University of Connecticut

Department of Economics Working Paper Series

Fixed Effects Estimation in Panel Nonlinear Fractional Response Models

Xiaoming Li
University of Connecticut

Working Paper 2011-11

June 2011

---

341 Mansfield Road, Unit 1063
Storrs, CT 06269–1063
Phone: (860) 486–3022
Fax: (860) 486–4463
http://www.econ.uconn.edu/

Fixed Effects Estimation in Panel Nonlinear Fractional Response Models

Xiaoming Li*

June 1, 2011

Abstract

Estimations of nonlinear panel models that include individual specific fixed effects are complicated by the incidental parameters problem, that is, the asymptotic bias in the estimation of typical fixed effects panel models generally results in inconsistent estimates. In this paper, I characterize the leading term of a large-T expansion of the biases in the nonlinear least square estimator (NLSE) and estimators of the average partial effects in panel fractional response models. The resulting estimator after analytical bias correction is robust to the incidental parameters bias and reduces the bias order from $O(T^{-1})$ to $O(T^{-2})$. I also examine the finite sample performance of the proposed estimator using a new data generating process in which panel fractional response variables are collapsed from repeated, clustered cross-sectional binary probit choices. A proof showing the generated data satisfies the identification assumption at the cluster level has been given. Simulation results suggest that, in the static case, the

*I would like to thank Stephen L. Ross, John M. Clapp, Dong_Jin Lee, and Gautam Tripathi for valuable guidance, Jeffrey M. Wooldridge for kindly providing data for the empirical application, as well as Tao Chen, Leshui He, and seminar participants at the University of Connecticut, for helpful discussion and/or comments. All errors are my own.

bias corrected estimator performs comparably to the quasi-maximum likelihood estimator (QMLE), which is the standard approach in the literature, for 8 or more periods, while in the dynamic case, the bias corrected estimators are substantially superior to those QMLE's.

JEL Classification Numbers: C23, C25, I22

Keywords: Fractional responses, Panel Data, Unobserved effects, Probit, Partial effects, Bias, Incidental parameters problem, Fixed effects, Bias Correction

1 Introduction

Fractional response models are widely studied in empirical economic studies. The traditional estimation approach, the log-odds ratio model, fails when we observe fractional response near the boundaries, and requires a strong independence assumption in order to consistently estimate the expected value of the fractional response. Papke and Wooldridge (1996) specify a logistic functional form model to estimate the conditional expected fractional response, for an application to 401(K) plan participation rates using cross sectional data. They apply the quasi-maximum likelihood estimation method proposed by Gourieroux et al. (1984) and maximize the Bernoulli log-likelihood function, to obtain a consistent and $\sqrt{N}$-asymptotically normal estimator.[1] With panel data, most fractional response analyses model unobserved heterogeneity as a random effect. Papke and Wooldridge (2008) extend the nonlinear functional form setting from cross sectional data to panel data by specifying an unobserved school district effect in an empirical study of test pass rates. They impose a structural relationship between explanatory variables and unobserved effects by employing Chamberlain (1980)'s device[2], and consistently estimate the average partial effects (hereafter, APEs) using a pooled probit analysis.[3] Assuming the distribution of individual heterogeneity conditional on the initial response variable and strictly exogenous regressors, Loudermilk (2007) specifies a "doubly censored" Tobit model under a dynamic setting for an application of firm dividend policy.

However, there are several situations where consistent estimation for fractional response

[1]Bernoulli-log likelihood function is a member of linear exponential family. Gourieroux et al. (1984) prove with the assumed conditional fractional response, a quasi-maximum likelihood estimator for a log-likelihood function belonging to a linear exponential family is consistent and asymptotically normal, regardless of the conditional distribution of y given x

[2]Technically this is a Mundlak (1978) version of Chamberlain's device, which assumes the unobserved effect is a function of the average of all explanatory variables, rather than a vector of all explanatory variables across all time periods.

[3]For pooled probit estimation, if the unobserved permanent heterogeneity is assumed normally distributed the estimated APEs are consistent, even though the parameter estimats are scaled and inconsistent. However, under more general assumptions concerning the unobserved effects, the APEs calculated from the pooled probit estimation are inconsistent. Although if one is willing to assume conditional independence of response variable over time, the estimated APEs from the random effects probit estimation will still be consistent.

models using panel data is challenging. Random effects models will be biased if either the unobserved heterogeneity is correlated with covariates, or the conditional distribution of unobserved heterogeneity is incorrectly specified. The natural solution to these problems is to estimate a fixed effects model, but in a short panel where the number of individuals, N, is assumed to go to infinity, and the number of periods, T, is fixed, these estimators will suffer from the incidental parameters bias, first noted by Neyman and Scott (1948) for static models and examined by Nickell (1981) for dynamic models.

Recently a number of alternative approaches that reduce the incidental parameters bias and can be applied in nonlinear models generally have been developed.[4] Hahn and Kuersteiner (2002) propose an analytical bias-correction of the structural parameters for linear dynamic fixed effects models. In dynamic models, time effects are often included to model a simple form of non-stationarity of the dependent variable or to represent an aggregation shock. Hahn and Moon (2006) show that the bias-correction method proposed in Hahn and Kuersteiner (2002) can be applied to dynamic linear panel models with both fixed individual and time effects included. The analytical bias-correction method is extended to nonlinear static fixed effects models with exogenous regressors in Hahn and Newey (2004). They also analyze the bias-correction for the APEs which are often the primary interest for nonlinear models. Fernandez-Val (2009) constructs an analytical expression of the bias formulae for dynamic fixed effects binary choice models using expected quantities, which has improved small sample properties in comparison to those using observed quantities. Further, he considers models with predetermined regressors by accounting for the serial correlation in the stochastic expansion of the estimates of the individual effects. The idea behind these approaches is to expand the incidental parameters bias of the structural parameters estimates on the order of T from the stochastic expansion of estimates of fixed effects, and then subtract an estimate of the leading term of the bias from the biased estimates.[5] The resulting

[4] A detailed survey of this bias-correction literature can be found in Arellano and Hahn (2005).

[5] There are two alternative approaches for reducing the bias of estimates of structural parameters. Carro (2007) derives a score-correction of the moment equation for dynamic panel data discrete choice models. Bester and Hansen (2009) propose a penalty function to correct the objective function involving cross-

bias-corrected estimator will hence reduce the order of the bias from $O(T^{-1})$ to $O(T^{-2})$.[6]

In this paper, I implement a nonlinear least square estimator (hereafter, NLSE) for the fixed effects fractional response models, and derive the bias correction for these model estimates. I compare the performance of this estimator using bias correction to the performance of the Quasi-MLE estimator, which is standard in the fractional response model literature. The fractional response variables are simulated by aggregating up to a panel from repeated, clustered cross-sections of binary choice outcomes using both static and dynamic experiments that have been designed for examining bias correction in binary choice models.[7] In the static case, the fixed effects NLSE with bias correction has an average partial effect bias that is less than 1% for eight or more time periods, and this bias is comparable to the bias for FE Quasi-MLE estimators. For smaller sample size, the Quasi-MLE outperforms the bias corrected NLSE. For the dynamic case, the average partial effect bias on the exogenous variable for the bias-corrected NLSE is less than 10% for $T = 8$, and Quasi-MLE outperforms the bias-corrected NLSE consistently, similar to the static case. However, the average partial effect bias on the dynamic variable is severely biased for the Quasi-MLE (33% for T=8, 39% for T=12, 43% for T=16, and 47% for T=20) while the bias in the bias corrected NLSE is much smaller (42% for T=8, 18% for T=12, 11% for T=16, and 6% for T=20).

The paper is organized as follows. Section 2 presents the analytical bias correction of parameter estimates and marginal effects for panel nonlinear fractional response models. Section 3 provides the design of generating fractional response variables and establishes the

products of scores and the Hessian matrix. These estimators are closely related with the bias-correction of the estimator as shown in Arellano and Hahn (2005).

[6]Smaller order of the bias such as $O(T^{-3})$ can be achieved through higher order of stochastic expansion of the estimator of fixed effects.

[7]In this paper, I assume the panel fractional responses are aggregated from repeated clustered cross sectional binary outcomes. I explicitly specify three kinds of unobservables for those binary outcomes, a cluster time invariant heterogeneity, a time-varying cluster-wise common shock, and an idiosyncratic error. Traditional binary choice probit models achieve consistency assuming those binary outcomes in a cluster are identically and independently distributed after controlling for the time invariant cluster heterogeneity. However, for any case in which a cluster time-varying common shock exists, those probit estimators suffer from the incidental parameter bias. Similarly, the Quasi-MLE estimator using panel fractional response data is also biased when cluster time-varying shocks exist. To correct the bias, knowledge of the distribution of conditional fraction response is required. On the other hand, fixed effects NLSE is robust to the distribution and therefore does not require such knowledge for bias correction.

improved small sample properties of bias-corrected fixed effects estimator to other estimators with Monte Carlo experiments. An empirical illustration is given in Section 4. Section 5 concludes, and the derivations of bias correction formulae are given in the Appendix.

