner

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP42/17

Fixed Effect Estimation of Large T Panel Data Models*

Iván Fernández-Val[‡]

Martin Weidner[§]

September 29, 2017

Abstract

This article reviews recent advances in fixed effect estimation of panel data models for long panels, where the number of time periods is relatively large. We focus on semiparametric models with unobserved individual and time effects, where the distribution of the outcome variable conditional on covariates and unobserved effects is specified parametrically, while the distribution of the unobserved effects is left unrestricted. Compared to existing reviews on long panels (Arellano & Hahn, 2007; a section in Arellano & Bonhomme, 2011) we discuss models with both individual and time effects, split-panel Jackknife bias corrections, unbalanced panels, distribution and quantile effects, and other extensions. Understanding and correcting the incidental parameter bias caused by the estimation of many fixed effects is our main focus, and the unifying theme is that the order of this bias is given by the simple formula p/n for all models discussed, with p the number of estimated parameters and n the total sample size.

Keywords: panel data, fixed effects, incidental parameter problem, bias correction, unobserved heterogeneity, nonlinear models, jackknife.

*We thank Stéphane Bonhomme, Geert Dhaene, and Koen Jochmans for helpful comments, and we are grateful to Siyi Luo for proofreading, and to Mario Cruz-Gonzalez for helping with the Monte Carlo simulations. 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-589-28-0001, and the European Research Council grant ERC-2014-CoG-646917-ROMIA 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

One of the main advantages of panel data over cross-sectional or time series data is the possibility of accounting for multiple sources of unobserved heterogeneity. This is accomplished by including individual and time unobserved effects into the model, which control for any unobserved covariate that is either time or cross-sectional invariant. In this article we consider fixed effect approaches to panel data models, where no distributional assumption on the unobserved effects is imposed, therefore allowing the unobserved effects to be arbitrarily related with the observed covariates. Fixed effects contrast with random effects approaches that impose restrictions on the distribution of the unobserved effects conditional on the observed covariates. We refer to Arellano (2003b), Baltagi (2008), Hsiao (2014), and Wooldridge (2010) for modern textbook treatments on the difference between fixed and random effects.

We consider semiparametric models that specify parametrically some characteristic of the distribution of the outcome variable of interest conditional on the observed covariates and unobserved effects. Examples include linear and nonlinear regression models for the conditional expectation, and generalized linear, distribution and quantile regression models for the conditional distribution. Generalized linear models include the most commonly used nonlinear models such as probit, logit, Poisson, negative binomial, proportional hazard, and tobit models. The nonparametric or unspecified part of the model consists of the joint distribution of the unobserved effects, exogenous covariates and initial conditions. We do not review nonparametric panel models such as the nonseparable models considered in Altonji & Matzkin (2005), Evdokimov (2010), Graham & Powell (2012), Hoderlein & White (2012), Chernozhukov, Fernández-Val, Hahn & Newey (2013), Chernozhukov, Fernandez-Val, Hoderlein, Holzmann & Newey (2015), Chernozhukov, Fernández-Val & Newey (2017a), Freyberger (2017), and Torgovitsky (2016), among others. We refer to Matzkin (2007) for an excellent survey on nonparametric identification of nonseparable models including results for panel data.

We analyze the properties of fixed effects estimators of model parameters. These estimators treat the unobserved effects as parameters to be estimated. In linear models with strictly exogenous covariates and individual effects, fixed effects is numerically equivalent to the within-group estimator that removes the individual effects by taking differences within each individual. We also analyze fixed effects estimation of average partial effects (APEs), which are averages of functions of the data, parameters and unobserved effects. APEs are often the quantities of interest in nonlinear models because they correspond to marginal effects of the covariates in some characteristic of the distribution of the outcome conditional on covariates and unobserved effects, averaged over the observed and unobserved heterogeneity. The main challenge with fixed effects estimators is to deal with the incidental parameter problem coming from the estimation of the unobserved effects. The estimated unobserved effects, usually called fixed effects, can be very noisy because there are few observations that are informative about them. Section 2 describes the incidental parameter problem in detail.

Motivated by the first panel data sets available, the econometric theory was initially designed for panels where the number of time periods, T , is small compared to the number of cross-sectional units, N . This theory analyzed identification, estimation and inference using a fixed- T asymptotic approximation that considers sequences of panels with increasing N and fixed T . Consistent approaches under this approximation usually consist of removing the unobserved effects from the model by some smart transformation such as differencing or conditioning on sufficient statistics. Examples include Mundlak (1978), Arellano & Bond (1991) and Moreira (2009) for linear models; Cox (1958), Rasch (1960), Andersen (1970), Chamberlain (1980), Manski (1987), Horowitz (1992), Honoré & Kyriazidou (2000), and Arellano & Carrasco (2003) for static and dynamic binary choice models; Hausman, Hall & Griliches (1984) for static Poisson models; Honoré (1992) for tobit models; Kyriazidou (1997, 2001) for sample selection models; and Muris (2017) and Botosaru & Muris (2017) for ordered logit and probit models; Bonhomme (2012) provided a unifying scheme for many of those methods.

The fixed- T approach has encountered several limitations. First, transformations to remove the unobserved effects are not available for some models such as probit and quantile regression. This issue gets worse in models with unobserved effects in multiple dimensions, where even if sufficient statistics exist conditioning on them can be computationally challenging (Hirji, Mehta & Patel, 1987; Charbonneau, 2014). Second, differencing over time in dynamic linear models does not work well when the data are very persistent because it removes most of the information about the parameters. Third, removing the unobserved effects from the model precludes the estimation of APEs in nonlinear models. Fourth, under fixed- T asymptotics the incidental parameter problem of fixed effects estimation is a consistency problem, which is difficult to tackle. Fifth, some models are not \sqrt{N} -estimable or not even point identified under the fixed- T approximation, see, e.g., Chamberlain (2010), Honoré & Tamer (2006), Chernozhukov, Fernández-Val, Hahn & Newey (2013), Shi, Shum & Song (2014), and Pakes & Porter (2013).

All the limitations of the fixed- T approximation mentioned in the previous paragraph, together with the recent availability of long panel datasets, has led to an alternative asymptotic approximation to panel data that considers sequences of panels where both N and T increase. This large- T approximation, developed by Phillips & Moon (1999) for linear models, deals with some of the shortcomings of the fixed- T approximation. For example, most models are point identified with large- T and the incidental parameter problem of fixed effects estimation becomes an asymptotic bias problem that is easier to tackle. Bias corrections have been developed to provide improved fixed effects estimators of model parameters and APEs. This large- T approximation can also be applied naturally to other types of data with a grouping structure similar to panel data such as network and trade data.

The two alternative asymptotic approximations to panel data should be viewed as complements rather than substitutes. For example, if a fixed T consistent estimator is available for the particular parameter under consideration, then it generally should be used. Otherwise it is likely that the model is not point identified for fixed- T . Then, one can either consider bounding the parameter if T is small, or drawing on

a large- T consistent estimator if T is moderately large with respect to N . Honoré & Tamer (2006) and Chernozhukov, Fernández-Val, Hahn & Newey (2013) showed that in some binary response models the bounds for parameters and APEs can be very tight and shrink very rapidly with T . Thus, unless T is very small, one can safely ignore that the model might only be set-identified and use large T consistent estimators. Alternatively, one might entertain (correlated) random effects approaches that achieve fixed- T point identification by imposing restrictions on the distribution of the unobserved effects.

This review focuses solely on fixed effects estimators of semiparametric panel models analyzed under a large- T asymptotic approximation. Available survey articles on panel data estimation under fixed- T approximations include Chamberlain (1984), Arellano & Honoré (2001), Honoré (2002), Arellano & Bonhomme (2011), and Arellano & Bonhomme (2017). We concentrate mainly on the developments since Arellano & Hahn (2007), which previously reviewed the large- T approach to panel data. This field has grown sufficiently fast in recent years to justify another review. In particular, Arellano & Hahn (2007) considered almost exclusively models with only unobserved individual effects, whereas we consider models with unobserved effects in multiple dimensions such as individual and time effects. We also review the estimation of distributional and quantile effects with panel data, which were not discussed in Arellano & Hahn (2007). Our review relies heavily upon the important foundational work in this literature by Neyman & Scott (1948), Nickell (1981), Cox & Reid (1987), Kiviet (1995), Phillips & Moon (1999), Lancaster (2000), Hahn & Kuersteiner (2002), Lancaster (2002), Woutersen (2002), Arellano (2003a), Alvarez & Arellano (2003), Li, Lindsay & Waterman (2003), Hahn & Newey (2004), Carro (2007), Fernández-Val (2009), Hahn & Kuersteiner (2011), Fernández-Val & Lee (2013) and Dhaene & Jochmans (2015b), among others.

Outline: In Section 2 we describe the bias of maximum likelihood estimators (MLEs) in cross-sectional samples, which allows us to establish some key results in a simplified setting. Those results are then generalized to panel data models in Section 3, where we discuss the incidental parameter bias and bias correction methods for fixed effected estimators of model parameters and average partial effects. A short heuristic derivation of the main formulas in Section 3 is provided in the appendix. Section 4 discusses unbalanced panels, multivariate fixed effects, distributional and quantile effects, and some further extensions of the methods, before we conclude in Section 5.

Notation: In the following we denote by \mathbb{E} the expectation with respect to the parametric part of the model. Averages over cross-sectional and time-series samples are denoted by $\mathbb{E}_n = n^{-1} \sum_{j=1}^n$, $\mathbb{E}_N = N^{-1} \sum_{i=1}^N$, $\mathbb{E}_T = T^{-1} \sum_{t=1}^T$, and $\mathbb{E}_{NT} = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T$. For example, we write $\mathbb{E}_n \ell_j(\theta)$ for $n^{-1} \sum_{j=1}^n \ell_j(\theta)$ in the next section. We also define $\mathbb{G}_n = \sqrt{n}(\mathbb{E}_n - \mathbb{E})$, $\mathbb{G}_N = \sqrt{N}(\mathbb{E}_N - \mathbb{E}_N \mathbb{E})$, $\mathbb{G}_T = \sqrt{T}(\mathbb{E}_T - \mathbb{E}_T \mathbb{E})$, and $\mathbb{G}_{NT} = \sqrt{NT}(\mathbb{E}_{NT} - \mathbb{E}_{NT} \mathbb{E})$. The notations \mathbb{E}_n and \mathbb{G}_n are used in Section 2 for random samples, whereas \mathbb{E}_N , \mathbb{E}_T , \mathbb{G}_N , \mathbb{G}_T and \mathbb{G}_{NT} are used in Section 3 for nonrandom panel samples, which explains the additional sample averages before the expectations \mathbb{E} . For example $\mathbb{G}_{NT} \ell_{it} = \sqrt{NT} \mathbb{E}_{NT}(\ell_{it} - \mathbb{E} \ell_{it})$.

In Section 2 we often just write a bar to denote the expectation E , e.g. $\bar{\ell}_{it} = E\ell_{it}$. Vector-valued variables are always arranged as column-vectors in the following. Thus, we write $\theta = (\beta, \alpha, \gamma)$ to form the column-vector θ as the concatenation of the column-vectors β , α and γ . Vector and matrix transposition is denoted by a prime. We use $\overset{a}{\sim}$ to denote asymptotic approximation to the distribution. For example, if $\sqrt{n}(\hat{\beta} - \beta_0 - B/a_n) \rightarrow_d \mathcal{N}(0, V)$ for some sequence a_n such that $a_n \rightarrow \infty$ as $n \rightarrow \infty$, then we write $\hat{\beta} - \beta_0 \overset{a}{\sim} \mathcal{N}(B/a_n, V/n)$.

2 THE INCIDENTAL PARAMETER PROBLEM

Incidental parameters are nuisance parameters whose dimension grows with the sample size. Neyman & Scott (1948) showed that maximum likelihood estimators (MLEs) can be asymptotically biased in models with incidental parameters. Here we provide a derivation based on a second-order asymptotic expansion that the order of the bias of the MLEs in random samples is

$$\text{bias} \sim \frac{p}{n}, \quad (1)$$

where p is the number of parameters and n is the sample size. This order corresponds with the inverse of the number of observations per parameter. In the following sections we will see that the order of the bias of fixed effects estimators of parameters and APEs can also be obtained using this simple heuristic formula in more complex panel data samples.

2.1 Bias of Maximum Likelihood Estimators

Given a random sample $\{z_j : 1 \leq j \leq n\}$ of a random variable z that has density $f(z, \theta_0)$ with respect to some dominating measure, let

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^p}{\operatorname{argmax}} \mathbb{E}_n \ell_j(\theta), \quad \ell_j(\theta) := \log f(z_j, \theta), \quad (2)$$

be the maximum likelihood estimator (MLE) of the value of the p -dimensional parameter θ that is identified by the population program

$$\theta_0 = \underset{\theta \in \mathbb{R}^p}{\operatorname{argmax}} E\ell_j(\theta). \quad (3)$$

In the following expansions we assume that θ_0 is uniquely determined by the first order conditions of the program (3), $\theta \mapsto \ell_j(\theta)$ is a.s. differentiable to sufficient order, and other regularity conditions that allow us to bound remainder terms. We also assume that $\hat{\theta}$ is consistent, i.e., $\hat{\theta} \rightarrow_P \theta_0$ under some norm. Establishing this consistency might require delicate arguments in high dimensional settings where we let $p \rightarrow \infty$ as $n \rightarrow \infty$. We denote derivatives of $\theta \mapsto \ell_j(\theta)$ using superscripts θ , that is, $\ell_j^\theta(\theta)$ is the p -dimensional gradient, $\ell_j^{\theta\theta}(\theta)$ is the $p \times p$ Hessian matrix, and, for $k \in \{1, \dots, p\}$, $\ell_j^{\theta\theta k}(\theta)$ is the $p \times p$ matrix

obtained by taking the partial derivative of the Hessian with respect to θ_k . We drop the argument θ from all the functions whenever they are evaluated at θ_0 , e.g. $\ell_j^\theta := \ell_j^\theta(\theta_0)$.