2 Bias Correction for Fixed Effects Estimator of Panel Fractional Response Models

2.1 Incidental parameters problem in fixed effects nonlinear models

In this paper, I propose a large-T consistent bias-corrected estimator for fixed effects fractional response models with general predetermined regressors. As shown below, traditional fixed effect estimators of nonlinear models are biased when N goes to infinity while T is fixed. Suppose we have a panel data nonlinear model with a structural parameter of interest, θ_0, individual specific fixed effects, $\alpha_n, n = 1, \cdots, N$, identification assumption $E[\epsilon_{nt}|x_n^t, \alpha_n] = 0$,[8] and the objective function for observation n in year t, $\varphi_{nt}(\theta, \alpha_n)$. Concentrating out the α_n's, one gets the fixed effects estimator,

$$\hat{\theta} \stackrel{\text{def}}{=} \arg\max_{\theta} (NT)^{-1} \sum_{n=1}^{N} \sum_{t=1}^{T} \varphi_{nt}(\theta, \hat{\alpha}_n(\theta)),$$

where,

$$\hat{\alpha}_n(\theta) \stackrel{\text{def}}{=} \arg\max_{\alpha} T^{-1} \sum_{t=1}^{T} \varphi_{nt}(\theta, \alpha).$$

As Fernandez-Val (2009), I assume the model is dynamically complete and regressors are predetermined. Lagged dependent variables are a leading example of predetermined variables. There are many instances of empirical studies where regressors are not strictly exogenous.

[8]In what follows, for any random variables z, z_{nt} denotes the observation for neighborhood n in year t, while z_n^t denotes a vector with all observations for neighborhood n up through t, i.e., $z_n^t = (z_{n1}, \cdots, z_{nt})'$.

For example, Li et al. (2010) argue against the exogeneity of market activities in mortgage markets and proposes a linear probability model including neighborhood fixed effects for use of repeated cross-sections with predetermined mortgage application volumes. Loudermilk (2007) derives a fixed T consistent conditional maximum likelihood estimator for panel fractional response Tobit models that permit lagged dependent variables as regressors. She assumes that the other explanatory variables are exogenous and imposes a parametric assumption to control for group heterogeneity. In studying the effects of spending on student performance, Papke and Wooldridge (2008) consider a single endogenous variable case and applies control function methods for consistent estimation, arguing spending could be correlated with the time-varying district unobservables, in addition to being correlated with the unobserved district specific heterogeneity.

Fixed effects estimators for nonlinear models, $\hat{\theta}$, generally suffer from severe incidental parameters bias (Heckman, 1981; Greene, 2004). The bias arises because there are a finite number of observations, T, available to estimate the fixed effects, and the asymptotic error of those estimates of individual effects does not vanish as the sample size, N, grows. As long as the structural parameters are not separately estimated from the estimation of those fixed effects, the noise in the individual effects estimates contaminates the estimation of the structural parameters and marginal effects.[9]

From the usual extreme estimator properties, as $N \to \infty$ with T fixed, $\hat{\theta} = \theta_T + o_p(1)$, where,

$$\theta_T \stackrel{\text{def}}{=} \arg\max_{\theta} E_N \left[T^{-1} \sum_{t=1}^{T} \varphi_{nt}(\theta, \hat{\alpha}_n(\theta)) \right], \tag{2.1}$$

[9] An extensive literature has developed to achieve fixed T consistent estimator. Chamberlain (1984) illustrates a conditional maximum likelihood estimation of static logit models with exogenous regressors. Manski (1987) was the first successfully to consistently estimate a nonlinear panel data model with individual specific fixed effects, based on the maximum score estimator. Honor and Kyriazidou (2000) propose a smoothed version consistent logit estimation that allows lagged dependent variables as the explanatory variables. However, those estimators are very model specific and in general not identifiable without auxiliary assumptions. Moreover, those identifiable fixed T consistent estimators are generally not at the standard $\sqrt{N}$-rate. Further, as the consistency of these estimators comes from avoiding estimating the individual effects, they provide no guidance towards estimating the APEs, which are often the ultimate interest in empirical studies.

with $E_N[m_n] \overset{\text{def}}{=} \lim_{n\to\infty}\sum_{n=1}^{N} m_n/N$. However, since $E_N\left[T^{-1}\sum_{t=1}^{T}\varphi_{nt}(\theta,\hat{\alpha}_n(\theta))\right]$ is generally not equal to $\mathrm{E}_N\left[T^{-1}\sum_{t=1}^{T}\varphi_{nt}(\theta,\bar{\alpha}_n(\theta))\right]$, in general, $\theta_T \neq \theta_0$, where

$$\bar{\alpha}_n(\theta) \overset{\text{def}}{=} \arg\max_{\alpha} E_T\left[T^{-1}\sum_{t=1}^{T}\varphi_{nt}(\theta,\alpha_n)\right] \tag{2.2}$$

with $E_T[m_{nt}] \overset{\text{def}}{=} \lim_{t\to\infty}\sum_{t=1}^{T} m_{nt}/T$. But as $T\to\infty$, $\hat{\alpha}_n(\theta)\to\bar{\alpha}_n(\theta)$, and $\theta_T\to\theta_0$. For smooth objective functions,

$$\theta_T = \theta_0 + \mathbf{B}/T + O\left(1/T^2\right), \tag{2.3}$$

for some $\mathbf{B}$. A derivation for the analytical expression of $\mathbf{B}$ is given in the Appendix. Under standard regularity conditions, as $N\to\infty$, it is generally true that

$$\sqrt{NT}(\hat{\theta}-\theta_T) \xrightarrow{d} \mathrm{N}(0, \mathrm{I}_\theta^{-1}\Omega\mathrm{I}_\theta^{-1}) \tag{2.4}$$

for some $\mathbf{I}_\theta$ and Ω defined below. Combining equations (2.3) and (2.4), one gets

$$\sqrt{\mathrm{NT}}(\hat{\theta}-\theta_0) = \sqrt{\mathrm{NT}}(\hat{\theta}-\theta_T) + \sqrt{NT}\mathbf{B}/T + O_p(\sqrt{N/T^3}). \tag{2.5}$$

As we can see, even T grows at the same rate as N, i.e., $N/T\to\rho$, $\hat{\theta}$ is consistent, but the limiting distribution is not centered around the true parameter value and the inference based on the asymptotic confidence interval is misleading.

2.2 Bias correction for fixed effects estimator

Hahn and Kuersteiner (2007) characterize the expression of the bias $\mathbf{B}$ for nonlinear fixed effects models using a first order stochastic expansion of $\hat{\alpha}_n$. Following their approach, I rederive these expressions for general nonlinear models with predetermined regressors using a second order stochastic expansion and the resulting notations will be intensively used in the fixed effects nonlinear fractional response models later.[10]

Specifically, let

$$u_{nt}(\theta,\alpha_n) \stackrel{\text{def}}{=} \partial\varphi_{nt}(\theta,\alpha_n)/\partial\theta, \qquad v_{nt}(\theta,\alpha_n) \stackrel{\text{def}}{=} \partial\varphi_{nt}(\theta,\alpha_n)/\partial\alpha_n,$$
$$U_{nt}(\theta,\alpha_n) \stackrel{\text{def}}{=} u_{nt}(\theta,\alpha_n) - v_{nt}(\theta,\alpha_n)E_T[v_{nt\theta}]/E_T[v_{nt\alpha}],$$

and additional subscripts denote partial derivatives, e.g., $u_{nt\alpha}(\theta,\alpha_n) \stackrel{\text{def}}{=} \partial u_{nt}(\theta,\alpha_n)/\partial\alpha_n$. Note, $U \stackrel{\text{def}}{=} E_N E_T[U_{nt}]$ and $\Omega \stackrel{\text{def}}{=} E_N E_T[\text{var}(T^{-1/2}\sum_{t=1}^{T} U_{nt})]$ are the score and Hessian matrix of θ after concentrating out of the individual fixed effects, respectively. Following Fernandez-Val (2009), as $T \to \infty$, a higher order stochastic expansion for the estimates of the individual effects is,

$$\begin{aligned}
&\hat{\alpha}_n = \alpha_n + \Psi_n/\sqrt{T} + \beta_n/T + o_p(1/T),\\
&\Psi_n = \sum_{t=1}^{T} \Psi_{nt}/\sqrt{T} \xrightarrow{d} \mathrm{N}(0,\sigma_n^2),\\
&\Psi_{nt} = -E_T[v_{nt\alpha}]^{-1}v_{nt}, \quad \sigma_n^2 = \bar{E}_T[\Psi_{nt}\Psi_{ns}], \quad \text{and} \qquad (2.6)\\
&\beta_n = -E_T[v_{nt\alpha}]^{-1}\{\bar{E}_T[v_{nt\alpha}\Psi_{ns}] + E_T[v_{nt\alpha\alpha}]\sigma_n^2/2\},
\end{aligned}$$

with $\bar{E}_T[h_{nt}k_{ns}] \stackrel{\text{def}}{=} \sum_{j=-\infty}^{j=\infty} E_T[h_{nt}k_{n,t-j}]$, where σ_n^2 and β_n are the asymptotic variance and bias components, respectively. As shown in the Appendix, the leading term of the large-T expansion of the asymptotic bias is

$$T(\hat{\theta}-\theta_0) \xrightarrow{p} T(\theta_T-\theta_0) = \mathbf{I}^{-1}\mathbf{b} \stackrel{\text{def}}{=} \mathbf{B},$$

[10] Fernandez-Val (2009) considers the case of likelihood setting where information identity and the Bartlett equality hold. Here, I examine the general case where the bias correction expressions are valid for any fixed effects m-estimator. See additional details in Arellano and Hahn (2005).

where, $\mathbf{I} = E_N[\mathbf{I}_n]$, the Fisher information for θ, with

$$\mathbf{I}_n = -E_T[U_{nt\theta}] = -\{E_T[u_{nt\theta}] - E_T[u_{nt\alpha}]E_T^{-1}[\upsilon_{nt\alpha}]E_T[\upsilon_{nt\theta}]\}, \tag{2.7}$$

and $\mathbf{b} = E_N[\mathbf{b}_n]$, the bias in the score of θ, with

$$\mathbf{b}_n = E_T[u_{nt\alpha}]\beta_n + \bar{E}_T[u_{nt\alpha}\Psi_{ns}] + E_T[u_{nt\alpha\alpha}]\sigma_n^2/2.\text{[11]} \tag{2.8}$$

The bias correction approach constructs the bias-corrected estimator by removing the estimated bias from the fixed effects estimate, i.e., $\tilde{\theta} = \hat{\theta} - \hat{\mathbf{B}}/T$, and reduces the order of the bias from $O(T^{-1})$ to $O(T^{-2})$. As we can see, under the assumption, $N/T^3 \to 0$, the bias-corrected estimator has the limiting distribution being centered at 0, i.e.,

$$\sqrt{NT}(\tilde{\theta} - \theta_0) = \sqrt{NT}(\tilde{\theta} - \theta_T^c) + O_p(\sqrt{N/T^3}),$$

where $\theta_T^c = \theta_0 + O(T^{-2})$ is the probability limit of $\tilde{\theta}$.