By consistency, the asymptotic properties of $\hat{\theta}$ are governed by the local properties of $\ell_j(\theta)$ around θ_0 . A second order Taylor expansion of the first order condition of (2) around $\theta = \theta_0$ yields

$$0 = \mathbb{E}_n \ell_j^\theta(\hat{\theta}) \approx \mathbb{E}_n \ell_j^\theta + \mathbb{E}_n \ell_j^{\theta\theta} (\hat{\theta} - \theta_0) + \frac{1}{2} \sum_{k=1}^p \mathbb{E}_n \ell_j^{\theta\theta\theta_k} (\hat{\theta} - \theta_0)(\hat{\theta}_k - \theta_{0,k}).$$

We have $\mathbb{E}_n \ell_j^\theta = n^{-1/2} \mathbb{G}_n \ell_j^\theta$, because $\mathbb{E} \ell_j^\theta = 0$ by the first order conditions of the population program (3). Decomposing $\mathbb{E}_n \ell_j^{\theta\theta} = \mathbb{E} \ell_j^{\theta\theta} + n^{-1/2} \mathbb{G}_n \ell_j^{\theta\theta}$ and $\mathbb{E}_n \ell_j^{\theta\theta\theta_k} = \mathbb{E} \ell_j^{\theta\theta\theta_k} + n^{-1/2} \mathbb{G}_n \ell_j^{\theta\theta\theta_k}$, and ignoring the term that includes $\mathbb{G}_n \ell_j^{\theta\theta\theta_k}$ because it is of small enough order, gives

$$n^{-1/2} \mathbb{G}_n \ell_j^\theta + \mathbb{E} \ell_j^{\theta\theta} (\hat{\theta} - \theta_0) + n^{-1/2} \mathbb{G}_n \ell_j^{\theta\theta} (\hat{\theta} - \theta_0) + \frac{1}{2} \sum_{k=1}^p \mathbb{E} \ell_j^{\theta\theta\theta_k} (\hat{\theta} - \theta_0)(\hat{\theta}_k - \theta_{0,k}) \approx 0. \quad (4)$$

The leading two terms of (4) give the standard *first-order* approximation $\hat{\theta} - \theta_0 \approx n^{-1/2} \psi_1$, where $\psi_1 := -(\mathbb{E} \ell_j^{\theta\theta})^{-1} \mathbb{G}_n \ell_j^\theta$ is the influence function. Plugging this first-order approximation into the third and fourth term of (4) and dropping low order terms yields the *second-order* approximation

$$\hat{\theta} - \theta_0 \approx n^{-1/2} \psi_1 + n^{-1} \psi_2, \quad (5)$$

where

$$\psi_2 := -(\mathbb{E} \ell_j^{\theta\theta})^{-1} \mathbb{G}_n \ell_j^{\theta\theta} \psi_1 - \frac{1}{2} \sum_{k=1}^p (\mathbb{E} \ell_j^{\theta\theta})^{-1} \mathbb{E} \ell_j^{\theta\theta\theta_k} \psi_1 \psi_{1,k}.$$

We now analyze the properties of the components of $\psi_1 = (\psi_{1,1}, \dots, \psi_{1,p})$ and $\psi_2 = (\psi_{2,1}, \dots, \psi_{2,p})$. For each $k \in \{1, \dots, p\}$, $\psi_{1,k}$ has zero mean, finite variance and asymptotic normal distribution under standard assumptions. Thus, ψ_1 is a variance term that does not contribute to the bias of $\hat{\theta}$. To determine the mean of the quadratic term $\psi_{2,k}$, we assume for simplicity that $\mathbb{E} \ell_j^{\theta\theta}$ is a diagonal matrix.¹ Then,

$$\mathbb{E} \psi_{2,k} = \sum_{l=1}^p \frac{\mathbb{E} \left(\ell_j^{\theta_k \theta_l} \ell_j^{\theta_l} + \frac{1}{2} \ell_j^{\theta_k \theta_l \theta_l} \right)}{\mathbb{E} \ell_j^{\theta_k \theta_k} \mathbb{E} \ell_j^{\theta_l \theta_l}}, \quad (6)$$

where we use the information equality $\mathbb{E} \ell_j^{\theta_l} \ell_j^{\theta_m} = -\mathbb{E} \ell_j^{\theta_l \theta_m}$ to simplify the expression. This result shows that $\mathbb{E} \psi_{2,k}$ is proportional to p , because of the sum over $l = 1, \dots, p$, and therefore

$$\mathbb{E} \hat{\theta} - \theta_0 \approx \underbrace{n^{-1/2} \mathbb{E} \psi_1}_{=0} + n^{-1} \mathbb{E} \psi_2 \sim n^{-1} p,$$

which verifies the order of the bias in (1). The term ψ_2 also has variance. However, the order of the variance of the term $n^{-1} \psi_2$ is lower than the order of the variance of the term $n^{-1/2} \psi_1$ in the expansion (5).

¹This simplification is without loss of generality because we can diagonalize the Hessian by reparametrizing the model.

2.2 Consequences for Inference

The asymptotic bias of the MLE has consequences for inference. Confidence intervals based on the first-order asymptotic distribution might severely undercover even in large samples if the sample size is not sufficiently large relative to the dimension of the parameters, more precisely when $n = O(p^2)$. To see this, let $\rho'\theta$, $\rho \in \mathbb{R}^p$, be the inferential object of interest, and also let $E\psi_2/p = B = (B_1, \dots, B_p)$. By standard theory for MLE, $\rho'\psi_1 \rightarrow_d \mathcal{N}(0, \rho'\Omega\rho)$ with $\Omega = -\left(E\ell_j^{\theta\theta}\right)^{-1}$, and $\rho'\psi_2/p \rightarrow_p \rho'B$, as $n \rightarrow \infty$. Then, as $n \rightarrow \infty$,

$$\rho'(\hat{\theta} - \theta_0) \overset{a}{\sim} \mathcal{N}\left(\frac{p}{n}\rho'B, \frac{1}{n}\rho'\Omega\rho\right).$$

Consider $\rho = e_k$, the unit vector with a one in position k . The standard asymptotic two-sided $(1 - \alpha)$ -confidence interval for $\theta_k = e_k'\theta$ is

$$CI_{1-\alpha}(\theta_k) = \hat{\theta}_k \pm z_{\alpha/2}\hat{\Omega}_{kk}^{1/2}/\sqrt{n},$$

where $\hat{\Omega}_{kk}$ is a consistent estimator of Ω_{kk} , the (k, k) element of Ω , and $z_{\alpha/2}$ is the $(1 - \alpha/2)$ -quantile of the standard normal distribution.

Consider $p/\sqrt{n} \rightarrow \kappa$ as $n \rightarrow \infty$. Then, the coverage of $CI_{1-\alpha}(\theta_k)$ in large samples is

$$\begin{aligned} \Pr(\theta_k \in CI_{1-\alpha}(\theta_k)) &= \Pr(\sqrt{n}|\hat{\theta}_k - \theta_k| \leq z_{\alpha/2}\hat{\Omega}_{kk}^{1/2}) \rightarrow \Pr(|\mathcal{N}(\kappa B_k, \Omega_{kk})| \leq z_{\alpha/2}\Omega_{kk}^{1/2}) \\ &= \Phi(z_{\alpha/2} - \kappa B_k/\Omega_{kk}^{1/2}) - \Phi(-z_{\alpha/2} - \kappa B_k/\Omega_{kk}^{1/2}) < 1 - \alpha \end{aligned}$$

if $\kappa B_k \neq 0$. This coverage can be much lower than the nominal level $1 - \alpha$. For example, if $\kappa B_k/\Omega_{kk}^{1/2} = 1$, then the coverage probability of a 95% interval is less than 83% in large samples.

3 SEMIPARAMETRIC PANEL MODELS

3.1 Model

We observe the panel data set $\{(y_{it}, x_{it}) : 1 \leq i \leq N, 1 \leq t \leq T\}$, for a scalar outcome variable of interest y_{it} and a vector of covariates x_{it} . The subscripts i and t index individuals and time periods in traditional panels, but they might index other dimensions in more general data structures such as firms and industries or siblings and families. The observations are independent across i and weakly dependent across t .² We consider the semiparametric model for each $i = 1, \dots, N$:

$$y_{it} \mid x_i^t, \alpha, \gamma \sim f(\cdot \mid x_{it}, \alpha_i, \gamma_t; \beta), \text{ independently over } t = 1, \dots, T, \quad (7)$$

where $x_i^t = (x_{i1}, \dots, x_{it})$, $\alpha = (\alpha_1, \dots, \alpha_N)$, $\gamma = (\gamma_1, \dots, \gamma_T)$, f is a known density with respect to some dominating measure, and β is a d_β -vector of parameters. The variables α_i and γ_t are scalar unobserved

²Long panel models with cross-sectional dependence are discussed in Pakel (2014).

individual and time effects that in economic applications capture individual heterogeneity and aggregate shocks, respectively.³ The model is semiparametric because it does not specify the distribution of these effects nor their relationship with the covariates. The conditional density f represents the parametric part of the model. The covariates x_{it} are predetermined with respect to y_{it} and might include lags of y_{it} to accommodate dynamic models. If the covariates x_{it} are strictly exogenous with respect to y_{it} , then x_i^t can be replaced by $x_i^T = (x_{i1}, \dots, x_{iT})$ in the conditioning set. Strict exogeneity rules out dynamics and more generally any feedback from the outcome to future values of the covariates, which might be restrictive in applications.

Example (i) The simplest example of this model is the normal linear model with additive individual and time effects $y_{it} = x_{it}'\delta + \alpha_i + \gamma_t + \varepsilon_{it}$, $\varepsilon_{it} \mid x_i^t, \alpha, \gamma \sim \mathcal{N}(0, \sigma^2)$, where

$$f(y \mid x_{it}, \alpha_i, \gamma_t; \beta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y - x_{it}'\delta - \alpha_i - \gamma_t)^2}{2\sigma^2}\right], \quad \beta = (\delta, \sigma^2). \quad (8)$$

Example (ii) An example of a nonlinear model is the panel binary response single index model with additive unobserved individual and time effects:

$$y_{it} = 1(x_{it}'\beta + \alpha_i + \gamma_t \geq \varepsilon_{it}), \quad \varepsilon_{it} \mid x_i^t, \alpha, \gamma \sim F_\varepsilon,$$

where F_ε is a known CDF such as the standard normal or logistic. In this case,

$$f(y \mid x_{it}, \alpha_i, \gamma_t; \beta) = F_\varepsilon(x_{it}'\beta + \alpha_i + \gamma_t)^y \times [1 - F_\varepsilon(x_{it}'\beta + \alpha_i + \gamma_t)]^{1-y} \times 1(y \in \{0, 1\}). \quad (9)$$

Example (iii) Another example is a panel Poisson count response single index model with additive unobserved individual and time effects where, for $\lambda_{it} = \exp(x_{it}'\beta + \alpha_i + \gamma_t)$,

$$f(y \mid x_{it}, \alpha_i, \gamma_t; \beta) = \frac{\lambda_{it}^y \exp(-\lambda_{it})}{y!} 1(y \in \{0, 1, \dots\}). \quad (10)$$

Let $\beta_0, \alpha_0 = (\alpha_{01}, \dots, \alpha_{0N})$, and $\gamma_0 = (\gamma_{01}, \dots, \gamma_{0T})$ denote the values of β, α and γ that generate the data. We assume that these values are identified by the population conditional maximum likelihood program

$$(\beta_0, \alpha_0, \gamma_0) \in \underset{(\beta, \alpha, \gamma) \in \mathbb{R}^{d_\beta + N + T}}{\operatorname{argmax}} \mathbb{E}_{NT} \bar{\ell}_{it}(\beta, \alpha_i, \gamma_t), \quad \ell_{it}(\beta, \alpha_i, \gamma_t) = \log f(y_{it} \mid x_{it}, \alpha_i, \gamma_t; \beta), \quad (11)$$

where $\bar{\ell}_{it}(\beta, \alpha_i, \gamma_t) := \mathbb{E} \ell_{it}(\beta, \alpha_i, \gamma_t)$, the expected conditional log-likelihood with respect to the parametric part of the model. This program can have multiple solutions for α_0 and γ_0 . For example, in models that

³We refer to Fernández-Val & Lee (2013) for GMM panel models defined by moment conditions. We discuss models with multivariate individual and time effects in Section 4.2.

are additively separable in α_i and γ_t such as the single index models above, $\ell_{it}(\beta, \alpha_i, \gamma_t) = \ell_{it}(\beta, \alpha_i + \gamma_t)$, the location translation $\alpha_i + c$ and $\gamma_t - c$ for a constant c does not change the objective function of the program. In this case we assume that there exists a normalization that selects α_0 and γ_0 (e.g. $\alpha_{01} = 0$ or $\gamma_{01} = 0$).

3.2 Fixed Effects Estimator and Incidental Parameter Problem

The fixed effects (FE) estimator treats the realizations of α and γ as parameters to be estimated. It is the solution to the sample conditional maximum likelihood program, for $\hat{\alpha} = (\hat{\alpha}_1, \dots, \hat{\alpha}_N)$ and $\hat{\gamma} = (\hat{\gamma}_1, \dots, \hat{\gamma}_T)$,

$$(\hat{\beta}, \hat{\alpha}, \hat{\gamma}) \in \underset{(\beta, \alpha, \gamma) \in \mathbb{R}^{d_\beta + N + T}}{\operatorname{argmax}} \mathbb{E}_{NT} \ell_{it}(\beta, \alpha_i, \gamma_t), \quad (12)$$

which is the sample analog of (11). This program can also have multiple solutions for $\hat{\alpha}$ and $\hat{\gamma}$. In that case we adopt the same normalization as in the population program to select $\hat{\alpha}$ and $\hat{\gamma}$. In the binary and count response examples, we can obtain the FE estimator using standard software routines including individual and time indicators for the unobserved effects. We discuss some computational aspects of the program (12) in Section 3.6.

The FE estimator suffers from the incidental parameter problem (Neyman & Scott, 1948, see also Lancaster, 2000 for a review). The incidental parameters are the individual and time fixed effects which have dimensions N and T , respectively. Unless T is large, each individual fixed effect is very noisy because only T observations are informative about it. Symmetrically, unless N is large, each time fixed effect is very noisy. The noise in the fixed effects generally contaminates the estimators of the other parameters. Exceptions include linear and Poisson models with strictly exogenous covariates where it is possible to separate the estimation of the fixed effects from the other parameters. The asymptotic consequences of the incidental parameter problem depend on the approximation adopted. It is a consistency problem under a fixed- T or fixed- N approximation, whereas it becomes an asymptotic bias problem under a large- T and large- N approximation, which we adopt here.

3.3 Asymptotic Bias

The FE program (12) can be seen as a special case of (2) with $j = it$, $n = NT$, $\theta = (\beta, \alpha, \gamma)$ and $p = d_\beta + N + T$. However, the random sampling assumption of Section 2.1 is not plausible for panel data. The presence of the unobserved effects introduces heterogeneity in both dimensions, and assuming independence is often too strong when one of the dimensions is time. Some adjustments to the expansion in (5) are therefore required. We can still derive a second-order expansion $\hat{\theta} - \theta_0 \approx (NT)^{-1/2} \psi_1 + (NT)^{-1} \psi_2$ for $\hat{\theta} = (\hat{\beta}, \hat{\alpha}, \hat{\gamma})$ and $\theta_0 = (\beta_0, \alpha_0, \gamma_0)$, but the expressions of ψ_1 and ψ_2 need to be modified to account for heterogeneity and weak serial dependence. We provide these modifications in the appendix.

To characterize the asymptotic bias of the component $\widehat{\beta}$ of $\widehat{\theta}$, it is convenient to make $\ell_{it}(\beta, \alpha, \gamma)$ information-orthogonal between β and the rest of the parameters. This can be achieved by the transformation⁴

$$\ell_{it}^*(\beta, \alpha_i, \gamma_t) := \ell_{it}(\beta, \alpha_i + \kappa'_i \beta, \gamma_t + \rho'_t \beta), \quad (13)$$

where the d_β -vectors κ_i and ρ_t are a solution to the system of equations⁵

$$\begin{aligned} \mathbb{E}_T \left[\bar{\ell}_{it}^{\beta \alpha_i} + \kappa_i \bar{\ell}_{it}^{\alpha_i \alpha_i} + \rho_t \bar{\ell}_{it}^{\alpha_i \gamma_t} \right] &= 0, & i = 1, \dots, N, \\ \mathbb{E}_N \left[\bar{\ell}_{it}^{\beta \gamma_t} + \kappa_i \bar{\ell}_{it}^{\gamma_t \alpha_i} + \rho_t \bar{\ell}_{it}^{\gamma_t \gamma_t} \right] &= 0, & t = 1, \dots, T. \end{aligned} \quad (14)$$

Here, as in Section 2.1, we drop the arguments of the partial derivatives of ℓ_{it} when they are evaluated at the true values $(\beta_0, \alpha_{0i}, \gamma_{0t})$, and analogously we drop the arguments of partial derivatives of ℓ_{it}^* when they are evaluated at the transformed true values $(\beta_0, \alpha_{0i} - \kappa'_i \beta_0, \gamma_{0t} - \rho'_t \beta_0)$. Solving the program (12) with $\ell_{it}^*(\beta, \alpha_i, \gamma_t)$ in place of $\ell_{it}(\beta, \alpha_i, \gamma_t)$ does not change the solution for β . The reparametrized program cannot be used to compute the estimator because $\ell_{it}^*(\beta, \alpha_i, \gamma_t)$ depends on θ_0 , but it is a convenient theoretical device to analyze the properties of $\widehat{\beta}$ because the Hessian $\mathbb{E}_{NT} \bar{\ell}_{it}^{*\theta\theta}$ is block-diagonal between β and (α, γ) . We proceed by obtaining the bias from the infeasible log-likelihood ℓ_{it}^* , and then we can express it in terms of the feasible likelihood ℓ_{it} using the one-to-one relationship between ℓ_{it} and ℓ_{it}^* . For example, $\ell_{it}^{*\beta} = \ell_{it}^\beta + \kappa_i \ell_{it}^{\alpha_i} + \rho_t \ell_{it}^{\gamma_t}$.

Under the orthogonal parametrization, the component of the influence function $\psi_1 = -(\mathbb{E}_{NT} \bar{\ell}_{it}^{*\theta\theta})^{-1} \mathbb{G}_{NT} \ell_{it}^{*\theta}$ corresponding to β simplifies to $\psi_{1,\beta} = -(\mathbb{E}_{NT} \bar{\ell}_{it}^{*\beta\beta})^{-1} \mathbb{G}_{NT} \ell_{it}^{*\beta}$. Under standard conditions for MLE, as $N, T \rightarrow \infty$,

$$\psi_{1,\beta} \rightarrow_d \mathcal{N}(0, H^{-1}), \quad H = - \operatorname{plim}_{N,T \rightarrow \infty} \mathbb{E}_{NT} \bar{\ell}_{it}^{*\beta\beta},$$

where we use that $\psi_{1,\beta}$ is a martingale difference over t , and the conditional information equality. As in the cross-sectional case, $\psi_{1,\beta}$ has zero mean and determines the asymptotic variance.

The component of the second term ψ_2 corresponding to β also simplifies with the reparametrization.

⁴This transformation corresponds to the reparameterization $\alpha_i^* = \alpha_i - \kappa'_i \beta$ and $\gamma_t^* = \gamma_t - \rho'_t \beta$. The log-likelihood with respect to these parameters is $\ell_{it}(\beta, \alpha_i^* + \kappa'_i \beta, \gamma_t^* + \rho'_t \beta) =: \ell_{it}^*(\beta, \alpha_i^*, \gamma_t^*)$, which gives (13) after renaming (α_i^*, γ_t^*) as (α_i, γ_t) again.

⁵The solution for κ_i and ρ_t may not be unique. For example, if $\ell_{it}(\beta, \alpha_i, \gamma_t) = \ell_{it}(\beta, \alpha_i + \gamma_t)$, then the system does not uniquely determine κ_i and ρ_t , but only $\kappa_i + \rho_t$ for all i, t . However, only $\kappa_i + \rho_t$ will appear in our formulas for asymptotic bias and variance of $\widehat{\beta}$ in that case, so the non-uniqueness of κ_i and ρ_t is not important.

After some calculations that are detailed in the appendix,

$$\psi_{2,\beta} := - \underbrace{\left(\mathbb{E}_{NT} \bar{\ell}_{it}^{*\beta\beta} \right)^{-1} \sum_{i=1}^N \left[-\frac{\mathbb{G}_T \ell_{it}^{*\beta\alpha_i} \mathbb{G}_T \ell_{it}^{\alpha_i}}{\mathbb{E}_T \bar{\ell}_{it}^{\alpha_i\alpha_i}} + \frac{\mathbb{E}_T \bar{\ell}_{it}^{*\beta\alpha_i\alpha_i} (\mathbb{G}_T \ell_{it}^{\alpha_i})^2}{2 \left(\mathbb{E}_T \bar{\ell}_{it}^{\alpha_i\alpha_i} \right)^2} \right]}_{=\psi_{2,\beta}^B} - \underbrace{\left(\mathbb{E}_{NT} \bar{\ell}_{it}^{*\beta\beta} \right)^{-1} \sum_{t=1}^T \left[-\frac{\mathbb{G}_N \ell_{it}^{*\beta\gamma_t} \mathbb{G}_N \ell_{it}^{\gamma_t}}{\mathbb{E}_N \bar{\ell}_{it}^{\gamma_t\gamma_t}} + \frac{\mathbb{E}_N \bar{\ell}_{it}^{*\beta\gamma_t\gamma_t} (\mathbb{G}_N \ell_{it}^{\gamma_t})^2}{2 \left(\mathbb{E}_N \bar{\ell}_{it}^{\gamma_t\gamma_t} \right)^2} \right]}_{=\psi_{2,\beta}^D}.$$

This second-order term is the primary source of bias. By the law of large numbers for heterogenous weakly dependent sequences as $N, T \rightarrow \infty$,

$$\frac{\psi_{2,\beta}^B}{N} \rightarrow_P H^{-1} \text{plim}_{N,T \rightarrow \infty} \mathbb{E}_N \left\{ \frac{-\mathbb{E}_T \sum_{s=t}^T \mathbb{E} \left(\ell_{it}^{\alpha_i} \ell_{is}^{*\beta\alpha_i} \right) - \mathbb{E}_T \bar{\ell}_{it}^{*\beta\alpha_i\alpha_i} / 2}{\mathbb{E}_T \bar{\ell}_{it}^{\alpha_i\alpha_i}} \right\} =: B, \quad (15)$$

and

$$\frac{\psi_{2,\beta}^D}{T} \rightarrow_P H^{-1} \text{plim}_{N,T \rightarrow \infty} \mathbb{E}_T \left\{ \frac{-\mathbb{E}_N \mathbb{E} \left(\ell_{it}^{\gamma_t} \ell_{it}^{*\beta\gamma_t} \right) - \mathbb{E}_N \bar{\ell}_{it}^{*\beta\gamma_t\gamma_t} / 2}{\mathbb{E}_N \bar{\ell}_{it}^{\gamma_t\gamma_t}} \right\} =: D. \quad (16)$$

The expressions for B and D are almost symmetric with respect to the indices i and t , but there is no double sum in D because we are assuming independence of observations across i , while we allow for predetermined covariates over t . The expressions become symmetric when the covariates are strictly exogenous because the terms of the double sum $\sum_{s=t+1}^T$ in B drop out. Hence, the order of the bias is

$$\frac{\psi_{2,\beta}}{NT} = \frac{\psi_{2,\beta}^B + \psi_{2,\beta}^D}{NT} \sim \frac{B}{T} + \frac{D}{N} \sim \frac{N+T}{NT},$$

which corresponds to the prediction from (1) since $n = NT$ and $p = d_\beta + N + T \sim N + T$.

The bias term B/T comes from the individual fixed effects since there are T observations that are informative about each of them. Symmetrically, the bias term D/N comes from the time fixed effects since there are N observations that are informative about each of them. Accordingly, the term D/N drops out in models with only individual effects or becomes negligible relative to B/T when $N \gg T$. Hahn & Kuersteiner (2002) and Alvarez & Arellano (2003) characterized B in dynamic linear panel models with individual effects. For nonlinear models, Hahn & Newey (2004) and Hahn & Kuersteiner (2011) derived B in the static and dynamic case, respectively. Fernández-Val & Weidner (2016) characterized B and D in static and dynamic nonlinear models with individual and time effects.

The expressions of B and D can be further characterized in specific models.

Example (i) In the linear model (8),

$$B = H^{-1} \operatorname{plim}_{N,T \rightarrow \infty} -\mathbb{E}_{NT} \left\{ \sum_{s=t+1}^T \varepsilon_{it} \tilde{x}_{is} \right\}, \quad D = 0, \quad H = \operatorname{plim}_{N,T \rightarrow \infty} \mathbb{E}_{NT} \{ \tilde{x}_{it} \tilde{x}'_{it} \},$$

where \tilde{x}_{it} is the residual of the linear projection of x_{it} on the space spanned by the individual and time effects (i.e. the two-way demeaned x_{it} in the linear model that corresponds to the orthogonal transformation). When the covariates are strictly exogenous, $B = 0$, i.e., there is no incidental parameter problem. Hahn & Kuersteiner (2002) obtained this large T bias expansion of the Nickell (1981) bias for dynamic linear panel models with only individual effects. Hahn & Moon (2006) showed that the bias expression carries over to dynamic linear panel models with individual and time effects because $D = 0$.

Example (ii) In the binary response model (9) with standard normal link $F_\varepsilon = \Phi$ or probit model, when the covariates are strictly exogenous

$$B = H^{-1} \mathbb{E}_{NT} \left\{ \frac{\omega_{it} \tilde{x}_{it} \tilde{x}'_{it}}{\mathbb{E}_T \omega_{it}} \right\} \beta_0, \quad D = H^{-1} \mathbb{E}_{NT} \left\{ \frac{\omega_{it} \tilde{x}_{it} \tilde{x}'_{it}}{\mathbb{E}_N \omega_{it}} \right\} \beta_0, \quad H = \mathbb{E}_{NT} \{ \omega_{it} \tilde{x}_{it} \tilde{x}'_{it} \},$$

where \tilde{x}_{it} is the residual of the linear projection of x_{it} on the space spanned by the individual and time effects under a metric weighted by ω_{it} , $\omega_{it} = \phi_{it}^2 / [\Phi_{it}(1 - \Phi_{it})]$, and ϕ_{it} and Φ_{it} are the standard normal PDF and CDF evaluated at $x'_{it}\beta_0 + \alpha_{0i} + \gamma_{0t}$. In this case there is bias, which is a positive definite matrix weighted average of β_0 . Fernández-Val (2009) derived this result in models with individual effects and Fernández-Val & Weidner (2016) in models with individual and time effects.

Example (iii) In the Poisson model (10),

$$B = H^{-1} \operatorname{plim}_{N,T \rightarrow \infty} -\mathbb{E}_{NT} \left\{ \frac{\sum_{s=t+1}^T (y_{it} - \lambda_{it}) \lambda_{is} \tilde{x}_{is}}{\mathbb{E}_N \lambda_{it}} \right\}, \quad D = 0, \quad H = \operatorname{plim}_{N,T \rightarrow \infty} \mathbb{E}_{NT} \{ \lambda_{it} \tilde{x}_{it} \tilde{x}'_{it} \},$$

where \tilde{x}_{it} is the residual of the linear projection of x_{it} on the space spanned by the individual and time effects under a metric weighted by λ_{it} . When the covariates are strictly exogenous, $B = 0$, i.e., there is no incidental parameter problem. This is a well-known result for models with only individual effects (e.g., Palmgren 1981),⁶ which was extended to models with individual and time effect in Fernández-Val & Weidner (2016).

Finally, combining the properties of $\psi_{1,\beta}$ and $\psi_{2,\beta}$, we conclude that as $N, T \rightarrow \infty$ such that $N/T \rightarrow \kappa$, $0 < \kappa < \infty$,

$$\hat{\beta} - \beta_0 \stackrel{a}{\sim} \mathcal{N} \left(\frac{B}{T} + \frac{D}{N}, \frac{H^{-1}}{NT} \right). \quad (17)$$

⁶Hausman, Hall & Griliches, 1984 use the conditional likelihood approach to eliminate the incidental parameters, which for the Poisson models turns out to be equivalent to fixed effect MLE, see Blundell, Griffith & Windmeijer (1999, 2002) and Lancaster (2002).

This asymptotic approximation prescribes that the fixed effects estimator can have significant bias relative to its standard deviation. Moreover, by the argument given in Section 2.2, confidence intervals constructed around the fixed effects estimator can severely undercover the components of β_0 even in large samples. Fernández-Val & Weidner (2016) showed that these prescriptions provide a good approximations to the behaviour of the fixed effects estimator for small sample sizes through analytical and simulation examples.

3.4 Bias Corrections

The goal of the bias corrections is to remove the bias from the asymptotic distribution (17), ideally without increasing the variance. In other words, we want to find a bias corrected estimator $\widehat{\beta}_{\text{BC}}$ such that as $N, T \rightarrow \infty$,

$$\widehat{\beta}_{\text{BC}} - \beta_0 \overset{a}{\sim} \mathcal{N}\left(0, \frac{H^{-1}}{NT}\right).$$

The corrected estimator therefore has small bias relative to its dispersion and the same variance as the FE estimator in large samples. Moreover, the confidence intervals constructed around the bias corrected estimator should have coverage close to their nominal level. We describe analytical and resampling methods to carry out the bias corrections. These methods rely on the asymptotic distribution (17), together with consistent estimators of the bias.

3.4.1 Analytical Bias Correction

The analytically bias corrected (ABC) estimator is

$$\widehat{\beta}_{\text{ABC}} = \widehat{\beta} - \frac{\widehat{B}}{T} - \frac{\widehat{D}}{N},$$

where \widehat{B} and \widehat{D} are consistent estimators of B and D , i.e. $\widehat{B} \rightarrow_P B$ and $\widehat{D} \rightarrow_P D$ as $N, T \rightarrow \infty$. In this case, as $N, T \rightarrow \infty$ such that $N/T \rightarrow \kappa$, $0 < \kappa < \infty$,

$$\widehat{\beta}_{\text{ABC}} - \beta_0 \approx \frac{\psi_{1,\beta}}{\sqrt{NT}} + \frac{1}{T} \left(\frac{\psi_{2,\beta}^B}{N} - \widehat{B} \right) + \frac{1}{N} \left(\frac{\psi_{2,\beta}^D}{T} - \widehat{D} \right) \approx \frac{\psi_{1,\beta}}{\sqrt{NT}},$$

since $\psi_{2,\beta}^B/N - \widehat{B} \rightarrow_P 0$ and $\psi_{2,\beta}^D/T - \widehat{D} \rightarrow_P 0$. Hence,

$$\widehat{\beta}_{\text{ABC}} - \beta_0 \overset{a}{\sim} \mathcal{N}\left(0, \frac{H^{-1}}{NT}\right).$$

The estimators \widehat{B} and \widehat{D} are constructed from the analytical expressions of B and D given in (15) and (16). Let $\widehat{\ell}_{it} = \ell_{it}(\widehat{\beta}, \widehat{\alpha}_i, \widehat{\gamma}_t)$ denote the log-likelihoods evaluated at the FE estimators, and define their derivatives evaluated at the FE estimators analogously. Let $\widehat{\ell}_{it}^*(\beta, \alpha_i, \gamma_t) := \ell_{it}(\beta, \alpha_i + \widehat{\kappa}_i' \beta, \gamma_t + \widehat{\rho}_t' \beta)$, with $\widehat{\kappa}_i$ and $\widehat{\rho}_t$ defined analogously to κ_i and ρ_t in (14), but using the fixed effect estimates instead of the true

value of the parameters. All derivatives of $\widehat{\ell}_{it}^*$ are evaluated at $(\widehat{\beta}, \widehat{\alpha}_i - \widehat{\kappa}'_i \widehat{\beta}, \widehat{\gamma}_t - \widehat{\rho}'_t \widehat{\beta})$ in the following. Then, the plug-in estimators of B and D are

$$\widehat{B} = \widehat{H}^{-1} \mathbb{E}_{NT} \left\{ \frac{\sum_{s=t}^{t+M \wedge T} \widehat{\ell}_{it}^{\alpha_i} \widehat{\ell}_{is}^{\beta \alpha_i} - \widehat{\ell}_{it}^{\beta \alpha_i \alpha_i} / 2}{\mathbb{E}_T \widehat{\ell}_{it}^{\alpha_i \alpha_i}} \right\},$$

and

$$\widehat{D} = \widehat{H}^{-1} \mathbb{E}_{NT} \left\{ \frac{\widehat{\ell}_{it}^{\gamma_t} \widehat{\ell}_{it}^{\beta \gamma_t} - \widehat{\ell}_{it}^{\beta \gamma_t \gamma_t} / 2}{\mathbb{E}_N \widehat{\ell}_{it}^{\gamma_t \gamma_t}} \right\},$$

where $\widehat{H} = -\mathbb{E}_{NT} \widehat{\ell}_{it}^{\beta \beta}$, and M is a trimming parameter such that $M/T \rightarrow 0$ and $M \rightarrow \infty$ as $T \rightarrow \infty$ (Hahn & Kuersteiner, 2007). We can set $M = 0$ when the covariates are strictly exogenous. These estimators use $\widehat{\beta}$, i.e. $\widehat{B} = \widehat{B}(\widehat{\beta})$ and $\widehat{D} = \widehat{D}(\widehat{\beta})$. It is possible to iterate the correction by (i) estimating B and D using $\widehat{\beta}_{\text{ABC}}$ to obtain $\widehat{B}_2 = \widehat{B}(\widehat{\beta}_{\text{ABC}})$, $\widehat{D}_2 = \widehat{D}(\widehat{\beta}_{\text{ABC}})$ and $\widehat{\beta}_{\text{ABC}2} = \widehat{\beta} - \widehat{B}_2/T - \widehat{D}_2/N$; (ii) estimating B and D using $\widehat{\beta}_{\text{ABC}2}$ to obtain $\widehat{B}_3 = \widehat{B}(\widehat{\beta}_{\text{ABC}2})$, $\widehat{D}_3 = \widehat{D}(\widehat{\beta}_{\text{ABC}2})$ and $\widehat{\beta}_{\text{ABC}3} = \widehat{\beta} - \widehat{B}_3/T - \widehat{D}_3/N$; and so on. The iteration does not affect the asymptotic distribution of the correction but might improve its small sample properties.