2.3 Bias correction of structural parameters in fixed effects fractional response models

I begin with the standard treatment of the fractional response models as described in Papke and Wooldridge (2008). The standard assumption for consistency of the traditional Quasi-MLE estimator is

$$E(y_{nt}|\mathbf{x}_{nt}, \alpha_n) = \Phi(\mathbf{x}_{nt}^{'}\theta + \alpha_n), \quad t = 1, \cdots, T, \tag{2.9}$$

[11] There are three sources of the bias in the score of θ: the first term coming from the asymptotic bias of $\hat{\alpha}_n$; the second term arising from the correlation between $u_{nt\alpha}$ and Ψ_{ns} because $\hat{\alpha}_n$ and $\hat{\theta}$ depend on the same data; and the third term being a pure random parameters bias corresponding to the standard bias formula for nonlinear functions of sample averages.

where, $\Phi(\cdot)$ denotes the normal cumulative distribution function. Under this assumption, the objective function of nonlinear least square estimation for cluster n at time t can be specified as

$$\varphi_{nt} = [y_{nt} - \Phi(\mathbf{x}_{nt}^{'}\theta + \alpha_n)]^2. \tag{2.10}$$

Proposition 1. *{Bias for Model Parameters}. Assume that (i) $E_T[\epsilon_{nt}|X_n^t, \alpha_n] = 0$ for $n = 1, \cdots, N$ and $t = 1, \cdots, T$; (ii) $E_T[W_{nt}W_{nt}^{'}]$ exists and is nonsingular for each n with $W_{nt} = (1, X_{nt}^{'})^{'}$; (iii) $\{Y_n^T, X_n^T, \alpha_n\}_{n=1}^N$ are independent; (iv) $N = o(T^3)$; (v) $\{X_n^T\}_{n=1}^N$ is identically distributed conditional on α_n. Then, $\boldsymbol{B} = \boldsymbol{I}^{-1}\boldsymbol{b}$, where $\boldsymbol{I} = E_N[\boldsymbol{I}_n]$ with $\boldsymbol{I}_n = -2E_T[\phi_{nt}^2 x_{nt} x_{nt}^{'}] + 2E_T[\phi_{nt}^2]^{-1} E_T[\phi_{nt}^2 x_{nt}] + E_T[\phi_{nt}^2 x_{nt}^{'}]$, and $\boldsymbol{b} = E_N[\boldsymbol{b}_n]$ with $\boldsymbol{b}_n = 2E_T[\phi_{nt}^2 x_{nt}]\beta_n - 3\sigma_n^2 E_T[\phi_{nt}^2(x_{nt}^{'}\theta + \alpha_n)x_{nt}] + 2\tilde{E}_T[\phi_{nt}^2 \Phi_{ns} x_{nt}]$.*

Proof. See Appendix. □

The one-step bias-estimate, $\hat{\mathbf{B}} = \hat{\mathbf{I}}(\hat{\theta})^{-1}\hat{\mathbf{b}}(\hat{\theta})$, is constructed by replacing Φ_{nt}, ϕ_{nt}, and $\tilde{E}_T[h_{nt}k_{ns}]$ with the corresponding sample analogs, $\hat{\Phi}_{nt}$, $\hat{\phi}_{nt}$, and $\hat{\tilde{E}}_{TJ}[h_{nt}k_{ns}]$, respectively, evaluated at $\hat{\theta}, \hat{\alpha}_1, \cdots, \hat{\alpha}_N$, e.g., $\hat{\Phi}_{nt} = \Phi_{nt}(\hat{\theta}, \hat{\alpha}_n(\hat{\theta}))$.[12]

2.4 Bias correction for APEs in fixed effects fractional response models

In the previous section, I demonstrate how to correct the incidental parameters bias for the index coefficient estimates in panel nonlinear fractional response models. However, in many empirical studies, one may be more interested in examining the marginal effects than the structural parameters themselves. Similar to the estimates of the structural parameters, estimating the APEs involves the estimates of individual effects, $\hat{\alpha}_n$, and as so even evaluated at the true structural parameter value, i.e., $\hat{\mu}(\theta_0, \hat{\alpha}_n)$, the estimated APEs are biased. In this section, I illustrate the bias-correction for the estimates of APEs.

[12] I follow Hahn and Kuersteiner (2007) to fix a bandwidth parameter J, for spectral expectation estimation satisfying $J/\sqrt{T} \to 0$ as $T \to \infty$, and compute $\hat{\tilde{E}}_{TJ}[h_{nt}k_{ns}] \stackrel{\text{def}}{=} \sum_{j=1}^{J}\sum_{t=j+1}^{T} h_{nt}k_{n,t-j}/(T-j)$.

Specifically, suppose $m(x_{nt}, \theta_0, \alpha_n)$ denotes the marginal effects, when x_{1nt} is continuous, the marginal effect on the conditional probability, y_{nt}, evaluated at $x_{nt}^{\circ} = (x_{1nt}^{\circ}, x_{-1nt}^{\circ\prime})'$, is given by

$$m(x_{nt}^{\circ}, \theta, \alpha_n) \stackrel{\text{def}}{=} \theta_1 \phi(x_{nt}^{\circ\prime}\theta + \alpha_n), \tag{2.11}$$

where $\theta = (\theta_1, \theta_{-1}^{'})'$; while when x_{1nt} is a discrete variable,

$$m(x_{nt}^{\circ}, \theta, \alpha_n) \stackrel{\text{def}}{=} \Phi(x_{nt}^{\circ\prime}\theta_1 + \alpha_n + \theta_1) - \Phi(x_{nt}^{\circ\prime}\theta + \alpha_n). \tag{2.12}$$

Clearly, the marginal effects in the nonlinear models depend on the level chosen for the regressor x_{nt} and the neighborhood unobserved heterogeneity α_n. Due to the nonlinearity, the effects for mean values of the regressors might be inappropriate. Chamberlain (1984) suggests reporting the average marginal effect for a neighborhood randomly drawn from the population. Thus,

$$\mu(\theta) = \int \mathrm{m}(x_{nt}, \theta, \alpha_n) d\mathrm{H}_{x_{nt},\alpha_n}(x_{nt}, \alpha_n) \tag{2.13}$$

for the continuous case or

$$\mu(\theta) = \int \mathrm{m}((x_1, x_{-1nt}^{'})', \theta, \alpha_n) d\mathrm{G}_{x_{-1nt},\alpha_n}(x_{-1nt}, \alpha_n) \tag{2.14}$$

for the discrete case, where H and G are the joint distributions of (x_{nt}, α_n) and (x_{-1nt}, α_n), respectively.

Proposition 2. $\{$*Bias for Marginal Effects*$\}$. *Let* $\tilde{\mu} \stackrel{def}{=} N^{-1}\sum_{n=1}^{N} \tilde{E}_T[m(x_{nt}, \tilde{\theta}, \tilde{\alpha}_n(\tilde{\theta}))]$ *and* $\mu_0 = E_N E_T[m(x_{nt}, \theta_0, \alpha_n(\theta_0))]$, *then under the conditions of Proposition 1, as* $N, T \to \infty$, $\tilde{\mu} = \mu_0 + \Delta/T + o_p(T^{-2})$, *where* $\Delta = E_N[\Delta_n]$ *with* $\Delta_n = E_T[m_\alpha(x_{nt}, \theta_0, \alpha_n)]\beta_n + \tilde{E}_T[m_\alpha(x_{nt}, \theta_0, \alpha_n)\Psi_{ns}] + E_T[m_{\alpha\alpha}(x_{nt}, \theta_0, \alpha_n)]\sigma_n^2/2.$[13]

[13]Subscripts on m denote partial derivatives. Fox example, for the continuous case, the marginal effect of x_{1nt}, evaluated at $x_{nt} = (x_{1nt}, x_{-1nt}^{'})'$ is $m(x_{nt}; \theta, \alpha_n) = \theta_1\phi(x_{nt}^{'}\theta + \alpha_n)$, so the first and second derivatives with respect to α_n can be expressed as $m_\alpha(x_{nt}; \theta, \alpha_n) = -\theta_1\phi(x_{nt}^{'}\theta + \alpha_n)$, and $m_{\alpha\alpha}(x_{nt}; \theta, \alpha_n) = \theta_1^2\phi(x_{nt}^{'}\theta + \alpha_n)$, respectively.