Hahn & Kuersteiner (2011) developed the ABC for general dynamic nonlinear models with unobserved individual effects, building on the analysis of Hahn & Kuersteiner (2002) for dynamic linear models and Hahn & Newey (2004) for static nonlinear models. Fernández-Val & Weidner (2016) extended the ABC to models with unobserved individual and time effects.

3.4.2 Leave-One-Out Jackknife Bias Correction

This correction is based on the Jackknife method introduced by Quenouille (1956) and Tukey (1958) for cross-sectional data. We start by giving some intuition on how this method works using the example of Section 2.⁷ Let $\widehat{\theta}_{(-j)}$ denote the estimator of θ that leaves out the j^{th} observation. The leave-one-out bias corrected estimator is

$$\widehat{\theta}_{\text{JC}} = n\widehat{\theta} - (n-1)\bar{\theta}_{n-1}, \quad \bar{\theta}_{n-1} = \frac{1}{n} \sum_{j=1}^n \widehat{\theta}_{(-j)}.$$

To understand how the correction works, assume the second-order expansion for the bias $\mathbb{E}\widehat{\theta} - \theta_0 = B_1/n + B_2/n^2 + o(n^{-2})$. Then, under identical distribution $\mathbb{E}\widehat{\theta}_{(-j)} - \theta_0 = B_1/(n-1) + B_2/(n-1)^2 + o(n^{-2})$ for all j , so that

$$\mathbb{E}\widehat{\theta}_{\text{JC}} - \theta_0 = B_1 + \frac{B_2}{n} - B_1 - \frac{B_2}{n-1} + o(n^{-1}) = o(n^{-1}).$$

In other words $(n-1)(\bar{\theta}_{n-1} - \widehat{\theta})$ is an estimator of the first-order bias since $(n-1)\mathbb{E}(\bar{\theta}_{n-1} - \widehat{\theta}) = B_1/n + o(n^{-1})$.

Hahn & Newey (2004) introduced the Jackknife to panel models with individual effects, and Fernández-Val & Weidner (2016) extended it to panel models with individual and time effects. To describe how to

⁷We refer to Shao & Tu (1995) for a rigorous analysis of the properties of the Jackknife.

apply the Jackknife to panel data, it is convenient to introduce some notation. Let $\mathbf{N} = \{1, \dots, N\}$ and $\mathbf{T} = \{1, \dots, T\}$ be the sets of indexes for the two dimensions of the panel. For the subsets of indexes $A \subseteq \mathbf{N}$ and $C \subseteq \mathbf{T}$, let $\widehat{\beta}_{A,C}$ denote the FE estimator of β in the subpanel with indexes $(i, t) \in A \times C = \{(i, t) : i \in A, t \in C\}$, that is

$$\widehat{\beta}_{A,C} = \operatorname{argmax}_{\beta \in \mathbb{R}^{d_\beta}} \max_{\alpha \in \mathbb{R}^{|A|}} \max_{\gamma \in \mathbb{R}^{|C|}} \sum_{i \in A, t \in C} \ell_{it}(\beta, \alpha_i, \gamma_t),$$

where $|A|$ denotes the cardinality of the set A . With this notation the FE estimator of β is $\widehat{\beta} = \widehat{\beta}_{\mathbf{N},\mathbf{T}}$. Define the average leave-one-out estimators in each dimension as

$$\bar{\beta}_{N-1,T} = \frac{1}{N} \sum_{i=1}^N \widehat{\beta}_{\mathbf{N} \setminus \{i\}, \mathbf{T}}, \quad \bar{\beta}_{N,T-1} = \frac{1}{T} \sum_{t=1}^T \widehat{\beta}_{\mathbf{N}, \mathbf{T} \setminus \{t\}}.$$

The panel Jackknife estimator for models with individual effects is $T\widehat{\beta}_{\mathbf{N},\mathbf{T}} - (T-1)\bar{\beta}_{N,T-1}$. The correction $(T-1)(\bar{\beta}_{N,T-1} - \widehat{\beta}_{\mathbf{N},\mathbf{T}})$ removes the bias term B/T , analogously to the cross-sectional case discussed above. A panel jackknife bias corrected (JBC) estimator that removes both bias terms can be formed as

$$\widehat{\beta}_{\text{JBC}} = (N+T-1)\widehat{\beta}_{\mathbf{N},\mathbf{T}} - (N-1)\bar{\beta}_{N-1,T} - (T-1)\bar{\beta}_{N,T-1},$$

This correction removes the bias in large samples because $(N-1)\mathbb{E}(\bar{\beta}_{N-1,T} - \widehat{\beta}_{\mathbf{N},\mathbf{T}}) \approx D/N$ and $(T-1)\mathbb{E}(\bar{\beta}_{N,T-1} - \widehat{\beta}_{\mathbf{N},\mathbf{T}}) \approx B/T$. Moreover, it can be shown that under suitable conditions

$$\widehat{\beta}_{\text{JBC}} - \beta_0 \stackrel{a}{\sim} \mathcal{N}\left(0, \frac{H^{-1}}{NT}\right).$$

An important limitation in the application of the leave-one-out Jackknife to panel data is that it requires independence in both dimensions of all the variables. This is a very restrictive condition for panel data, specially when one of the dimensions is time, as it rules out lagged-dependent variables or serially correlated variables as covariates.

3.4.3 Split-Sample Jackknife Bias Correction

This correction is based on the split-sample Jackknife method introduced by Quenouille (1949) for time series data. We again start by giving an intuitive description of this method in the context of the simple model of Section 2. Assume that the sample size n is even and split the sample in two halves: $\{z_j : 1 \leq j \leq n/2\}$ and $\{z_j : n/2 + 1 \leq j \leq n\}$. Let $\widehat{\theta}^1$ and $\widehat{\theta}^2$ denote the estimator of θ in each of the half-samples. The split-sample jackknife bias corrected estimator is

$$\widehat{\theta}_{\text{SC}} = 2\widehat{\theta} - \widetilde{\theta}_{1/2} = \widehat{\theta} - (\widetilde{\theta}_{1/2} - \widehat{\theta}), \quad \widetilde{\theta}_{1/2} = (\widehat{\theta}^1 + \widehat{\theta}^2)/2.$$

Assume the first-order expansion for the bias $\mathbb{E}\widehat{\theta} - \theta_0 = B/n + o(n^{-1})$. The correction works asymptotically because under stationarity $\mathbb{E}\widehat{\theta}^1 = \mathbb{E}\widehat{\theta}^2 = \theta_0 + 2B/n + o(n^{-1})$, so that

$$\mathbb{E}\widehat{\theta}_{\text{SC}} - \theta_0 = \frac{2B}{n} - \frac{2B}{n} + o(n^{-1}) = o(n^{-1}).$$

In other words $\tilde{\theta}_{1/2} - \hat{\theta}$ is an estimator of the bias since $E(\tilde{\theta}_{1/2} - \hat{\theta}) = B/n + o(n^{-1})$.

To define the panel version of this Jackknife correction we introduce

$$\tilde{\beta}_{N/2,T} = \frac{1}{2} \left[\hat{\beta}_{\{i \leq \lceil N/2 \rceil\}, \mathbf{T}} + \hat{\beta}_{\{i \geq \lfloor N/2+1 \rfloor\}, \mathbf{T}} \right], \quad \tilde{\beta}_{N,T/2} = \frac{1}{2} \left[\hat{\beta}_{\mathbf{N}, \{t \leq \lceil T/2 \rceil\}} + \hat{\beta}_{\mathbf{N}, \{t \geq \lfloor T/2+1 \rfloor\}} \right],$$

where $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ are the ceiling and floor functions, and we use the same notation as in the previous section. Dhaene & Jochmans (2015b) introduced the split-sample jackknife to panel models with individual effects. In this case the split-sample bias corrected (SBC) estimator is $2\hat{\beta}_{\mathbf{N}, \mathbf{T}} - \tilde{\beta}_{N,T/2}$. The correction $\tilde{\beta}_{N,T/2} - \hat{\beta}_{\mathbf{N}, \mathbf{T}}$ removes the bias term B/T , analogously to the cross-sectional case. Fernández-Val & Weidner (2016) extended the SBC estimator to panel models with individual and time effects, in which case the correct linear combination is

$$\hat{\beta}_{\text{SBC}} = 3\hat{\beta}_{\mathbf{N}, \mathbf{T}} - \tilde{\beta}_{N/2,T} - \tilde{\beta}_{N,T/2}.$$

This correction removes both bias terms in large samples because $E(\tilde{\beta}_{N/2,T} - \hat{\beta}_{\mathbf{N}, \mathbf{T}}) \approx D/N$ and $E(\tilde{\beta}_{N,T/2} - \hat{\beta}_{\mathbf{N}, \mathbf{T}}) \approx B/T$. The intuition here is simple. The estimator $\tilde{\beta}_{N,T/2}$ has double the bias than $\hat{\beta}_{\mathbf{N}, \mathbf{T}}$ coming from the estimation of the individual effects, because there are only half of observations, $T/2$, informative about each of them. However, $\tilde{\beta}_{N,T/2}$ has the same bias as $\hat{\beta}_{\mathbf{N}, \mathbf{T}}$ coming from the estimation of the time effects, because there are the same number of observations, N , informative about each of them. A similar argument shows that the bias of $\tilde{\beta}_{N/2,T}$ is $2D/N + B/T$. Under suitable conditions,

$$\hat{\beta}_{\text{SBC}} - \beta_0 \overset{a}{\sim} \mathcal{N} \left(0, \frac{H^{-1}}{NT} \right).$$

Quenouille (1956) already noted that Jackknife correction can be extended to higher order, and Dhaene & Jochmans (2015b) discussed high-order Jackknife corrections, that is, elimination not only of the leading bias of order T^{-1} , but also of the next order bias term of order T^{-2} , which can again be achieved by an appropriate linear combination of split-panel fixed-effect estimates. A formal discussion of those higher-order bias terms is given in Bun & Kiviet (2003) for dynamic linear models panel models and in Sun & Dhaene (2017) for non-linear models.

3.4.4 Hybrid Jackknife Bias Correction

When the data are independent across one of the dimensions of the panel, it is not clear how to carry out the sample split in this dimension. One possibility is to make multiple splits and average the resulting corrected estimators for each split. Another possibility is a hybrid method that uses leave-one-out along the independent dimension and split-sample along the weakly dependent dimension. For example, if we assume independence over the cross-sectional dimension, then a hybrid jackknife estimator is

$$\hat{\beta}_{\text{HBC}} = (N+1)\hat{\beta}_{\mathbf{N}, \mathbf{T}} - (N-1)\bar{\beta}_{N-1,T} - \tilde{\beta}_{N,T/2}.$$

This correction removes the bias in large samples because $(N-1)\mathbb{E}(\bar{\beta}_{N-1,T} - \hat{\beta}_{\mathbf{N},\mathbf{T}}) \approx D/N$ and $\mathbb{E}(\tilde{\beta}_{N,T/2} - \hat{\beta}_{\mathbf{N},\mathbf{T}}) \approx B/T$. Again, under suitable conditions

$$\hat{\beta}_{\text{HBC}} - \beta_0 \overset{a}{\sim} \mathcal{N}\left(0, \frac{H^{-1}}{NT}\right).$$

We refer to Cruz-González, Fernández-Val & Weidner (2015) for other examples of hybrid corrections.

3.4.5 Comparison

All the corrections produce estimators with the same asymptotic distribution, but they rely on different sampling conditions and might involve the choice of tuning parameters. Compared to the JBC; ABC, SBC and HBC work under more general sampling conditions because they do not require independence of the data along both dimensions. However, SBC and HBC usually increase dispersion in small samples because they estimate the parameters in smaller panels than JBC. Compared to ABC, SBC and HBC work under less general sampling conditions because they rely on homogeneity in both dimensions in order for the bias to be the same in all the subpanels. ABC does not use homogeneity, but requires coding the estimators of the bias, which involves a delicate choice for the trimming parameter M when the covariates are not strictly exogenous. In practice, we recommend to always carry out a sensitivity analysis reporting results from both jackknife and analytical corrections and to try several values of M in the analytical correction starting with $M = 0, 1, 2, 3, 4$.

Some of the differences can be appreciated in a simple cross sectional example where we can characterize the moments of all the estimators. Consider the normal model $z \sim \mathcal{N}(\mu, \sigma^2)$, $\sigma^2 > 0$. The MLE of σ^2 is the sample variance $\hat{\sigma}^2 = n^{-1} \sum_{j=1}^n (z_j - \bar{z})^2$, where $\bar{z} = n^{-1} \sum_{j=1}^n z_j$. As it is well-known, $\hat{\sigma}^2$ is biased with $\text{Bias}(\hat{\sigma}^2) = -\sigma^2/n$ and $\text{Var}(\hat{\sigma}^2) = 2\sigma^4(n-1)/n^2$. The ABC is $\tilde{\sigma}_{\text{ABC}}^2 = (n+1)\hat{\sigma}^2/n$, which has $\text{Bias}(\tilde{\sigma}_{\text{ABC}}^2) = -\sigma^2/n^2$ and $\text{Var}(\tilde{\sigma}_{\text{ABC}}^2) = (n+1)^2\text{Var}(\hat{\sigma}^2)/n^2$. Iterating k times yields $\tilde{\sigma}_{\text{ABC}k}^2 = (\sum_{r=0}^k n^{-r})\hat{\sigma}^2$, which has $\text{Bias}(\tilde{\sigma}_{\text{ABC}k}^2) = -\sigma^2/n^{k+1}$ and $\text{Var}(\tilde{\sigma}_{\text{ABC}k}^2) = (\sum_{r=0}^k n^{-r})^2\text{Var}(\hat{\sigma}^2)$. The JBC can be shown to be $\tilde{\sigma}_{\text{JBC}}^2 = n\hat{\sigma}^2/(n-1)$, the degrees-of-freedom corrected estimator, which has $\text{Bias}(\tilde{\sigma}_{\text{JBC}}^2) = 0$ and $\text{Var}(\tilde{\sigma}_{\text{JBC}}^2) = n^2\text{Var}(\hat{\sigma}^2)/(n-1)^2$. Let \bar{z}_1 and \bar{z}_2 be the sample means of z in the first and second halves of the sample. Simple algebra shows that the SBC is $\tilde{\sigma}_{\text{SBC}}^2 = \hat{\sigma}^2 + \bar{z}^2 - \bar{z}_1\bar{z}_2$, which has $\text{Bias}(\tilde{\sigma}_{\text{SBC}}^2) = 0$ and $\text{Var}(\tilde{\sigma}_{\text{SBC}}^2) = (n+2)\text{Var}(\hat{\sigma}^2)/n$. A comparison of biases and variances reveals that in this case $\text{Bias}(\hat{\sigma}^2) > \text{Bias}(\tilde{\sigma}_{\text{ABC}}^2) > \text{Bias}(\tilde{\sigma}_{\text{ABC}k}^2) > \text{Bias}(\tilde{\sigma}_{\text{JBC}}^2) = \text{Bias}(\tilde{\sigma}_{\text{SBC}}^2)$, while $\text{Var}(\hat{\sigma}^2) < \text{Var}(\tilde{\sigma}_{\text{ABC}}^2) < \text{Var}(\tilde{\sigma}_{\text{ABC}k}^2) < \text{Var}(\tilde{\sigma}_{\text{JBC}}^2) < \text{Var}(\tilde{\sigma}_{\text{SBC}}^2)$.