Proof. See Appendix. $\square$

Similarly, the bias-corrected estimator for average partial effects can be constructed as $\tilde{\mu} = \hat{\mu}(\tilde{\theta}) - \hat{\Delta}/T$ where $\hat{\Delta}$ and $\hat{\bar{E}}_T[m_\alpha(x_{nt}, \tilde{\theta}, \tilde{\alpha}_n)\tilde{\Psi}_{ns}]$ are the sample analogs of Δ and $\bar{E}_T[m_\alpha(x_{nt}, \theta_0, \alpha_n)\Psi_{ns}]$, evaluated at $\tilde{\theta}$, respectively.

3 Monte Carlo Experiments

In order to evaluate the small sample properties of proposed bias-corrected estimator, I conduct some Monte Carlo experiments. For fractional response models, as the dependent variable has to lie in the unit interval and the two end points come with positive probabilities, how to generate the simulated data is a concern. In this section, I propose a novel data generating process, in which I first generate repeated clustered cross sectional binary choices, and then aggregate each group of observations for one cluster-time to get the fractional response. The generated data satisfies the identification assumption I make in equation (2.9), the same condition as required for Quasi-MLE in Papke and Wooldridge (2008).

Proposition 3. {*Identification Assumption Satisfied Data Generating Process*}. *Assume that (i) individual mortgage underwriting decisions follow a latent binary process, i.e.,* $y_{nti} = \mathbf{1}\{x_{nti}\theta + \alpha_n + \epsilon_{nti} \geq 0\}$*; (ii)* $x_{nti} \equiv x_{nt}$*;*[14] *(iii)* $\{u_{nt}\}_{t=1}^{T}$ *is a sequence of random variables such that* $u_{nt}|x_n^t, \alpha_n \sim i.i.d.N(0, \rho)$ *with* $\rho \in [0, 1]$ *for each n; (iv)* $\{v_{nti}\}_{i=1}^{I_{nt}}$ *is a sequence of random variables such that* $v_{nti}|x_{nj}^t, \alpha_n, u_{nt} \sim i.i.d.N(0, 1-\rho)$ $\forall j = 1, \cdots, I_{nt}$ *for each i, where* I_{nt} *is the size of cluster n at time t; (v)* $\epsilon_{nti} = u_{nt} + v_{nti}$*; (vi)* $y_{nt} = \frac{1}{I_{nt}} \sum_{i=1}^{I_{nt}} y_{nti}$*. Then,*

$$E[y_{nt}|x_{nt}, \alpha_n] = \Phi(x_{nt}\theta + \alpha_n). \tag{3.1}$$

[14] Technically the proof can go through when individuals in a cluster-time group have different characteristics but are assumed to be normally distributed. In that case, the product of individual characteristic deviations from the group averages and the structural parameters for each individual is unobserved at the cluster level and becomes a random component embedded in the conditional expectation of panel fractional responses. Similar to the pooled analysis, the estimator of APEs using group average characteristics is consistent even parameters estimators are scaled and inconsistent. Here, as I focus on the fractional response models using variables at the group level, for simplicity, I restrict individuals in a cluster to share the same characteristics and explicitly make the scale to be 1.

Proof. See Appendix. □

For computational efficiency, I conduct the simulation in the following way. I first generate the number of trials, I_{nt}, following a half-normal distribution with a variance 10, conditional on the cluster heterogeneity.[15] I assume the cluster heterogeneity, α_n, and the time-varying cluster common shock, u_{nt}, together capture all the correlation among the cross-sections in a cluster. Note, after controlling for the cluster heterogeneity, α_n, and the time-varying cluster-wise common shock, u_{nt}, cross-sections in cluster n at time t experience idiosyncratic shocks, i.e., $v_{nti} \sim \mathrm{N}(0, \sigma_v^2)$, then $\Phi((x_{nt}\theta + \alpha_n + u_{nt})/\sigma_v)$ could be treated as the fundamental probability of success for all cross sections in a cluster-time group. A binomial process with the number of trials and the fundamental probability of success as arguments gives the number of successes, which is used to calculate the fractional response, y_{nt}. This process is much faster than generating the binary choices for each cross-section, y_{nti}, and aggregating them up for the fractional response.

3.1 Static fractional response model

The model design for a static case is

$$
\begin{aligned}
& y_{nti} = \mathbf{1}\{x_{nt}\theta_0 + \alpha_n + \epsilon_{nti} \geq 0\}; \quad \alpha_n \sim \mathrm{i.i.d.N}(0,1); \\
& I_{nt} \sim 1 + \mathrm{round}(\mathrm{i.i.d.HN}(\alpha_n, 10)); \quad \epsilon_{nti} = u_{nt} + v_{nti} \sim \mathrm{N}(0,1); \\
& u_{nt} \sim \mathrm{N}(0, \rho); \quad v_{nti} \sim \mathrm{N}(0, 1-\rho); \quad \rho = 0.5; \quad u_{nt} \perp\!\!\!\perp v_{nti}; \\
& x_{nt} = .1t + .5x_{nt-1} + \omega_{nt} \quad \text{for } t = 1, \cdots, T; \quad x_{n0} = \omega_{n0}; \\
& \omega_{nt} \sim \mathrm{i.i.d.U}(-.5, .5); \quad N = 100; \quad T = 4, 8, 12; \quad \theta_0 = 1;
\end{aligned}
\tag{3.2}
$$

where, u_{nt} is a cluster-time common shock, v_{nti} represents an individual idiosyncratic shock, α_n is a cluster unobserved effect, and N and U denote normal and uniform distributions,

[15] All generated cluster sizes are incremented by 1 to avoid the case where there is no observation in a cluster at a time. Current the average cluster size is around 9.1 with a variance of 10. The results are robust when I increase the cluster size to 40 by increasing the variance to 50.

respectively.[16] Through out the tables reported, SD represents the standard deviation of the estimator over 1000 replications; P:# is a rejection frequency of 1000 replications with # denoting the nomial value; SE/SD gives the ratio of the average standard error to the standard deviation of the estimators. For the proposed bias-corrected estimator, I choose the bandwidth parameter $M = 1$ for 12 or less periods, while $M = 2$ for 16 or more periods.

Table 1 presents the Monte Carlo results for the estimates of index coefficients and marginal effects[17] for panel fractional response models in the static case. Throughout the tables, QMLE represents the estimator maximizing Bernoulli log-likelihood function; Pooled NLSE and Pooled QMLE are estimators pooling the unobserved cluster specific effects and the error term together; FE NLSE and FE QMLE are estimators treating those cluster specific effects as parameters to be estimated; and Corr. FE NLSE corresponds to the bias corrected estimator based on FE NLSE. For pooled estimators, as the heterogeneity values are drawn from a standard normal distribution, the variance is two and the true parameter value for pooled estimators becomes to $1/\sqrt{2} = .7071$. For convenience, those pooled estimates are normalized according to the true parameter value. Not surprisingly, those pooled estimators are subject to only negligible bias once the scale is considered as the required assumption that unobserved cluster specific effects are conditionally normally distributed is matched by the generated data. Similar to Hahn and Newey (2004) and Fernandez-Val (2009), the fixed effects estimates are over-estimated with severe biases, more than 17% for 4 periods. In addition, column 5 and 6 indicate the limiting distribution of those fixed effects estimators are significantly biased, causing misleading inference. Bias-correction significantly reduces the estimated bias associated with fixed effects NLSE estimators, more than 10% generally, and substantially increases the accurancy of the inference. However, the fixed effects Quasi-MLE estimators consistently outperform the bias-corrected NLSE

[16] Traditionally, when specifying a probit model using repeated clustered cross-sections, one assumes $\epsilon_{nti} \sim$ i.i.d.N$(0, 1)$ to achieve consistent estimation. This corresponds to a special case where $u_{nt} = 0$ under the general data generating process. Obviously, for any case in which a cluster common shock exists, i.e., $u_{nt} \neq 0$, as $n \to \infty$ with T fixed, treating α_n as parameters to be estimated suffers from the incidental parameters problem.

[17] Estimates are normalized by dividing the true marginal effect values.

estimators for the very short panel, $T = 4$. Not surprisingly, the APEs estimates for pooled estimators are very close to the true value of marginal effects, even for the very short panel with $T = 4$. Wooldridge(2002) conjectures that the APEs estimate might have reasonable properties because the incidental parameters bias of the structural parameters might be offset by the asymptotic bias of the estimates of the fixed effects. However, this is not the case for the FE NLSE, which has a bias more than 19% for $T = 4$. The APEs estimates are significantly improved by the bias-correction, with comparable bias to the other estimates for eight or more periods.[18] Note, in terms of inference, the estimates of APEs and their standard errors do not have the proper size, which is standard in nonlinear models of this sort (Hahn and Newey, 2004; Bester and Hansen, 2009; Fernandez-Val, 2009).

There are three key parameters affecting the estimators in fractional response models, the coefficients for observed explanatory variables and two unobservables, cluster time-invariant heterogeneity and cluster specific time-varying errors. To examine the robustness of the proposed bias-corrected fixed effects estimator, Table 2 presents the mean estimates for the parameter and APEs under different scenarios with $T = 8$. Column 1 examines the case where time-varying cluster common shock are more variated, while column 2 corresponds to a common shock with a smaller variance. Column 3 and 4 examine the relative importance between time-invariant and time-varying cluster heterogeneities, while the influence of observed explanatory variables are tested in column 5 and 6. Column 7 and 8 present the examinations of the persistence in the explanatory variables. Generally, the proposed estimator is robust across specifications and performs comparable to the pooled Quasi-MLE in terms of estimates of APEs for $T = 8$.

3.2 Dynamic fractional response model

In the design of dynamic fractional response models, similarly, we first generate the data for individual binary choices which depends on a lagged fractional response in that cluster,

[18] Note, even the bias correction reduces a significant amount of bias, more than 7%, the APEs estimate after bias correction performs still worse than the other estimators in a very short panel, $T = 4$.

the common cluster explanatory variable, unobserved cluster time-invariant heterogeneity,[19] and an unobserved composite shock.

$$
\begin{aligned}
& y_{n0i} = \mathbf{1}\{x_{n0}\theta_0^x + \alpha_n + \epsilon_{n0i} \geq 0\}; \quad \alpha_n = \frac{1}{4}\sum_{t=0}^{3} x_{nt}; \quad I_{nt} \sim 1 + \text{round}(\text{i.i.d.HN}(\alpha_n, 10)); \\
& x_{nt} \sim \text{i.i.d.N}(0,1); \quad y_{nt} = \sum_{i=1}^{I_{nt}} y_{nti}/I_{nt}; \quad \epsilon_{nti} = u_{nt} + v_{nti} \sim \text{N}(0,1); \quad u_{nt} \sim \text{N}(0,\rho); \\
& v_{nti} \sim \text{N}(0, 1-\rho); \;\; \rho = 0.5; \;\; \text{and} \;\; u_{nt} \perp\!\!\!\perp v_{nti} \;\; \text{for } t = 0, \cdots, T-1; \qquad (3.3) \\
& y_{nti} = \mathbf{1}\{y_{nt-1}\theta_0^y + x_{nt}\theta_0^x + \alpha_n + \epsilon_{nti} \geq 0\} \quad \text{for } t = 1, \cdots, T-1; \\
& N = 100; \quad T = 8, 12, 16, 20; \quad \theta_0^y = .5; \quad \theta_0^x = 1;
\end{aligned}
$$

Dynamic QMLE corresponds to the estimator proposed by Wooldridge (2005) for dynamic nonlinear panel data random effects models. The initial condition specifies the conditional density of heterogeneities, e.g.,

$$
\alpha_n | (y_{n0}, x_n) \sim \text{N}(\eta + \bar{x}_n'\xi + y_{n0}\zeta, \sigma_a^2), \tag{3.4}
$$

where, $\bar{x}_n \equiv T^{-1}\sum_{t=1}^{T} x_{nt}$ is the time averages.[20]

Table 3 reports the Monte Carlo results for the parameter estimates of exogneous variable θ_x^0 and the ratio of the estimates of the APEs of μ_x^0 to the true value, respectively. Not surprisingly, as endogeneity is introduced through the dynamic variable, y_{nt-1}, those pooled estimators are no longer consistent and over-estimated for both estimates of index coefficients and marginal effects of exogneous variable, about 10%. Fixed effects estimators are robust to such a arbitrary correlation between explanatory variables and unobserved factors, but suffer from incidental parameters bias and come with misleading inferences. Similar to the static

[19] In particular, I follow the literature (Honor and Kyriazidou, 2000; Fernandez-Val, 2009) to generate the cluster heterogeneity.

[20] Note, with this specified initial condition, the pooled estimators are consistently estimated with a scale, $\sqrt{1+\sigma_a^2}$. However, as σ_a^2 is unobserved, we cannot recover the true parameter estimates and as so we do not report the unscaled parameter estimates for dynamic QMLE in the Monte Carlo experiments.

models, bias-correction significantly reduces the estimate biases and increases the accurancy of the inference. However, as before, bias-corrected estimators consistently underperform fixed effects QMLE estimators for the exogenous variable.

Table 4 reports the Monte Carlo results for the parameter estimates of endogenous variable θ_y^0 and the ratio of the estimates of the APEs of μ_y^0 to the true value, respectively. Similar to previous studies, (Carro, 2007; Fernandez-Val, 2009; Bester and Hansen, 2009), the identification of the dynamic effects is severely biased. The biases of those estimates for the lagged dependent variables are heterogenous, with upward biases for pooled estimates and downward biases for fixed effects estimates. In identifying the marginal effects as well as conducting inference on the lagged dependent variable, the bias-corrected estimates of APEs consistently outperform the other estimates, especially for 12 or more periods.[21] Interestingly, pooled estimators are more biased than the fixed effects estimators, indicating endogeneity as a problem is more severe than incidental parameters bias in dynamic studies. Overall, bias-correction approach substantially improves the estimation of index coefficients and achieves amazingly better inference properties for identifying true state dependence.

The bias correction estimator achieves consistency under the assumption that $N/T^3 \to 0$. It would be interesting to know how the estimator performs as the ratio of N/T changes and Table 5 reports the Monte Carlo results for robustness examination with different $N = 100, 200,$ and 400 but same $T = 12$. Not surprisingly, as the proposed bias-corrected estimator is large-T consistent, increasing N with T fixed does not improve the performance of the point estimator. However, bias correction substantially reduces the accurancy of the limiting distribution as shown by the increasing empirical rejection frequency, in estimates either for index coefficients or marginal effects, for both exogenous and endogenous variables.

Table 6 and 7 present the Monte Carlo results for robustness examination under different

[21]Note, here the cluster heterogeneity is specified as the average of the explanatory variables in the first 4 periods. For the dynamic Quasi-MLE, I specify the initial conditions conditional on the average of explanatory variables over time as well as the first period fractional response, which is standard in the literature (Wooldridge, 2002, 2005; Loudermilk, 2007), so that as more periods of data are available, the worse the performance of this estimator becomes.

scenarios with $T = 12$. These settings are same to those examined in Table 2, except in column 7 and 8, I examine the persistence of the dynamic effects. For the parameter estimates of either exogenous or endogenous variable, bias-corrected fixed effects estimators continue to outperform all the other estimators. Interestingly, for the case with very low persistence, column 7 in both tables, non-bias corrected fixed effects estimators come with sign reversal, while pooled estimators come with the right sign, but are more than 10 times larger in magnitude than the true value.

4 Empirical illustration: Effects of spending on student performance

In this section, I apply my bias corrected fixed effects NLSE in an empirical study to examine the extent to which spending per pupil will affect the student achievement. Research on the causality of spending on student performance is complicated because both variables can be jointly determined by unobserved factors. For example, unobserved parental efforts could help local schools to exploit more resources on education while also are helpful for students' achievement. Moreover, as test pass rates, when measured as a proportion, are necessarily bounded between zero and one, linear models might not capture the whole effect of spending on pass rates over the entire distribution of spending.

I use the data considered in Papke and Wooldridge (2008) in the QMLE estimation of test pass rates. The data consists of an annual panel of math test outcomes for fourth graders of 501 school districts in Michigan from 1995 to 2001. Descriptive statistics are given in Table 8. Following immediately the major change in the way K-12 education is financed in 1995 where the lowest spending districts have been provided with a foundation allowance significantly above their previous per-student funding. On average, test pass rates jump from 49.25% to 61.75%. After that, the average test pass rates continue to rise over time, to a peak, 77.45% in 2000. Nonetheless, there still exist large variations in terms of test pass rates as well as spending on education across school districts, the minimum pass rates being lower than 26% during the whole sample period while the maximum pass rates being always more than 97% and even 100% for most years. This highlights the necessity of controlling for unobserved school district specific effects. Unlike Papke and Wooldridge (2008) which specifies an explicit distributional form for these school district unobservables and models them as random effects, I specify a nonlinear probability model and control for these district unobservables with fixed effects that allows for arbitrary correlation between spending and unobserved school district specific effects.

In particular, I estimate the following equation:

$$E[Y_{it}|P_{it}, X_{it}, \alpha_i] = \Phi(\delta_t + P_{it}\theta_p + X_{it}^{'}\theta_x + \alpha_i), \quad t = 1, \cdots, T, \tag{4.1}$$

where, Y_{it} is the Michigan Education Assessment Program (MEAP) fourth-grade math test passing rate in district i in year t; P_{it} denotes the "foundation grant" per-student awarded by the state based funding in 1993/1994 in district i in year t; X_{it} is a three-dimensional vector of real expenditure per pupil (in logarithmic form), the fraction of students eligible for the free and reduced-price lunch programs, and district enrollment (in logarithmic form) in district i in year t.[22]

Table 9 reports the estimation results of index coefficients and marginal effects obtained from a variety of static specifications. Pooled estimators assume that unobserved school district specific effects conditional on explanatory variables are normally distributed, implying that unobserved school district specific effects and explanatory variables are independent. This is clearly violated when these effects such as unobserved parental efforts are likely to be correlated with both student performance as well as educational spending. Omission of this correlation subtantially biases the analyses of the effect of spending on student perfor-

[22]For computation simplicity, I focus on only normal cumulative distribution function (CDF). Use of a general nonlinear function form causes no conceptual or theoretical difficulity, but the computation is cumbersome.

mance.[23] Interestingly, there are no big changes from the RE specifications to FE specifications, either in the estimation of index coefficients (increase by 10%) or in marginal effects (increases by 6%), indicating that the conditional distributional form assumed explicitly in Papke and Wooldridge (2008) capture most of the correlation between school district unobservables and regressors. Surprisingly, bias correction does not change the estimation results much, especially given the potential endogeneity evidenced from the distinct estimation results for those pooled estimations, suggesting that the time-varying school district common shock only accounts for small share of the variation in the determination of each student' test pass result. While the estimation results are different in magnitude, these results tell a consistent theory that educational spending positively and significantly influences student performance. In my preferred specification, the bias-corrected one, if log spending increases by 0.10 - about a 10% increase in spending - the estimated pass rate increases by about 0.317 or 3.17%. The heteroscadasticity-robust t-statistic on this estimate is 2.03.