Table 1 reports the results of a numerical simulation for a logit model with strictly exogenous covariates and unobserved individual and time effects. The design is calibrated to the female labor force participation (LFP) application in Fernández-Val (2009). Thus, we draw 500 panels of size $N = 664$ and $T = 9$, with

$$y_{it} = 1(\beta'x_{it} + \alpha_i + \gamma_t \geq \varepsilon_{it}), \quad i = 1, \dots, N, \quad t = 1, \dots, T,$$

Table 1: Logit Model with Strictly Exogenous Covariates ($N = 664, T = 9$)

	Bias	SD	RMSE	p; 95	Bias	SD	RMSE	p; 95
	FE				ABC			
kids0_2	14.3	9.6	17.3	0.64	-0.2	8.3	8.3	0.96
kids3_5	14.1	14.4	20.1	0.79	-0.2	12.6	12.5	0.96
kids6_17	14.5	34.0	36.9	0.93	0.7	29.8	29.8	0.97
	SBC				HBC			
kids0_2	-0.3	11.7	11.7	0.86	-0.6	11.7	11.6	0.84
kids3_5	-0.0	18.4	18.4	0.84	1.2	19.3	19.3	0.81
kids6_17	-0.7	48.1	48.0	0.80	0.4	43.1	42.9	0.92

Notes: 500 simulations calibrated to the PSID 1980–1988.

Bias, SD and RMSE are in percentage of true value of the parameter.

Calculations in `Stata` with the command `logitfe` (Cruz-González et al., 2015).

where x_{it} includes three fertility variables (the numbers of children aged 0-2, 3-5, and 6-17), the logarithm of the husband’s earnings in 1995 thousands of dollars, and a quadratic function of age in years divided by 10, whose values are taken for the PSID 1980–1988; and ε_{it} are independent draws from the standard logistic distribution.⁸ The parameters $(\beta, \alpha_i, \gamma_t)$ are calibrated to the FE logit estimates in the PSID 1980–1988 with the observed LFP as the dependent variable. The table reports biases, standard deviations, root mean square errors, and empirical coverage probabilities of confidence intervals with nominal level of 95% for the FE, ABC, SBC and HBC estimators of the coefficients of the three fertility variables. The SBC is the average over 50 partitions over the cross sectional dimension. All the results, except for the coverage probabilities, are in percentage of the true value of the parameter. We find that the ABC drastically reduce bias, dispersion and rmse, and have coverage probabilities close to their nominal level. The evidence for the SBC and HBC is more mixed. While always reducing bias, they increase dispersion resulting in higher rmse and lower coverage than the FE for the kids6_17 fertility variable.

3.4.6 Other methods

The methods discussed so far correct the estimator. The same methods can also be applied to correct the first order conditions of the FE estimator of β . Namely, let $L(\beta) = \max_{(\alpha, \gamma) \in \mathbb{R}^{N+T}} \mathbb{E}_{NT} \ell_{it}(\beta, \alpha_i, \gamma_t)$ be the profile objective function of (12), where all the fixed effects have been concentrated out. Then, a

⁸The original PSID sample includes 1,461 women, but only 664 of them have variation in the LFP variable over the years of the panel.

second-order expansion similar to (17) yields

$$\frac{dL(\beta_0)}{d\beta} \stackrel{a}{\sim} \mathcal{N}\left(-\frac{b}{T} - \frac{d}{N}, \frac{H}{NT}\right). \quad (18)$$

where $b = HB$ and $d = HD$. This expression was derived in Bester & Hansen (2009) for models with individual effects and Sun (2016a) for models with individual and time effects. Thus, the score of $L(\beta)$ at the true parameter value is not centered at zero, which is the source of the bias in $\hat{\beta}$. A ‘‘profile-score bias corrected estimator’’ is the solution to

$$dL(\hat{\beta}_{\text{PSBC}})/d\beta + \hat{b}/T + \hat{d}/N = 0,$$

where $\hat{b} = \hat{b}(\hat{\beta})$ and $\hat{d} = \hat{d}(\hat{\beta})$ are consistent estimators for b and d , respectively. An alternative score correction that estimates simultaneously the parameter and bias terms can be formed as the solution to

$$dL(\hat{\beta}_{\text{PSBC2}})/d\beta + \hat{b}(\hat{\beta}_{\text{PSBC2}})/T + \hat{d}(\hat{\beta}_{\text{PSBC2}})/N = 0.$$

The analytical properties of this corrected profile-score are model specific, since they depend on the functional form of $\hat{b}(\beta)$ and $\hat{d}(\beta)$. In particular, they may have multiple or no solution, even in the case of concave log-likelihood, where the solution for the previous equation for $\hat{\beta}_{\text{PSBC}}$ is unique. Dhaene & Jochmans (2016) discussed this issue in detail for linear autoregressive panel models, and Dhaene & Jochmans (2015a) explored the behaviour of these profile-score adjustments for other panel data models with individual effects. Dhaene & Jochmans (2015b) developed jackknife methods to correct the scores in models with individual effects.

Finally, Gonalves & Kaffo (2015) show that bootstrap methods can correct for the Nickell bias in dynamic linear panel models, and Kim & Sun (2016) show how to use the bootstrap to construct bias-corrected estimators in non-linear panel models.

3.5 Average Partial Effects

The objects of interest in panel models are often *ceteris paribus* or partial effects, i.e. effects in the outcome of changing each covariate while holding the rest of covariates and unobserved effects fixed (Chamberlain, 1984). In linear models the parameters correspond to partial effects. For example, in the model (8), the components of β measure the partial effects of each of the covariates. This effect is the same for all the individuals and time periods by linearity. In nonlinear models the partial effects depend not only on the parameters but also on the covariates and unobserved effects. For example in the nonlinear model for the conditional expectation $E[y_{it} \mid x_{it}, \alpha_i, \gamma_t] = m(x_{it}, \alpha_i, \gamma_t, \beta)$, the partial effect of changing the covariates from x_{it}^0 to x_{it}^1 is

$$\delta_{it}(\alpha_i, \gamma_t, \beta) = m(x_{it}^1, \alpha_i, \gamma_t, \beta) - m(x_{it}^0, \alpha_i, \gamma_t, \beta),$$

where x_{it}^1 and x_{it}^0 might depend on x_{it} . If x_{it} is continuous and $x \mapsto m(x, \alpha_i, \gamma_t, \beta)$ a.s. differentiable, the partial effect of a marginal change in the covariates is

$$\delta_{it}(\alpha_i, \gamma_t, \beta) = \partial m(x, \alpha_i, \gamma_t) / \partial x \Big|_{x=x_{it}}.$$

In both cases the partial effects are heterogenous across i and t .

Example (ii) In the binary response model, if the k th covariate $x_{k,it}$ is binary, then the partial effect of changing $x_{k,it}$ from 0 to 1 on the probability of $y_{it} = 1$ conditional on the rest of the covariates and unobserved effects is

$$\delta_{it}(\alpha_i, \gamma_t, \beta) = F_\varepsilon(\beta_k + x'_{it,-k}\beta_{-k} + \alpha_i + \gamma_t) - F_\varepsilon(x'_{it,-k}\beta_{-k} + \alpha_i + \gamma_t),$$

where β_k is the k th element of β , and $x_{it,-k}$ and β_{-k} include all elements of x_{it} and β except for the k th element.

Example (iii) In the count response example, if the k th covariate $x_{k,it}$ is continuous, then the partial effect of a marginal change in $x_{k,it}$ on the average of y_{it} conditional on the rest of the covariates and unobserved effects is

$$\delta_{it}(\alpha_i, \gamma_t, \beta) = \beta_k \exp(x'_{it}\beta + \alpha_i + \gamma_t),$$

where β_k is the k th element of β .

One way to summarize the heterogeneity in the partial effects is to average them across the individuals and time periods in the panel. This yields the in-sample average partial effect (APE)

$$\delta_{NT} = \mathbb{E}_{NT} \delta_{it}(\alpha_i, \gamma_t, \beta).$$

Another possibility is to assume that there is a population of individuals and time periods from where the observed panel is drawn and consider the APE in this population. Under weak conditions this in-population APE corresponds to

$$\delta = \text{plim}_{N,T \rightarrow \infty} \delta_{NT}.$$

Yet another possibility is to average the partial effects of the individuals in the population over the time periods in the sample

$$\delta_T = \text{plim}_{N \rightarrow \infty} \delta_{NT}.$$

The choice of the relevant APE is application-specific.

The fixed effect estimator is the same for all the previous APEs. Thus, applying the plug-in principle,

$$\widehat{\delta}_{NT} = \mathbb{E}_{NT} \delta_{it}(\widehat{\alpha}_i, \widehat{\gamma}_t, \widehat{\beta}).$$

However, the asymptotic distribution of $\widehat{\delta}_{NT}$ depends on the APE of interest. The asymptotic distribution around δ_{NT} is

$$\widehat{\delta}_{NT} - \delta_{NT} \overset{a}{\sim} \mathcal{N}\left(\frac{E}{T} + \frac{F}{N}, \frac{\Sigma}{NT}\right),$$

where the expressions for the bias terms E , F , and the variance Σ are given in Fernández-Val & Weidner (2016). The bias and variance here come from the estimation of the parameters and unobserved effects. The distribution around δ becomes

$$\widehat{\delta}_{NT} - \delta \overset{a}{\sim} \mathcal{N}\left(\frac{E}{T} + \frac{F}{N}, \frac{\Sigma + a_{NT}\Omega}{NT}\right), \quad (19)$$

where the additional variance term, $\Omega = \text{plim}_{N,T \rightarrow \infty} (\mathbb{E}_{NT}[\delta_{it}(\alpha_i, \gamma_t, \beta) - \delta])^2 / a_{NT}$, comes from the estimation of the population mean δ using the sample mean δ_{NT} , and $a_{NT} = N \vee T$ because of the correlation of $\delta_{it}(\alpha_i, \gamma_t, \beta)$ with $\delta_{jt}(\alpha_j, \gamma_t, \beta)$ and $\delta_{is}(\alpha_i, \gamma_s, \beta)$ induced by the individual and time effects. Similarly, the distribution around δ_T is

$$\widehat{\delta}_{NT} - \delta_T \overset{a}{\sim} \mathcal{N}\left(\frac{E}{T} + \frac{F}{N}, \frac{\Sigma + T\Omega_T}{NT}\right),$$

where $\Omega_T = \text{plim}_{N \rightarrow \infty} (\mathbb{E}_{NT}[\delta_{it}(\alpha_i, \gamma_t, \beta) - \delta_T])^2 / T$. Analogously to the parameters, the bias in the asymptotic distribution of the estimators of the APEs can be removed using analytical and jackknife corrections (Fernández-Val & Weidner, 2016).

The rate of convergence of $\widehat{\delta}_{NT}$ differs depending on the APE of interest. Thus, δ_{NT} can be estimated at the rate \sqrt{NT} , the same rate as the parameter β , whereas δ can be estimated at the rate $\sqrt{NT/a_{NT}} = \sqrt{N} \wedge \sqrt{T}$ and δ_T at the rate \sqrt{T} . As a consequence of the slower rate of convergence, the asymptotic distribution (19) simplifies to

$$\widehat{\delta}_{NT} - \delta \overset{a}{\sim} \mathcal{N}\left(0, \frac{a_{NT}\Omega}{NT}\right), \quad (20)$$

after dropping terms of lower order.⁹ In other words the estimation of the parameters does not have first-order effect on the distribution around δ . The reason is that the bias and standard deviation introduced by the parameter estimation are of lower order compared to $\sqrt{NT/a_{NT}}$. Despite this possible asymptotic simplification, we recommend using the higher-order distribution in (19) to perform inference on δ , because it provides a more accurate approximation to the finite-sample distribution of $\widehat{\delta}_{NT}$ than the first-order distribution (20).

3.6 Computation

The FE program (12) might look like computationally challenging due to the high dimension of the fixed effects. However, there are two aspects of the program that greatly facilitate the computation. First, the objective function is concave and smooth in the most commonly used cases such as the linear, Poisson, logit,

⁹Similarly the asymptotic distribution around δ_T simplifies to $\widehat{\delta}_{NT} - \delta_T \overset{a}{\sim} \mathcal{N}(0, \Omega_T/N)$.

probit, ordered probit, and tobit models. Second, the design matrix is sparse allowing the use of sparse algebra methods to speed up the computation. The SBC correction preserves the computational properties of the FE estimator as it only involves solving the program (12) in a small number of subsamples. The JBC and HBC can be computationally more intensive than the SBC when the cross-sectional dimension is large. The ABC also preserves the computational properties of the FE estimator, but requires to code the estimators of the bias that are model specific.

The bias corrections are available in the statistical software packages **Stata** and **R** for some models. Cruz-González, Fernández-Val & Weidner (2015) provided two **Stata** commands that implement the analytical and several jackknife corrections for estimators of parameters and APEs in logit and probit models with individual and time effects. These commands also report standard errors for the estimates of the parameter and APEs constructed using the asymptotic distribution. Sun (2016b) implemented in **Stata** the split-sample jackknife estimator and score corrections for parameters in linear, logit and probit models with individual effects. The command also has functionality for other single index user-supplied models and reports standard errors based on the asymptotic distribution. All the previous **Stata** commands provide functionality for unbalanced panels. Stammann, Heiß & McFadden (2016) implemented the analytical correction in **R** for estimators of parameters and APEs in logit and probit models with exogenous covariates and individual effects.

4 EXTENSIONS

4.1 Unbalanced Panels

In the previous section, we assumed that the observations for all the combinations of the two indexes i and t were observed, i.e. the panel was balanced. In empirical applications, however, it is common to have unbalanced panels where some of the observations are missing due to sample attrition. This does not introduce special theoretical complications provided that the source of the missing observations is random. Perhaps due to this reason, we are not aware of any work in the panel literature that extends the bias corrections explicitly to unbalanced panels. To fill this void, we provide without proof the asymptotic distribution of the FE for unbalanced panels.