5 Conclusion

This paper derives the analytical expression for the leading terms of large-T expansion of the biases associated with fixed effects NLSE and marginal effects in fractional response models. With the new, computationally efficient data generating process, this paper shows for static models, Quasi-MLE estimators with or without fixed effects work well and in fact in very small samples outperform NLSE estimators with bias correction, but for dynamic models, the bias-corrected estimators substantially outperform those Quasi-MLE estimators for inference based on the parameter estimates and for estimating the marginal effects of the dynamic regressors. In the future, it would be useful to know whether this proposed bias-correction approach is robust to the introduction of heterogeneities in the observed individual characteristics, that is, whether the estimate of APEs adjusted with the bias-correction at

[23]The estimates of index coefficients are scaled due to the omission of unobserved school district effects. The marginal effects is consistent only if these effects are randomly distributed across school districts.

the cluster level is comparable to the estimate using repeated clustered cross-sections?

Appendix

A.1. Derivation of the Fisher information for θ: I

Note that $E_T[v_{nt}(\theta, \alpha_n)] = 0$. Under regularity conditions (expectation and differentiation can exchange the order), totally differentiating this equation and the objective function w.r.t. θ gives

$$\frac{d\alpha_n}{d\theta} = -\frac{E_T[v_{nt\theta}]}{E_T[v_{nt\alpha}]}, \tag{5.1}$$
$$\frac{d\varphi_{nt}}{d\theta} = \frac{\partial \varphi_{nt}}{\partial \theta} + \frac{\partial \varphi_{nt}}{\partial \alpha_n}\frac{d\alpha_n}{d\theta}. \tag{5.2}$$

Plugging equation (5.1) into equation (5.2) produces U_{nt} and as $N \to \infty$ and $T \to \infty$, it converges to $\mathbf{I} = E_N[\mathbf{I}_n] = -E_N E_T[U_{nt\theta}]$, the Fisher information for θ, where $\mathbf{I}_n = -\{E_T[u_{nt\theta}] - E_T[u_{nt\alpha}]E_T^{-1}[v_{nt\alpha}]E_T[v_{nt\theta}]\}$.

A.2. Derivation of the bias of the score for θ: B

Consider a score based on an infeasible estimator $\bar{\theta}$, solving

$$E_N E_T[U_{nt}(\bar{\theta}, \hat{\alpha}_n(\theta_0))] = 0. \tag{5.3}$$

Standard arguments suggest

$$U_{nt}(\bar{\theta}, \hat{\alpha}_n(\theta_0)) \approx U_{nt}(\theta_0, \hat{\alpha}_n(\theta_0)) + (\bar{\theta} - \theta_0)\partial U_{nt}(\theta_0, \alpha_n)/\partial \theta', \tag{5.4}$$

which combines equation (5.3) and implies

$$\bar{\theta} - \theta_0 = \mathbf{I}^{-1} E_N E_T[U_{nt}(\theta_0, \hat{\alpha}_n(\theta_0))]. \tag{5.5}$$

Note, making use of $E_T[v_{nt}(\theta_0, \hat{\alpha}_n(\theta_0))] = 0$ and expanding $E_N E_T[U_{nt}(\theta_0, \hat{\alpha}_n(\theta_0))]$ about $\hat{\alpha}_n(\theta_0)$ around α_n yield

$$\begin{aligned} E_N E_T[U_{nt}(\theta_0, \hat{\alpha}_n(\theta_0))] &= E_N E_T[u_{nt}(\theta_0, \hat{\alpha}_n(\theta_0))] \\ &\approx E_N E_T\{u_{nt}(\theta_0, \alpha_n) + u_{nt\alpha}(\hat{\alpha}_n(\theta_0) - \alpha_n) + u_{nt\alpha\alpha}(\hat{\alpha}_n(\theta_0) - \alpha_n)^2/2\} \\ &= E_N E_T\{u_{nt}(\theta_0, \alpha_n) + u_{nt\alpha}(\sum_{t=1}^{T} \Psi_{nt} + \beta_n)/T + \sum_{t=1}^{T}\sum_{s=1}^{T} \Psi_{ns} u_{nt\alpha\alpha}/(2T)\}. \end{aligned} \tag{5.6}$$

Combining equations (5.5) and (5.6), and noting $E_N E_T[u_{nt}(\theta_0, \alpha_n)] = 0$, one gets what we need to derive: $\mathbf{B} = \mathbf{I}^{-1}\mathbf{b}$, where $\mathbf{b} = E_N[\mathbf{b}_n]$ with $\mathbf{b}_n = E_T[u_{nt\alpha}]\beta_n + \bar{E}_T[u_{nt\alpha}\Psi_{ns}] + E_T[u_{nt\alpha\alpha}]\sigma_n^2/2$.

A.3. Proof of Proposition 1

Proof. By the Law of Iterated Expectations, the expressions for the bias can be specialized as the following:

$$u_{nt} = -2(y_{nt} - \Phi_{nt}(\mathbf{x}_{nt}^{'}\theta + \alpha_n))\phi_{nt}(\mathbf{x}_{nt}^{'}\theta + \alpha_n)x_{nt}, \quad \text{and} \tag{5.7}$$

$$v_{nt} = -2(y_{nt} - \Phi_{nt}(\mathbf{x}_{nt}^{'}\theta + \alpha_n))\phi_{nt}(\mathbf{x}_{nt}^{'}\theta + \alpha_n), \tag{5.8}$$

where $\phi(\cdot)$ corresponds to the normal probability density function, further,

$$\begin{aligned} &\Psi_{nt} = E_T[\phi_{nt}^2]^{-1}(y_{nt} - \Phi_{nt})\phi_{nt}, \quad \sigma_n^2 = E_T^2[\phi_{nt}^2]^{-1}E_T[(y_{nt} - \Phi_{nt})^2\phi_{nt}^2], \\ &\beta_n = -E_T^2[\phi_{nt}^2]^{-1}E_T[(y_{nt}, -\Phi_{nt})^2\phi_{nt}^2(x_{nt}^{'}\theta + \alpha_n)] - E_T[\phi_{nt}^2]^{-1}\tilde{E}_T[\phi_{nt}^2\Phi_{ns}] \\ &\quad + 3/2E_T^3[\phi_{nt}^2]^{-1}E_T[(y_{nt} - \Phi_{nt})^2\phi_{nt}^2]E_T[(y_{nt} - \Phi_{nt})^2\phi_{nt}^2(x_{nt}^{'}\theta + \alpha_n)], \\ &\mathbf{b}_n = 2E_T[\phi_{nt}^2 x_{nt}]\beta_n - 3\sigma_n^2 E_T[\phi_{nt}^2(x_{nt}^{'}\theta + \alpha_n)x_{nt}] + 2\tilde{E}_T[\phi_{nt}^2\Phi_{ns}x_{nt}] \\ &\quad + 2E_T[\phi_{nt}^2]^{-1}E_T[(y_{nt} - \Phi_{nt})^2\phi_{nt}^2(x_{nt}^{'}\theta + \alpha_n)x_{nt}], \quad \text{and} \\ &\mathbf{I}_n = -2E_T[\phi_{nt}^2 x_{nt}x_{nt}^{'}] + 2E_T[\phi_{nt}^2]^{-1}E_T[\phi_{nt}^2 x_{nt}] + E_T[\phi_{nt}^2 x_{nt}^{'}], \end{aligned} \tag{5.9}$$

where $\tilde{E}_T(h_{nt}k_{ns}) \overset{\text{def}}{=} \sum_{j=1}^{\infty} E_T[h_{nt}k_{n,t-j}]$. □

A.4. Proof of Proposition 2

Proof. First, note that by the Law of Large Number and Continuous Mapping Theorem, as $N/T^3 \to 0$, since $\tilde{\theta} \xrightarrow{p} \theta_0$, so $\tilde{\mu} \xrightarrow{p} \bar{\mu}$, where $\bar{\mu} = N^{-1}\sum_{n=1}^{N} \hat{E}_T[m(x_{nt}, \theta_0, \hat{\alpha}_n(\theta_0))]$. So the estimator of marginal effects is large-T consistent.

From the asymptotic stochastic expansion of $\hat{\alpha}_n(\theta_0)$, the leading term of the large-T expansion of the bias of APEs is,

$$T(\bar{\mu} - E_N E_T[m(x_{nt}, \theta_0, \alpha_n)]) \xrightarrow{p} \Delta, \tag{5.10}$$

where $\Delta = E_N[\Delta_n]$ with

$$\Delta_n = E_T[m_\alpha(x_{nt}, \theta_0, \alpha_n)]\beta_n + \tilde{E}_T[m_\alpha(x_{nt}, \theta_0, \alpha_n)\Psi_{ns}] + E_T[m_{\alpha\alpha}(x_{nt}, \theta_0, \alpha_n)]\sigma_n^2/2 \tag{5.11}$$

□

A.5. Proof of Proposition 3

Proof. First, conditions (i) and (v) imply the identification assumption for a probit model, $E[y_{nti}|x_{nti}, \alpha_n] = \Phi(x_{nti}\theta + \alpha_n)$. Note, combining condition (ii) into this identification assumption, we get

$$E[y_{nti}|x_{nti}, \alpha_n] = \Phi(x_{nti}\theta + \alpha_n) = \Phi(x_{nt}\theta + \alpha_n). \tag{5.12}$$

Now substituting condition (vi) into Equation (5.12), yields

$$\begin{aligned} E[y_{nt}|x_{nt}, \alpha_n] &= E[I_{nt}^{-1}\sum_{i=1}^{I_{nt}} y_{nti}|x_{nt}, \alpha_n] = I_{nt}^{-1}\sum_{i=1}^{I_{nt}} E[y_{nti}|x_{nt}, \alpha_n] \\ &= I_{nt}^{-1}\sum_{i=1}^{I_{nt}} \Phi(x_{nt}\theta + \alpha_n) = \Phi(x_{nt}\theta + \alpha_n), \end{aligned}$$

which is just the identification assumption (2.9) required for consistent estimation. □

Table 1: Estimators of θ ($\theta_0 = 1$) and μ ($\mu_0 = 1$) in Static Cases

Estimators	**Structure Parameters** $\theta(\theta_0 = 1)$					**Marginal Effects** $\mu(\mu_0 = 1)$				
	Mean	**SD**	**SE/SD**	**P:.05**	**P:.10**	**Mean**	**SD**	**SE/SD**	**P:.05**	**P:.10**
T = 4										
Pooled NLSE	1.01	0.23	0.89	0.08	0.13	1.01	0.22	0.80	0.11	0.19
Pooled QMLE	1.01	0.22	0.89	0.08	0.13	1.01	0.22	0.80	0.11	0.18
FE QMLE	1.17	0.21	0.86	0.19	0.27	1.00	0.17	0.80	0.12	0.19
FE NLSE	1.53	0.47	0.50	0.51	0.60	1.19	0.25	0.60	0.34	0.41
Corr. FE NLSE	1.35	0.68	0.33	0.34	0.42	1.12	0.24	0.60	0.30	0.37
T = 8										
Pooled NLSE	1.00	0.12	0.94	0.06	0.12	1.00	0.10	0.72	0.16	0.24
Pooled QMLE	1.00	0.11	0.96	0.06	0.11	1.00	0.10	0.73	0.15	0.22
FE QMLE	1.08	0.09	0.91	0.18	0.25	0.99	0.07	0.72	0.16	0.25
FE NLSE	1.20	0.14	0.75	0.46	0.57	1.05	0.09	0.69	0.23	0.31
Corr. FE NLSE	1.07	0.13	0.76	0.17	0.25	1.01	0.09	0.68	0.18	0.26
T = 12										
Pooled NLSE	1.01	0.09	0.92	0.06	0.13	1.00	0.06	0.51	0.32	0.41
Pooled QMLE	1.01	0.08	0.90	0.08	0.14	1.00	0.06	0.50	0.32	0.40
FE QMLE	1.06	0.06	0.94	0.16	0.26	0.99	0.05	0.49	0.35	0.41
FE NLSE	1.15	0.10	0.77	0.50	0.60	1.02	0.05	0.47	0.37	0.45
Corr. FE NLSE	1.04	0.09	0.76	0.17	0.25	0.99	0.05	0.47	0.34	0.42

[a] Note: 1000 replications. Pooled estimators are estimated without fixed effects included. If the heterogeneities are drawn from a standard normal distribution, together with the original error term, a new normal distribution with variance 2 is formed. Here, for the convenience to read, all pooled estimates are normalized by multiplying root 2. For finite sample, complete separation of data happens with positive probability. In this case, the objective function is either ∞ or $-\infty$ and the estimation cannot converge. Under this situation, I drop all estimators for this replication. In the simulations, the numbers of dropped replications are 12, 0, and 0, respectively.

Table 2: Robustness check for static cases with $T = 8$

Scenarios	$\sigma_u^2 = 0.25$ $\sigma_v^2 = 0.75$	$\sigma_u^2 = 0.75$ $\sigma_v^2 = 0.25$	$\sigma_u^2 = 0.25$ $\sigma_v^2 = 0.25$	$\sigma_u^2 = 0.75$ $\sigma_v^2 = 0.75$	$\theta_0 = 0.75$	$\theta_0 = 1.25$	$\rho = 0.05$	$\rho = 0.95$
Panel 1: Parameters ($\theta_0 = 1$)								
Pooled NLSE	1.00	1.00	1.01	1.00	1.00	1.00	1.00	1.01
Pooled QMLE	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.01
FE QMLE	1.06	1.10	1.09	1.07	1.07	1.08	1.07	1.10
FE NLSE	1.15	1.29	1.29	1.18	1.18	1.24	1.18	1.39
Corr. FE NLSE	1.05	1.10	1.12	1.06	1.06	1.10	1.06	1.14
Panel 2: APEs ($\mu_0 = 1$)								
Pooled NLSE	1.00	1.00	1.00	1.00	0.99	1.00	0.99	1.00
Pooled QMLE	1.00	1.00	1.00	1.00	0.99	1.00	0.99	1.00
FE QMLE	1.00	0.99	0.99	0.99	0.99	0.99	0.99	0.99
FE NLSE	1.03	1.06	1.05	1.05	1.05	1.05	1.05	1.06
Corr. FE NLSE	1.01	1.01	0.97	1.02	1.02	1.00	1.02	0.92
# of dropped replications	0	2	3	0	0	0	0	13

[a] Notes: 1000 replications. Pooled estimators are estimated without fixed effects included. If the heterogeneities are drawn from a standard normal distribution, together with the original error term, a new normal distribution with variance 2 is formed. Here, for the convenience to read, all pooled estimates are normalized by multiplying root 2, except those in column 3 and 4, where multiplying by root 1.5 and root 2.5, respectively. Further, those fixed effects estimates in column 3 and 4 are normalized by multiplying root 0.5 and root 1.5, respectively. Similarly, those estimates in column 5 and 6 are divided by the corresponding true parameter values.

Table 3: Estimators of θ_x ($\theta_x^0 = 1$) and μ_x ($\mu_x^0 = 1$) in Dynamic Cases

Estimators	**Structure Parameter $\theta_x(\theta_x^0 = 1)$**					**Marginal Effects $\mu_x(\mu_x^0 = 1)$**				
	Mean	**SD**	**SE/SD**	**P:.05**	**P:.10**	**Mean**	**SD**	**SE/SD**	**P:.05**	**P:.10**
T = 8										
Pooled NLSE	1.13	0.07	1.00	0.39	0.52	1.09	0.04	0.95	0.57	0.68
Pooled QMLE	1.12	0.06	1.01	0.50	0.63	1.08	0.04	0.90	0.68	0.78
Dynamic QMLE	n/a	n/a	n/a	n/a	n/a	0.99	0.04	0.91	0.07	0.13
FE QMLE	1.09	0.06	0.93	0.33	0.44	0.98	0.04	0.80	0.15	0.23
FE NLSE	1.38	0.20	0.48	0.92	0.96	1.06	0.06	0.63	0.33	0.41
Corr. FE NLSE	1.21	0.35	0.26	0.61	0.69	0.91	0.13	0.27	0.56	0.65
T = 12										
Pooled NLSE	1.09	0.06	1.02	0.31	0.45	1.06	0.03	0.96	0.47	0.58
Pooled QMLE	1.08	0.05	1.03	0.41	0.54	1.05	0.03	0.91	0.55	0.64
Dynamic QMLE	n/a	n/a	n/a	n/a	n/a	1.00	0.03	0.97	0.06	0.11
FE QMLE	1.06	0.05	0.94	0.24	0.37	0.99	0.03	0.87	0.10	0.16
FE NLSE	1.17	0.08	0.76	0.72	0.81	1.02	0.04	0.86	0.16	0.24
Corr. FE NLSE	1.07	0.07	0.77	0.24	0.34	0.95	0.05	0.57	0.35	0.45
T = 16										
Pooled NLSE	1.07	0.05	0.97	0.25	0.39	1.04	0.03	0.92	0.35	0.46
Pooled QMLE	1.06	0.04	0.98	0.33	0.46	1.04	0.03	0.89	0.42	0.53
Dynamic QMLE	n/a	n/a	n/a	n/a	n/a	1.00	0.02	0.95	0.07	0.11
FE QMLE	1.04	0.04	0.94	0.20	0.28	1.00	0.02	0.90	0.07	0.14
FE NLSE	1.10	0.06	0.85	0.53	0.65	1.01	0.03	0.94	0.11	0.18
Corr. FE NLSE	1.03	0.05	0.87	0.12	0.19	0.97	0.03	0.75	0.29	0.38
T = 20										
Pooled NLSE	1.05	0.04	0.95	0.25	0.36	1.03	0.03	0.91	0.28	0.38
Pooled QMLE	1.05	0.04	0.95	0.31	0.43	1.03	0.02	0.88	0.35	0.45
Dynamic QMLE	n/a	n/a	n/a	n/a	n/a	1.00	0.02	0.96	0.06	0.11
FE QMLE	1.03	0.04	0.91	0.18	0.27	1.00	0.02	0.91	0.08	0.13
FE NLSE	1.08	0.05	0.86	0.42	0.55	1.01	0.02	0.96	0.10	0.14
Corr. FE NLSE	1.02	0.05	0.87	0.10	0.16	0.97	0.03	0.78	0.24	0.33

[a] Note: Notes: 1000 replications. Pooled estimators and Dynamic QMLE are estimated without fixed effects included. If the heterogeneities are drawn from a standard normal distribution, together with the original error term, a new normal distribution with variance 2 is formed. Here, for the convenience to read, all pooled estimates are normalized by multiplying root 2. Dynamic QMLE also requires an initial condition for the distribution of the cluster heterogeneities, which I assume is a function of the mean of all regressors including the constant but excluding the lag. For finite sample, complete separation of data happens with positive probability. In this case, the objective function is either ∞ or $-\infty$ and the estimation cannot converge. Under this situation, I drop all estimators for this replication. In the simulations, the numbers of dropped replications are 17, 0, 1, and 1, respectively.