Let \mathcal{D} be the set of all observed pairs (i, t) , and $n = |\mathcal{D}|$ be the sample size. Define the operators $\mathbb{E}_n := n^{-1} \sum_{(i,t) \in \mathcal{D}}$, $\mathbb{E}_{T,i} = |\mathcal{D}_i|^{-1} \sum_{t \in \mathcal{D}_i}$, and $\mathbb{E}_{N,t} = |\mathcal{D}_t|^{-1} \sum_{i \in \mathcal{D}_t}$, where $\mathcal{D}_i = \{t : (i, t) \in \mathcal{D}\}$ and $\mathcal{D}_t = \{i : (i, t) \in \mathcal{D}\}$. Let a_{it} be an attrition indicator for the observation (i, t) . The key regularity conditions that are needed to derive the bias expressions are (i) $|\mathcal{D}_i|/T \rightarrow c_i > 0$ as $T \rightarrow \infty$ for all i ; (ii) $|\mathcal{D}_t|/N \rightarrow c_t > 0$ as $N \rightarrow \infty$ for all t ; and (iii) y_{it} is independent of a_{it} conditional on (x_i^t, α, γ) . The first two conditions guarantee that the number of observations for each unobserved effects increases with the sample size. The third condition imposes conditional missing at random in the attrition process. The

asymptotic distribution of $\widehat{\beta}$ then becomes

$$\widehat{\beta} - \beta_0 \stackrel{a}{\sim} \mathcal{N} \left(\frac{B_u}{\bar{T}} + \frac{D_u}{\bar{N}}, \frac{H_u^{-1}}{n} \right),$$

where $H_u = -\text{plim}_{n \rightarrow \infty} \mathbb{E}_n \bar{\ell}_{it}^{*\beta\beta}$, $\bar{T} = n/N$ is the average number of observations available for each cross-sectional unit,

$$B_u = H_u^{-1} \text{plim}_{N,T \rightarrow \infty} \mathbb{E}_N \left\{ \frac{-\mathbb{E}_{T,i} \sum_{t \leq s \in \mathcal{D}_i} \mathbb{E} \left(\ell_{it}^{\alpha_i} \ell_{is}^{*\beta\alpha_i} \right) - \mathbb{E}_{T,i} \bar{\ell}_{it}^{*\beta\alpha_i\alpha_i} / 2}{\mathbb{E}_{T,i} \bar{\ell}_{it}^{\alpha_i\alpha_i}} \right\},$$

$\bar{N} = n/T$ is the average number of observations available for each time period, and

$$D_u = H_u^{-1} \text{plim}_{N,T \rightarrow \infty} \mathbb{E}_T \left\{ \frac{-\mathbb{E}_{N,t} \mathbb{E} \left(\ell_{it}^{\gamma_t} \ell_{it}^{*\beta\gamma_t} \right) - \mathbb{E}_{N,t} \bar{\ell}_{it}^{*\beta\gamma_t\gamma_t} / 2}{\mathbb{E}_{N,t} \bar{\ell}_{it}^{\gamma_t\gamma_t}} \right\}.$$

This result agrees with formula (1), because the order of the bias is $\bar{T}^{-1}B + \bar{N}^{-1}D \sim (N + T)/n$. The magnitude of the bias depends on the averages \bar{T} and \bar{N} , so having some units i and t with a very small number of observations is not necessarily a serious problem from the perspective of incidental parameter bias.

Regarding bias correction, the ABC for the case of unbalanced panels can be constructed using the empirical analog of the bias expressions evaluated at the FE estimator. Jackknife corrections can be formed by partitioning the panel in the same way as in the balanced case, i.e. without taking into account the attrition. We refer to Dhaene & Jochmans (2015b), Cruz-González, Fernández-Val & Weidner (2015), Chudik, Pesaran & Yang (2016) and Sun (2016b) for details.

4.2 Multivariate Fixed Effects

We now briefly discuss the case where the unobserved effects are vector-valued, i.e. $\alpha_i \in \mathbb{R}^{d_\alpha}$ and $\gamma_t \in \mathbb{R}^{d_\gamma}$ with $d_\alpha \geq 1$ and $d_\gamma \geq 1$. We start by providing some motivating examples.

Example (iv) An important case is the normal linear model with interactive effects or factor structure, $y_{it} = x'_{it}\delta + \alpha'_i\gamma_t + \varepsilon_{it}$, $\varepsilon_{it} \mid x_{it}^t, \alpha, \gamma, \beta \sim \mathcal{N}(0, \sigma^2)$, where

$$f(y \mid x_{it}, \alpha_i, \gamma_t; \beta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(y - x'_{it}\delta - \alpha'_i\gamma_t)^2}{2\sigma^2} \right], \quad \beta = (\delta, \sigma^2),$$

Here, $d_\alpha = d_\gamma$ is the number of interactive effects. Bai (2009) showed that this model does not suffer from the incidental parameter problem when the covariates are strictly exogenous and the semiparametric model is correctly specified. However, he found that the FE least squares estimator has an asymptotic bias structure of the form $B/T + D/N$ under conditional heteroskedasticity and derived the expressions for B

and D that can be used for analytic bias correction. Moon & Weidner (2017) showed that the FE least squares estimator of δ suffers from a Nickell (1981) type incidental parameter bias when the covariates are predetermined, even when the model is correctly specified. We refer to Bai & Wang (2016) for a recent review on this model.

Example (v) Another example is a nonlinear single index factor model where

$$f(y \mid x_{it}, \alpha_i, \gamma_t; \beta) = f(y, x'_{it}\beta + \alpha'_i\gamma_t), \quad d_\alpha = d_\gamma,$$

where f is a known function such as $f(y, u) = F_\varepsilon(u)^y(1 - F_\varepsilon(u))^{1-y}1(y \in \{0, 1\})$ with a CDF F_ε , for a binary response model, or $f(y, u) = [\exp(u)^y \exp(-\exp u)/y!]1(y \in \{0, 1, \dots\})$ for a count response model. Chen, Fernandez-Val & Weidner (2014) characterized the bias of the FE estimator for models where the function $u \mapsto f(\cdot, u)$ is log-concave including the probit, logit, ordered probit and Poisson. See also Chen (2016), Boneva & Linton (2017) and Wang (2017) for other articles discussing this model.

The analysis of Section 3 carries over to the multivariate case with some minor adjustments. For the second-order expansion of $\widehat{\beta}$ it is still convenient to define $\ell_{it}^*(\beta, \alpha_i, \gamma_t)$ as in (13), but now κ_i is a $d_\beta \times d_\alpha$ matrix and ρ_t is a $d_\beta \times d_\gamma$ matrix, which are solutions to

$$\begin{aligned} \mathbb{E}_T \left[\bar{\ell}_{it}^{\beta\alpha'_i} + \kappa_i \bar{\ell}_{it}^{\alpha_i\alpha'_i} + \rho_t \bar{\ell}_{it}^{\alpha_i\gamma'_t} \right] &= 0, \quad i = 1, \dots, N, \\ \mathbb{E}_N \left[\bar{\ell}_{it}^{\beta\gamma'_t} + \kappa_i \bar{\ell}_{it}^{\alpha_i\gamma'_t} + \rho_t \bar{\ell}_{it}^{\gamma_t\gamma'_t} \right] &= 0, \quad t = 1, \dots, T. \end{aligned}$$

Here we need to be more careful with the column and row dimensions of matrices in the notation. For example, $\bar{\ell}_{it}^{\beta\alpha'_i} := \partial(\partial\bar{\ell}_{it}/\partial\beta)/\partial\alpha'_i$ is the $d_\beta \times d_\alpha$ matrix of second derivatives of $\bar{\ell}_{it}$. A second-order asymptotic expansion similar to Section 3.3 gives

$$\widehat{\beta} - \beta_0 \stackrel{a}{\sim} \mathcal{N} \left(\frac{B}{T} + \frac{D}{N}, \frac{H^{-1}\Omega H^{-1}}{NT} \right),$$

where $H = -\text{plim}_{N,T \rightarrow \infty} \mathbb{E}_{NT} \bar{\ell}^{*\beta\beta'}$, $\Omega = \text{plim}_{N,T \rightarrow \infty} \mathbb{E}_{NT} \mathbb{E} \left(\bar{\ell}^{*\beta} \bar{\ell}^{*\beta'} \right)$, $B = H^{-1}b$ and $D = H^{-1}d$, and b and d are d_β -vectors with elements

$$\begin{aligned} b_k := \mathbb{E}_N \left\{ -\text{Tr} \left[\left(\mathbb{E}_T \bar{\ell}_{it}^{\alpha_i\alpha'_i} \right)^{-1} \mathbb{E}_T \sum_{s=t}^T \mathbb{E} \left(\ell_{it}^{\alpha_i} \ell_{is}^{*\beta_k\alpha'_i} \right) \right] \right. \\ \left. + \frac{1}{2} \text{Tr} \left[\left(\mathbb{E}_T \bar{\ell}_{it}^{\alpha_i\alpha'_i} \right)^{-1} \left(\mathbb{E}_T \bar{\ell}_{it}^{*\beta_k\alpha_i\alpha'_i} \right) \left(\mathbb{E}_T \bar{\ell}_{it}^{\alpha_i\alpha'_i} \right)^{-1} \mathbb{E}_T \mathbb{E} \left(\ell_{it}^{\alpha_i} \ell_{it}^{\alpha'_i} \right) \right] \right\}, \end{aligned}$$

and

$$\begin{aligned} d_k := \mathbb{E}_T \left\{ -\text{Tr} \left[\left(\mathbb{E}_N \bar{\ell}_{it}^{\gamma_t\gamma'_t} \right)^{-1} \mathbb{E}_N \mathbb{E} \left(\ell_{it}^{\gamma_t} \ell_{it}^{*\beta_k\gamma'_t} \right) \right] \right. \\ \left. + \frac{1}{2} \text{Tr} \left[\left(\mathbb{E}_N \bar{\ell}_{it}^{\gamma_t\gamma'_t} \right)^{-1} \left(\mathbb{E}_N \bar{\ell}_{it}^{*\beta_k\gamma_t\gamma'_t} \right) \left(\mathbb{E}_N \bar{\ell}_{it}^{\gamma_t\gamma'_t} \right)^{-1} \mathbb{E}_N \mathbb{E} \left(\ell_{it}^{\gamma_t} \ell_{it}^{\gamma'_t} \right) \right] \right\}. \end{aligned}$$

In contrast to Section 3.3, the expressions of the asymptotic bias and variance here do not use the information equality to simplify terms, e.g., $H^{-1}\Omega H^{-1} = H^{-1}$. These formulas remain valid in conditional moment models where only a characteristic of the conditional distribution such as the expectation is correctly specified. This covers for example the linear model with interactive effects under heteroskedasticity, that is, a version of Example (iv) that only imposes

$$\mathbb{E}[y_{it} \mid x_{it}, \alpha_i, \gamma_t; \beta] = x'_{it}\beta + \alpha'_i\gamma_t.$$

The above formulas for asymptotic bias and variance thus contain those in Bai (2009) and Moon & Weidner (2017) as special cases.¹⁰ However, these results rely on an auxiliary consistency proof, which is usually model specific and therefore can be delicate. For example, it is difficult to show consistency in models with interactive effects because the log-likelihood $(\beta, \alpha, \gamma) \mapsto \mathbb{E}_{NT}\ell_{it}(\beta, \alpha_i, \gamma_t)$ is not concave.

Analytical and jackknife corrections can be formed analogously to Section 3.4. For details we refer to Arellano & Hahn (2016) for models with individual effects, Bai (2009) and Moon & Weidner (2017) for linear models with interactive fixed effects, and to Chen, Fernandez-Val & Weidner (2014) for single-index factor models. Fernández-Val & Vella (2011) developed bias corrections for two-step estimators of selection and other control variable models with unobserved individual effects in both steps.

4.3 Distributional and Quantile Effects

Chernozhukov, Fernández-Val & Weidner (2017b) used nonlinear panel data methods to estimate distributional and quantile effects. The idea is to model the distribution of the outcome conditional on the covariates and unobserved effects via distribution regression, that is

$$\Pr(y_{it} \leq y \mid x_{it}, \alpha_i, \gamma_t) = F_y(x'_{it}\delta(y) + \alpha'_i\pi(y) + \gamma'_t\xi(y)), \quad \beta(y) = (\delta(y), \pi(y), \xi(y)),$$

where F_y is a CDF such as the normal or logistic, $y \mapsto \beta(y)$ is a function-valued parameter, and the dimension of α_i and γ_t is unrestricted. Both F_y and $\beta(y)$ can vary with y to accommodate heterogeneity across the distribution. At each y , the parameters are estimated by a binary response regression of the outcome $1(y_{it} \leq y)$ on the covariates and fixed effects with the parametrization $\alpha_i(y) = \alpha'_i\pi(y)$ and $\gamma_t(y) = \gamma'_t\xi(y)$. The distributional and quantile effects are functionals of APEs. For example, the marginal distribution of the potential outcome corresponding to $x_{it} = x_{it}^0$ is the APE

$$F^0(y) = (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T F_y(\delta(y)'x_{it}^0 + \alpha_i(y) + \gamma_t(y)),$$

and the corresponding τ -quantile is the functional

$$Q^0(\tau) = \inf\{y \in \mathbb{R} : F^0(y) \geq \tau\}.$$

¹⁰See Sun (2016a) for a related derivation.

The τ -quantile effect of a change from $x_{it} = x_{it}^0$ to $x_{it} = x_{it}^1$ is

$$Q^1(\tau) - Q^0(\tau), \tag{21}$$

where $Q^1(\tau)$ is constructed using the same procedure as $Q^0(\tau)$ but with x_{it}^1 in place of x_{it}^0 . Chernozhukov, Fernández-Val & Weidner (2017b) applied bias corrections to estimate and make inference on quantile effects uniformly over quantile indexes. These methods applied to continuous, discrete and mixed continuous-discrete outcomes.

The effect in (21) is a marginal quantile effect that has causal interpretation under standard conditional ignorability conditions. Alternative conditional effects can be estimated by modelling the distribution of the outcome conditional on the covariates and unobserved effects via quantile regression. The FE quantile regression estimator, introduced by Koenker (2004), in general suffers from the incidental parameter problem. Kato, Galvao & Montes-Rojas (2012) showed asymptotic normality and unbiasedness of the FE quantile regression estimator in models with only individual effects under sequences where $N/T \rightarrow 0$ as $N, T \rightarrow \infty$. Galvao & Kato (2016) derived the asymptotic distribution of the estimator in the same models, including the leading order $1/T$ bias, for a smoothed version of the quantile regression objective function under sequences for which N and T grow at the same rate. Arellano & Weidner (2017) showed that the bias of the FE estimator without smoothing is the same as in Galvao & Kato (2016) in models with strictly exogenous regressors. Thus, in terms of leading order bias and variance the FE quantile regression estimator behaves analogous to any other non-linear panel data model. However, the analysis in Section 3 does not directly carry over to this case, because the quantile regression objective function is not sufficiently smooth. We refer to Galvao & Kato (2016) for a recent review of quantile regression methods for panel data.

4.4 Three Way Fixed Effects

Another natural extension of the panel semiparametric model of Section 3 is a model with three-way effects for data with a multidimensional structure. Let $\ell_{ijt}(\beta, \alpha_i, \gamma_j, \delta_t) = \log f(y_{ijt} | x_{ijt}, \alpha_i, \gamma_j, \delta_t; \beta)$ be the log-likelihood of an outcome y_{ijt} conditional on x_{ijt} , with common parameter β and unobserved effects α_i , γ_j and δ_t . Here, we have three panel indices $i = 1, \dots, I$, $j = 1, \dots, J$ and $t = 1, \dots, T$. For example in an international trade application, Helpman, Melitz & Rubinstein (2008) are interested in analyzing the volume of trade from country i to country j at year t . The corresponding fixed effect estimator is

$$(\hat{\beta}, \hat{\alpha}, \hat{\gamma}, \hat{\delta}) \in \underset{(\beta, \alpha, \gamma, \delta) \in \mathbb{R}^{d_\beta + I + J + T}}{\operatorname{argmax}} \sum_{i=1}^I \sum_{j=1}^J \sum_{t=1}^T \ell_{ijt}(\beta, \alpha_i, \gamma_j, \delta_t).$$

Example (vi) As an extension of Example (ii) above, consider a binary response single index model

$$y_{ijt} = 1(x'_{ijt}\beta + \alpha_i + \gamma_j + \delta_t \geq \varepsilon_{it}), \quad \varepsilon_{it} | x_{it}^t, \alpha, \gamma, \delta \sim F_\varepsilon,$$

where F_ε is a known CDF. For example, y_{ijt} can be an indicator of trade from country i to country j at year t , x_{ijt} the level of tariffs from country i to country j at year t , α_i is an exporter country effect, γ_j is an importer country effect, and δ_t is a year effect.

Similar to the problem of unbalanced panels above, we are not aware of any formal discussion of three-way fixed effect models in the econometrics literature on large panels, but such estimation problems certainly occur in applications. We provide a brief discussion here. In particular, we highlight the applicability of the formula (1) to this case. According to that formula, the expected order of the incidental parameter bias in $\widehat{\beta}$ is $(I + J + T)/(IJT)$, that is, we expect a leading order bias structure $B_1/(IJ) + B_2/(IT) + B_3/(JT)$. Compared to the order of the standard deviation of $\widehat{\beta}$, $1/\sqrt{IJT}$, the bias is small, as long as all three panel dimensions are sufficiently large, in which case the incidental parameter problem can be ignored. Thus, if $I, J, T \rightarrow \infty$ such that $I/JT \rightarrow 0$, $J/IT \rightarrow 0$ and $T/IJ \rightarrow 0$, then $\widehat{\beta}$ is asymptotically normal and unbiased, and standard MLE inference results are applicable to $\widehat{\beta}$, without requiring any bias correction.

So far we have discussed a relatively benign three-way fixed effect model. A more difficult case is $\ell_{ijt}(\beta, \alpha_{ij}, \gamma_{it}, \delta_{jt}) = \log f(y_{ijt} | x_{ijt}, \alpha_{ij}, \gamma_{it}, \delta_{jt}; \beta)$ and

$$(\widehat{\beta}, \widehat{\alpha}, \widehat{\gamma}, \widehat{\delta}) \in \underset{(\beta, \alpha, \gamma, \delta) \in \mathbb{R}^{d_\beta + IJ + IT + JT}}{\operatorname{argmax}} \sum_{i=1}^I \sum_{j=1}^J \sum_{t=1}^T \ell_{ijt}(\beta, \alpha_{ij}, \gamma_{it}, \delta_{jt}).$$

Here, the unobserved effects α_{ij} , γ_{it} and δ_{jt} are pair-specific, only constant along one of the panel dimensions. It is still possible to estimate them consistently if $I, J, T \rightarrow \infty$. Applying the formula (1), we expect the leading order incidental parameter bias to be of the form $B_1/I + B_2/J + B_3/T$ since $p = IJ + IT + JT$ and $n = IJT$. The order of this bias is never negligible compared to the order of the standard deviation of $\widehat{\beta}$, $1/\sqrt{IJT}$. For example, if $I = J = T$, then the order of the bias is I^{-1} , greater than the order of the standard deviation $I^{-3/2}$. It is therefore particularly useful to develop bias corrections for models with fixed effects in multiple dimensions.

5 CONCLUSION

We have reviewed developments for large panel data models since Arellano & Hahn (2007). A key insight is that the order of the incidental parameter bias of the fixed effects estimators can be obtained from the simple formula (1). This formula is useful to get an initial idea of the relevance of the bias in a given application, but the exact magnitude of the bias is still very much model and data generating process dependent. In particular, it is impossible to give any general answer to the question of how large T needs to be for the large- T methods to perform well, see for example Greene (2004) for simulation results on various non-linear panel data models. We have discussed bias correction methods that estimate the actual size of the leading order bias. If this leading order bias is small (i.e. if the bias corrected estimator is close

to the uncorrected estimator), then it is plausible to assume that the higher order bias terms are also small and can be neglected. If the leading order bias is large (i.e. if the bias corrected estimator is very different from the uncorrected estimator), then it is useful to conduct a Monte Carlo simulation calibrated to the application of interest, in order to verify that the remaining bias is small and that the inference procedure works well.

Fixed effect estimation methods for long panels have recently been applied to other data structures such as network and trade data. Examples include Harrigan (1996), Anderson & van Wincoop (2003), Santos Silva & Tenreyro (2006) and Helpman, Melitz & Rubinstein (2008), which estimate gravity equations with unobserved importer and exporter country effects; and Graham (2017), Dzemski (2017), Shi & Chen (2016) and Candelaria (2016), which apply large- T panel methods to network data with unobserved sender and receiver effects. The large- T panel methods are particularly well-suited for this type of data because typically $N = T$. Thus, trade/network usually correspond to square panel data sets where the two dimensions index the same set of countries/individuals as senders and receivers.

There are alternative semiparametric methods to FE that are not reviewed here. An interesting recent approach is the grouped fixed effects (GFE) of Hahn & Moon (2010) and Bonhomme & Manresa (2015) for models with individual effects. Compared to FE, GFE is less affected by the incidental parameter problem, because it restricts the distribution of the individual effects to be discrete. Large T is still required to consistently estimate the group membership of individuals, but if the true heterogeneity structure is indeed discrete, then no asymptotic bias correction is required for the parameters of interest, even if N is much larger than T . The GFE approach imposes restrictions on the distribution of the individual effects, but is more flexible in other dimensions than FE. For example, it permits to include individual effects that change over time. Bonhomme, Lamadon, Manresa et al. (2017) characterized the approximation properties of GFE when the distribution of the individual effects is continuous. They showed that GFE is an effective dimension reduction device in this case, but suffers from the incidental parameter problem.

APPENDIX: BIAS DERIVATION FOR PANEL MODELS

We want to provide a heuristic derivation of the bias formulas (15) and (16). Compared to section 2.1, one needs to define $\theta = (\beta, \alpha, \gamma)$, replace the observation index j by the double index it , and account for nonrandom sampling across i and t . The appropriate generalization of the second-order expansion in (5) then reads

$$\hat{\theta} - \theta_0 \approx (NT)^{-1/2}\psi_1 + (NT)^{-1}\psi_2, \tag{22}$$

where

$$\begin{aligned}\psi_1 &:= - \left(\mathbb{E}_{NT} \bar{\ell}_{it}^{\theta\theta} \right)^{-1} \mathbb{G}_{NT} \ell_{it}^{\theta}, \\ \psi_2 &:= - \left(\mathbb{E}_{NT} \bar{\ell}_{it}^{\theta\theta} \right)^{-1} \mathbb{G}_{NT} \ell_{it}^{\theta\theta} \psi_1 - \frac{1}{2} \sum_{k=1}^p \left(\mathbb{E}_{NT} \bar{\ell}_{it}^{\theta\theta} \right)^{-1} \left(\mathbb{E}_{NT} \bar{\ell}_{it}^{\theta\theta k} \right) \psi_1 \psi_{1,k}.\end{aligned}$$

Here, the expected Hessian $\mathbb{E} \ell_j^{\theta\theta}$ in (5) was replaced with the sample average of the expected Hessian $\mathbb{E}_{NT} \bar{\ell}_{it}^{\theta\theta} = (NT)^{-1} \sum_{i,t} \mathbb{E} \ell_{it}^{\theta\theta}$, and analogously for the third derivative term $\mathbb{E} \ell_j^{\theta\theta k}$. Similarly, we need to define $\mathbb{G}_{NT} \ell_{it}^{\theta\theta}$ here as $\mathbb{G}_{NT} \ell_{it}^{\theta\theta} = (NT)^{-1/2} \sum_{i,t} \left(\ell_{it}^{\theta\theta} - \bar{\ell}_{it}^{\theta\theta} \right)$. Apart from those changes the derivation in section 2.1 still applies. One always first needs an additional consistency argument for $\hat{\theta}$ before the expansion (22) becomes applicable.

One-way Fixed Effects

We start by considering the estimation problem (12) with only individual specific effects, that is, $(\hat{\beta}, \hat{\alpha}) = \operatorname{argmax}_{(\beta, \alpha) \in \mathbb{R}^{d_\beta + N}} \mathbb{E}_{NT} \ell_{it}(\beta, \alpha_i)$. It is convenient to define

$$\ell_{it}^*(\beta, \alpha_i) := \ell_{it}(\beta, \alpha_i^*(\beta, \alpha_i)), \quad \alpha_i^*(\beta, \alpha_i) = \alpha_i - \left(\mathbb{E}_T \bar{\ell}_{it}^{\alpha_i \alpha_i} \right)^{-1} \left(\mathbb{E}_T \bar{\ell}_{it}^{\beta \alpha_i} \right)'. \quad (23)$$

The profile objective functions $\max_{\alpha} \mathbb{E}_{NT} \ell_{it}(\beta, \alpha_i)$ and $\max_{\alpha} \mathbb{E}_{NT} \ell_{it}^*(\beta, \alpha_i)$ have the same maximizer $\hat{\beta}$, because we have simply re-parameterized α_i . $\mathbb{E}_{NT} \ell_{it}^*(\beta, \alpha_i)$ is information-orthogonal between β and α_i , i.e. $\mathbb{E}_{NT} \bar{\ell}_{it}^{* \beta \alpha_i} = 0$ for all $i = 1, \dots, N$, where the omitted parameter means evaluation at the true parameters after the parameter transformation.¹¹ Thus, the expected Hessian of the objective function $\mathbb{E}_{NT} \ell_{it}^*(\beta, \alpha_i)$ is a block-diagonal matrix, with one $d_\beta \times d_\beta$ block $\mathbb{E}_{NT} \bar{\ell}_{it}^{* \beta \beta}$, and the rest of the matrix diagonal. By applying the expansion (22) to $\mathbb{E}_{NT} \ell_{it}^*(\beta, \alpha_i)$ with $\theta = (\beta, \alpha)$ we find

$$\hat{\beta} - \beta_0 \approx (NT)^{-1/2} \psi_{1,\beta} + (NT)^{-1} \psi_{2,\beta}, \quad (24)$$

with

$$\psi_{1,\beta} = H_{NT}^{-1} \mathbb{G}_{NT} \ell_{it}^{* \beta}, \quad \psi_{2,\beta} \approx H_{NT}^{-1} \sum_{i=1}^N \left[- \frac{\left(\mathbb{G}_T \ell_{it}^{* \beta \alpha_i} \right) \mathbb{G}_T \ell_{it}^{\alpha_i}}{\mathbb{E}_T \bar{\ell}_{it}^{\alpha_i \alpha_i}} + \frac{\left(\mathbb{E}_T \bar{\ell}_{it}^{* \beta \alpha_i \alpha_i} \right) \left(\mathbb{G}_T \ell_{it}^{\alpha_i} \right)^2}{2 \left(\mathbb{E}_T \bar{\ell}_{it}^{\alpha_i \alpha_i} \right)^2} \right],$$

where $H_{NT} = -\mathbb{E}_{NT} \bar{\ell}_{it}^{* \beta \beta}$. In $\psi_{2,\beta}$ we have dropped the terms that would originate from $\mathbb{G}_{NT} \ell_{it}^{\beta \beta}$, $\bar{\ell}_{it}^{\beta \beta}$ and $\bar{\ell}_{it}^{\beta \beta \alpha_i}$, because they only give smaller order terms. We also used that partial derivatives with respect

¹¹We can express partial derivatives of $\ell_{it}^*(\beta, \alpha_i)$ in terms of partial derivatives of $\ell_{it}(\beta, \alpha_i)$, for example, $\ell_{it}^{* \beta \alpha_i}(\beta, \alpha_i) = \ell_{it}^{\beta \alpha_i}(\beta, \alpha_i^*(\beta, \alpha_i)) - \ell_{it}^{\alpha_i \alpha_i}(\beta, \alpha_i^*(\beta, \alpha_i)) \left(\mathbb{E}_T \bar{\ell}_{it}^{\alpha_i \alpha_i} \right)^{-1} \left(\mathbb{E}_T \bar{\ell}_{it}^{\beta \alpha_i} \right)$. From this we also immediately find that $\mathbb{E}_T \bar{\ell}_{it}^{* \beta \alpha_i} = 0$, which also implies $\mathbb{E}_{NT} \bar{\ell}_{it}^{* \beta \alpha_i} = 0$.

to only α_i are equal for ℓ_{it} and ℓ_{it}^* . The leading bias of order T^{-1} in $\widehat{\beta}$ is thus given by $N^{-1}\mathbb{E}\psi_{2,\beta} \approx B_{NT}$, where

$$B_{NT} := H_{NT}^{-1}\mathbb{E}_N \left[-\frac{\mathbb{E}_T \sum_{s=t}^T \mathbb{E} \left(\ell_{it}^{\alpha_i} \ell_{is}^{*\beta\alpha_i} \right)}{\mathbb{E}_T \bar{\ell}_{it}^{\alpha_i\alpha_i}} + \frac{\left(\mathbb{E}_T \bar{\ell}_{it}^{*\beta\alpha_i\alpha_i} \right) \mathbb{E}_T \mathbb{E} \left(\ell_{it}^{\alpha_i} \right)^2}{2 \left(\mathbb{E}_T \bar{\ell}_{it}^{\alpha_i\alpha_i} \right)^2} \right]. \quad (25)$$

The last formula is also valid for conditional moment models, where the information equality $\mathbb{E} \left(\ell_{it}^{\alpha_i} \right)^2 = -\bar{\ell}_{it}^{\alpha_i\alpha_i}$ may not hold. Assuming a correctly specified likelihood and applying the information equality gives the result for $B = \text{plim}_{N,T \rightarrow \infty} B_{NT}$ in equation (15).

Two-way Fixed Effects

We now include the time effects as well, as in (12), that is, we use the expansion (22) with $\theta = (\beta, \alpha, \gamma)$. Some normalization of α and γ may be required, and this can also result in the Hessian matrix $\mathbb{E}_{NT} \bar{\ell}_{it}^{\theta\theta}$ to be singular, which needs to be accounted for in (22) by, for example, applying a pseudo-inverse instead of the regular inverse, but for our heuristic discussion here this is not important.

Consider $\ell_{it}^*(\beta, \alpha_i, \gamma_t)$ as defined in (13). This definition of ℓ_{it}^* guarantees that the expected Hessian matrix of $\mathbb{E}_{NT} \ell_{it}^*(\beta, \alpha_i, \gamma_t)$ is a block-diagonal matrix when evaluated at the true parameters, with two non-zero blocks $\mathbb{E}_{NT} \bar{\ell}_{it}^{*\beta\beta}$ and $\mathbb{E}_{NT} \bar{\ell}_{it}^{*\phi\phi}$, where $\phi = (\alpha, \gamma)$. Here, $\mathbb{E}_{NT} \bar{\ell}_{it}^{*\phi\phi}$ is not a diagonal matrix, but its diagonal elements dominate, and in Fernández-Val & Weidner (2016) it is shown that its (pseudo-) inverse can be approximated by the inverse of its diagonal part. The expansion in (24) is thus almost unchanged, we just need to add the terms in $\psi_{2,\beta}$ that stem from summing over the parameters γ_t as well,

$$\widehat{\beta} - \beta_0 \approx (NT)^{-1/2} \psi_{1,\beta} + (NT)^{-1} \psi_{2,\beta},$$

where

$$\begin{aligned} \psi_{1,\beta} &= H_{NT}^{-1} \mathbb{G}_{NT} \ell_{it}^{*\beta}, & \psi_{2,\beta} &\approx H_{NT}^{-1} \sum_{i=1}^N \left[-\frac{\mathbb{G}_T \ell_{it}^{*\beta\alpha_i} \mathbb{G}_T \ell_{it}^{\alpha_i}}{\mathbb{E}_T \bar{\ell}_{it}^{\alpha_i\alpha_i}} + \frac{\mathbb{E}_T \bar{\ell}_{it}^{*\beta\alpha_i\alpha_i} (\mathbb{G}_T \ell_{it}^{\alpha_i})^2}{2 \left(\mathbb{E}_T \bar{\ell}_{it}^{\alpha_i\alpha_i} \right)^2} \right] \\ & & &+ H_{NT}^{-1} \sum_{t=1}^T \left[-\frac{\mathbb{G}_N \ell_{it}^{*\beta\gamma_t} \mathbb{G}_N \ell_{it}^{\gamma_t}}{\mathbb{E}_N \bar{\ell}_{it}^{\gamma_t\gamma_t}} + \frac{\mathbb{E}_N \bar{\ell}_{it}^{*\beta\gamma_t\gamma_t} (\mathbb{G}_N \ell_{it}^{\gamma_t})^2}{2 \left(\mathbb{E}_N \bar{\ell}_{it}^{\gamma_t\gamma_t} \right)^2} \right], \end{aligned}$$

with $H_{NT} = -\mathbb{E}_{NT} \bar{\ell}_{it}^{*\beta\beta}$. We again dropped terms from $\psi_{2,\beta}$ that are asymptotically irrelevant. We now find $(NT)^{-1}\mathbb{E}\psi_{2,\beta} \approx T^{-1}B_{NT} + N^{-1}D_{NT}$, where the formula for B_{NT} in (25) is unchanged, and the time effects incidental parameter bias reads

$$D_{NT} = H_{NT}^{-1} \mathbb{E}_T \left[-\frac{\mathbb{E}_N \mathbb{E} \left(\ell_{it}^{\gamma_t} \ell_{it}^{*\beta\gamma_t} \right)}{\mathbb{E}_N \bar{\ell}_{it}^{\gamma_t\gamma_t}} + \frac{\left(\mathbb{E}_N \bar{\ell}_{it}^{*\beta\gamma_t\gamma_t} \right) \mathbb{E}_N \mathbb{E} \left(\ell_{it}^{\gamma_t} \right)^2}{2 \left(\mathbb{E}_N \bar{\ell}_{it}^{\gamma_t\gamma_t} \right)^2} \right],$$

where we have not used the information equality, yet, so that the formula holds for conditional moment models as well. Using $E(\ell_{it}^{\gamma_t})^2 = -\bar{\ell}_{it}^{\gamma_t}$ then gives the result for $D = \text{plim}_{N,T \rightarrow \infty} D_{NT}$ in equation (16) of the main text.

References

- Altonji JG, Matzkin RL. 2005. Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica* 73:1053–1102
- Alvarez J, Arellano M. 2003. The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71:1121–1159
- Andersen E. 1970. Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society. Series B (Methodological)* 32:283–301
- Anderson JE, van Wincoop E. 2003. Gravity with gravitas: A solution to the border puzzle. *American Economic Review* 93:170–192
- Arellano M. 2003a. Discrete choices with panel data. *Investigaciones Economicas* 27:423–458
- Arellano M. 2003b. Panel data econometrics. Oxford university press
- Arellano M, Bond S. 1991. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies* 58:277–297
- Arellano M, Bonhomme S. 2011. Nonlinear panel data analysis. *Annu. Rev. Econ.* 3:395–424
- Arellano M, Bonhomme S. 2017. Nonlinear panel data methods for dynamic heterogeneous agent models. *Annu. Rev. Econ.* 9
- Arellano M, Carrasco R. 2003. Binary choice panel data models with predetermined variables. *Journal of Econometrics* 115:125–157
- Arellano M, Hahn J. 2007. Understanding bias in nonlinear panel models: Some recent developments. *Econometric Society Monographs* 43:381
- Arellano M, Hahn J. 2016. A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects. *Global Economic Review* 45:251–274
- Arellano M, Honoré B. 2001. Panel data models: some recent developments. *Handbook of econometrics* 5:3229–3296

- Arellano M, Weidner M. 2017. Instrumental variable quantile regressions in large panels with fixed effects. *Unpublished manuscript*
- Bai J. 2009. Panel data models with interactive fixed effects. *Econometrica* 77:1229–1279
- Bai J, Wang P. 2016. Econometric analysis of large factor models. *Annual Review of Economics* 8:53–80
- Baltagi B. 2008. Econometric analysis of panel data. John Wiley & Sons
- Bester CA, Hansen C. 2009. A penalty function approach to bias reduction in nonlinear panel models with fixed effects. *Journal of Business & Economic Statistics* 27:131–148
- Blundell R, Griffith R, Windmeijer F. 1999. Individual effects and dynamics in count data models. *Working paper series no. W99/3, Institute for Fiscal Studies*.
- Blundell R, Griffith R, Windmeijer F. 2002. Individual effects and dynamics in count data models. *Journal of Econometrics* 108:113–131
- Boneva L, Linton O. 2017. A discrete-choice model for large heterogeneous panels with interactive fixed effects with an application to the determinants of corporate bond issuance. *Journal of Applied Econometrics*
- Bonhomme S. 2012. Functional differencing. *Econometrica* 80:1337–1385
- Bonhomme S, Lamadon T, Manresa E, et al. 2017. Discretizing unobserved heterogeneity. *Unpublished manuscript*
- Bonhomme S, Manresa E. 2015. Grouped patterns of heterogeneity in panel data. *Econometrica* 83:1147–1184
- Botosaru I, Muris C. 2017. Binarization for panel models with fixed effects. *Unpublished manuscript*
- Bun MJ, Kiviet JF. 2003. On the diminishing returns of higher-order terms in asymptotic expansions of bias. *economics Letters* 79:145–152
- Candelaria LE. 2016. A semiparametric network formation model with multiple linear fixed effects. *Unpublished manuscript*
- Carro J. 2007. Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics* 140:503–528
- Chamberlain G. 1980. Analysis of covariance with qualitative data. *The Review of Economic Studies* 47:225–238

- Chamberlain G. 1984. Panel Data. *Griliches and M. Intrilligator, eds., Handbook of Econometrics, Chapter 22* :1247–1318
- Chamberlain G. 2010. Binary Response Models for Panel Data: Identification and Information. *Econometrica* 78:159–168
- Charbonneau KB. 2014. Multiple fixed effects in binary response panel data models. *The Econometrics Journal*
- Chen M. 2016. Estimation of nonlinear panel models with multiple unobserved effects. *Unpublished manuscript*
- Chen M, Fernandez-Val I, Weidner M. 2014. Nonlinear panel models with interactive effects. *arXiv preprint arXiv:1412.5647*
- Chernozhukov V, Fernández-Val I, Hahn J, Newey W. 2013. Average and quantile effects in nonseparable panel models. *Econometrica* 81:535–580
- Chernozhukov V, Fernandez-Val I, Hoderlein S, Holzmann H, Newey W. 2015. Nonparametric identification in panels using quantiles. *Journal of Econometrics* 188:378–392
- Chernozhukov V, Fernández-Val I, Newey W. 2017a. Nonseparable multinomial choice models in cross-section and panel data. *arXiv preprint arXiv:1706.08418*
- Chernozhukov V, Fernández-Val I, Weidner M. 2017b. Network and panel quantile effects via distribution regression. *Unpublished manuscript*
- Chudik A, Pesaran MH, Yang JC. 2016. Half-panel jackknife fixed effects estimation of panels with weakly exogenous regressor. *Unpublished manuscript*
- Cox DR. 1958. The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)* :215–242
- Cox DR, Reid N. 1987. Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society. Series B (Methodological)* 49:1–39
- Cruz-González M, Fernández-Val I, Weidner M. 2015. probitfe and logitfe: Bias corrections for fixed effects estimators of probit and logit panel models. *Unpublished manuscript*
- Dhaene G, Jochmans K. 2015a. Profile-score adjustments for incidental-parameter problems. *Unpublished manuscript*
- Dhaene G, Jochmans K. 2015b. Split-panel jackknife estimation of fixed-effect models. *The Review of Economic Studies* 82:991–1030

- Dhaene G, Jochmans K. 2016. Likelihood inference in an autoregression with fixed effects. *Econometric Theory* 32:1178–1215
- Dzemski A. 2017. An empirical model of dyadic link formation in a network with unobserved heterogeneity. *Unpublished manuscript*
- Evdokimov K. 2010. Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. *Department of Economics, Princeton University*
- 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, Lee J. 2013. Panel data models with nonadditive unobserved heterogeneity: estimation and inference. *Quant. Econ.* 4:453–481
- Fernández-Val I, Vella F. 2011. Bias corrections for two-step fixed effects panel data estimators. *J. Econometrics* 163:144–162
- Fernández-Val I, Weidner M. 2016. Individual and time effects in nonlinear panel models with large n , t . *Journal of Econometrics* 192:291–312
- Freyberger J. 2017. Nonparametric panel data models with interactive fixed effects. *Review of Economic Studies (forthcoming)*
- Galvao AF, Kato K. 2016. Quantile regression methods for longitudinal data. *Unpublished manuscript*
- Gonçalves S, Kaffo M. 2015. Bootstrap inference for linear dynamic panel data models with individual fixed effects. *Journal of Econometrics* 186:407–426
- Graham BS. 2017. An econometric model of link formation with degree heterogeneity. *Econometrica (forthcoming)*
- Graham BS, Powell JL. 2012. Identification and estimation of average partial effects in irregular correlated random coefficient panel data models. *Econometrica* 80:2105–2152
- Greene W. 2004. The behavior of the fixed effects estimator in nonlinear models. *The Econometrics Journal* 7:98–119
- Hahn J, Kuersteiner G. 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica* 70:1639–1657
- Hahn J, Kuersteiner G. 2007. Bandwidth choice for bias estimators in dynamic nonlinear panel models. *Unpublished manuscript*

- Hahn J, Kuersteiner G. 2011. Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory* 27:1152–1191
- Hahn J, Moon HR. 2006. Reducing bias of mle in a dynamic panel model. *Econometric Theory* 22:499–512
- Hahn J, Moon HR. 2010. Panel data models with finite number of multiple equilibria. *Econometric Theory* 26:863–881
- Hahn J, Newey W. 2004. Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72:1295–1319
- Harrigan J. 1996. Openness to trade in manufactures in the oecd. *Journal of international economics* 40:23–39
- Hausman J, Hall BH, Griliches Z. 1984. Econometric Models for Count Data with an Application to the Patents-R & D Relationship. *Econometrica* 52:909–938
- Helpman E, Melitz M, Rubinstein Y. 2008. Estimating trade flows: Trading partners and trading volumes. *The Quarterly Journal of Economics* 123:441–487
- Hirji KF, Mehta CR, Patel NR. 1987. Computing distributions for exact logistic regression. *J. Amer. Statist. Assoc.* 82:1110–1117
- Hoderlein S, White H. 2012. Nonparametric identification in nonseparable panel data models with generalized fixed effects. *Journal of Econometrics* 168:300–314
- Honoré B. 1992. Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects. *Econometrica: Journal of the Econometric Society* 60:533–565
- Honoré B, Tamer E. 2006. Bounds on parameters in panel dynamic discrete choice models. *Econometrica* 74:611–629
- Honoré BE. 2002. Nonlinear models with panel data. *Portuguese Economic Journal* 1:163–179
- Honoré BE, Kyriazidou E. 2000. Panel data discrete choice models with lagged dependent variables. *Econometrica* 68:839–874
- Horowitz JL. 1992. A smoothed maximum score estimator for the binary response model. *Econometrica: journal of the Econometric Society* :505–531
- Hsiao C. 2014. Analysis of panel data. Cambridge University Press
- Kato K, Galvao AF, Montes-Rojas GV. 2012. Asymptotics for panel quantile regression models with individual effects. *Journal of Econometrics* 170:76–91

- Kim MS, Sun Y. 2016. Bootstrap and k-step bootstrap bias corrections for the fixed effects estimator in nonlinear panel data models. *Econometric Theory* 32:1523–1568
- Kiviet JF. 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *J. Econometrics* 68:53–78
- Koenker R. 2004. Quantile regression for longitudinal data. *Journal of Multivariate Analysis* 91:74–89
- Kyriazidou E. 1997. Estimation of a panel data sample selection model. *Econometrica: Journal of the Econometric Society* :1335–1364
- Kyriazidou E. 2001. Estimation of dynamic panel data sample selection models. *The Review of Economic Studies* 68:543–572
- Lancaster T. 2000. The incidental parameter problem since 1948. *Journal of Econometrics* 95:391–413
- Lancaster T. 2002. Orthogonal parameters and panel data. *The Review of Economic Studies* 69:647–666
- Li H, Lindsay BG, Waterman RP. 2003. Efficiency of projected score methods in rectangular array asymptotics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65:191–208
- Manski C. 1987. Semiparametric analysis of random effects linear models from binary panel data. *Econometrica: Journal of the Econometric Society* 55:357–362
- Matzkin RL. 2007. Nonparametric identification. *Handbook of Econometrics* 6:5307–5368
- Moon HR, Weidner M. 2017. Dynamic linear panel regression models with interactive fixed effects. *Econometric Theory* 33:158–195
- Moreira MJ. 2009. A maximum likelihood method for the incidental parameter problem. *The Annals of Statistics* 37:3660–3696
- Mundlak Y. 1978. On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society* :69–85
- Muris C. 2017. Estimation in the fixed-effects ordered logit model. *The Review of Economics and Statistics* 99:465–477
- Neyman J, Scott E. 1948. Consistent estimates based on partially consistent observations. *Econometrica* 16:1–32
- Nickell SJ. 1981. Biases in dynamic models with fixed effects. *Econometrica* 49:1417–26
- Pakel C. 2014. Bias reduction in nonlinear and dynamic panels in the presence of cross-section dependence. *Unpublished manuscript*

- Pakes A, Porter J. 2013. Moment inequalities for semiparametric multinomial choice with fixed effects. *Work. Pap., Harvard Univ., Cambridge, MA*
- Palmgren J. 1981. The fisher information matrix for log linear models arguing conditionally on observed explanatory variable. *Biometrika* 68:563–566
- Phillips PCB, Moon H. 1999. Linear regression limit theory for nonstationary panel data. *Econometrica* 67:1057–1111
- Quenouille MH. 1949. Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society. Series B (Methodological)* 11:68–84
- Quenouille MH. 1956. Notes on bias in estimation. *Biometrika* 43:353–360
- Rasch G. 1960. Probabilistic models for some intelligence and achievement tests. *Copenhagen: Danish Institute for Educational Research*
- Santos Silva J, Tenreyro S. 2006. The log of gravity. *The Review of Economics and statistics* 88:641–658
- Shao J, Tu D. 1995. The jackknife and bootstrap. *Springer Series in Statistics, New York* 85:8
- Shi X, Shum M, Song W. 2014. Estimating multinomial choice models using cyclic monotonicity. *Unpublished manuscript*
- Shi Z, Chen X. 2016. A structural pairwise network model with individual heterogeneity. *Unpublished manuscript*
- Stammann A, Hei F, McFadden D. 2016. Estimating fixed effects logit models with large panel data. *Unpublished manuscript*
- Sun Y. 2016a. Likelihood-based inference for nonlinear models with both individual and time effects. *Unpublished manuscript*
- Sun Y. 2016b. xtspj: Split-panel jackknife estimation for nonlinear fixed-effect models. *Unpublished manuscript*
- Sun Y, Dhaene G. 2017. Second-order corrected likelihood for nonlinear models with fixed effects. *Unpublished manuscript*
- Torgovitsky A. 2016. Nonparametric inference on state dependence with applications to employment dynamics. *Unpublished manuscript*
- Tukey JW. 1958. Bias and confidence in not-quite large samples. *Ann. Math. Statist.* 29:614

- Wang F. 2017. Maximum likelihood estimation and inference for high dimensional nonlinear factor models. *Unpublished manuscript*
- Wooldridge JM. 2010. Econometric analysis of cross section and panel data. MIT Press, Cambridge, MA, 2nd ed.
- Woutersen T. 2002. Robustness against incidental parameters. *Unpublished manuscript*