Table 4: Estimators of θ_y ($\theta_y^0 = 0.5$) and μ_y ($\mu_y^0 = 1$) in Dynamic Cases

Estimators	**Structure Parameters** $\theta_y(\theta_y^0 = 0.5)$					**Marginal Effects** $\mu_y(\mu_y^0 = 1)$				
	Mean	**SD**	**SE/SD**	**P:.05**	**P:.10**	**Mean**	**SD**	**SE/SD**	**P:.05**	**P:.10**
T = 8										
Pooled NLSE	1.09	0.13	0.93	1.00	1.00	2.11	0.24	0.83	1.00	1.00
Pooled QMLE	1.09	0.12	0.92	1.00	1.00	2.11	0.23	0.79	1.00	1.00
Dynamic QMLE	n/a	n/a	n/a	n/a	n/a	1.33	0.23	0.92	0.35	0.48
FE QMLE	0.19	0.12	0.95	0.72	0.83	0.34	0.28	0.76	0.88	0.93
FE NLSE	0.22	0.19	0.75	0.52	0.60	0.34	0.28	0.76	0.80	0.87
Corr. FE NLSE	0.41	0.17	0.83	0.15	0.22	0.58	0.41	0.48	0.50	0.59
T = 12										
Pooled NLSE	1.03	0.11	0.86	1.00	1.00	2.00	0.21	0.75	1.00	1.00
Pooled QMLE	1.03	0.11	0.84	1.00	1.00	2.01	0.21	0.71	1.00	1.00
Dynamic QMLE	n/a	n/a	n/a	n/a	n/a	1.39	0.19	0.88	0.64	0.73
FE QMLE	0.31	0.09	0.96	0.56	0.68	0.58	0.17	0.93	0.74	0.83
FE NLSE	0.33	0.12	0.87	0.41	0.52	0.57	0.20	0.86	0.68	0.77
Corr. FE NLSE	0.46	0.11	0.95	0.08	0.14	0.82	0.20	0.80	0.25	0.34
T = 16										
Pooled NLSE	1.00	0.10	0.86	1.00	1.00	1.95	0.19	0.74	1.00	1.00
Pooled QMLE	1.00	0.09	0.84	1.00	1.00	1.96	0.18	0.71	1.00	1.00
Dynamic QMLE	n/a	n/a	n/a	n/a	n/a	1.43	0.16	0.87	0.82	0.90
FE QMLE	0.36	0.08	0.98	0.50	0.60	0.68	0.15	0.94	0.62	0.72
FE NLSE	0.37	0.09	0.92	0.36	0.48	0.68	0.16	0.90	0.58	0.69
Corr. FE NLSE	0.48	0.09	0.96	0.07	0.13	0.89	0.17	0.85	0.15	0.23
T = 20										
Pooled NLSE	0.99	0.09	0.80	1.00	1.00	1.94	0.18	0.68	1.00	1.00
Pooled QMLE	0.99	0.09	0.78	1.00	1.00	1.95	0.18	0.64	1.00	1.00
Dynamic QMLE	n/a	n/a	n/a	n/a	n/a	1.47	0.15	0.80	0.93	0.96
FE QMLE	0.39	0.07	0.96	0.37	0.49	0.76	0.13	0.92	0.50	0.63
FE NLSE	0.40	0.08	0.93	0.30	0.39	0.76	0.14	0.90	0.47	0.59
Corr. FE NLSE	0.49	0.08	0.96	0.06	0.11	0.94	0.15	0.86	0.12	0.19

[a] Notes: 1000 replications. Pooled estimators and Dynamic QMLE are estimated without fixed effects included. If the heterogeneities are drawn from a standard normal distribution, together with the original error term, a new normal distribution with variance 2 is formed. Here, for the convenience to read, all pooled estimates are normalized by multiplying root 2. Dynamic QMLE also requires an initial condition for the distribution of the cluster heterogeneities, which I assume is a function of the mean of all regressors including the constant but excluding the lag. For finite sample, complete separation of data happens with positive probability. In this case, the objective function is either ∞ or $-\infty$ and the estimation cannot converge. Under this situation, I drop all estimators for this replication. In the simulations, the numbers of dropped replications are 17, 0, 1, and 1, respectively.

Table 5: Robustness check for static cases with $T = 12$

Scenarios	$\theta_x^0 = 1$			$\mu_x^0 = 1$			$\theta_y^0 = 0.5$			$\mu_y^0 = 1$		
	Mean	P:.05	P.10	Mean	P:.05	P.10	Mean	P:.05	P.10	Mean	P:.05	P.10
Panel 1: N=100												
Pooled NLSE	1.09	0.31	0.45	1.06	0.47	0.58	1.03	1.00	1.00	2.00	1.00	1.00
Pooled QMLE	1.08	0.41	0.54	1.05	0.55	0.64	1.03	1.00	1.00	2.01	1.00	1.00
Dynamic QMLE	n/a	n/a	n/a	1.00	0.06	0.11	n/a	n/a	n/a	1.39	0.64	0.73
FE QMLE	1.06	0.24	0.37	0.99	0.10	0.16	0.31	0.56	0.68	0.58	0.74	0.83
FE NLSE	1.17	0.72	0.81	1.02	0.16	0.24	0.33	0.41	0.52	0.57	0.68	0.77
Corr. FE NLSE	1.07	0.24	0.34	0.95	0.35	0.45	0.46	0.08	0.14	0.82	0.25	0.34
Panel 2: N=200												
Pooled NLSE	1.08	0.53	0.67	1.05	0.69	0.77	1.03	1.00	1.00	2.01	1.00	1.00
Pooled QMLE	1.08	0.66	0.77	1.05	0.78	0.85	1.03	1.00	1.00	2.01	1.00	1.00
Dynamic QMLE	n/a	n/a	n/a	1.00	0.07	0.11	n/a	n/a	n/a	1.40	0.91	0.95
FE QMLE	1.06	0.40	0.51	0.99	0.10	0.17	0.31	0.84	0.90	0.58	0.95	0.97
FE NLSE	1.16	0.93	0.97	1.02	0.21	0.29	0.33	0.66	0.77	0.57	0.90	0.94
Corr. FE NLSE	1.06	0.33	0.44	0.95	0.56	0.65	0.46	0.11	0.18	0.82	0.36	0.46
Panel 3: N=400												
Pooled NLSE	1.08	0.85	0.90	1.06	0.93	0.96	1.03	1.00	1.00	2.02	1.00	1.00
Pooled QMLE	1.08	0.91	0.96	1.05	0.97	0.98	1.03	1.00	1.00	2.02	1.00	1.00
Dynamic QMLE	n/a	n/a	n/a	1.00	0.07	0.14	n/a	n/a	n/a	1.41	0.99	1.00
FE QMLE	1.06	0.66	0.78	0.99	0.14	0.20	0.30	0.99	0.99	0.57	1.00	1.00
FE NLSE	1.16	1.00	1.00	1.02	0.32	0.42	0.33	0.91	0.94	0.57	1.00	1.00
Corr. FE NLSE	1.06	0.53	0.64	0.95	0.80	0.86	0.46	0.13	0.22	0.82	0.56	0.67

[a] Notes: 1000 replications. Pooled estimators and Dynamic QMLE are estimated without fixed effects included. If the heterogeneities are drawn from a standard normal distribution, together with the original error term, a new normal distribution with variance 2 is formed. Here, for the convenience to read, all pooled estimates are normalized by multiplying root 2. Dynamic QMLE also requires an initial condition for the distribution of the cluster heterogeneities, which I assume is a function of the mean of all regressors including the constant but excluding the lag. For finite sample, complete separation of data happens with positive probability. In this case, the objective function is either or and the estimation cannot converge. Under this situation, I drop all estimators for this replication. In the simulations, the numbers of dropped replications are 0, 1, and 2, respectively.

Table 6: Robustness check for dynamic cases with $T = 12$

Scenarios	$\sigma_u^2 = 0.25$ $\sigma_v^2 = 0.75$	$\sigma_u^2 = 0.75$ $\sigma_v^2 = 0.25$	$\sigma_u^2 = 0.25$ $\sigma_v^2 = 0.25$	$\sigma_u^2 = 0.75$ $\sigma_v^2 = 0.75$	$\theta_0 = 0.75$	$\theta_0 = 1.25$	$\rho = 0.05$	$\rho = 0.95$
Panel 1: Parameters of x **(**$\theta_x^0 = 1$**)**								
Pooled NLSE	1.09	1.09	1.17	1.08	1.11	1.08	1.09	1.09
Pooled QMLE	1.08	1.08	1.17	1.08	1.10	1.07	1.08	1.08
FE QMLE	1.04	1.08	1.10	1.05	1.05	1.07	1.06	1.06
FE NLSE	1.12	1.25	1.14	1.13	1.13	1.22	1.16	1.19
Corr. FE NLSE	1.05	1.10	0.98	1.05	1.05	1.08	1.06	1.08
Panel 2: Parameters of lag y **(**$\theta_y^0 = 0.5$**)**								
Pooled NLSE	1.09	0.98	1.08	1.00	3.23	1.68	10.61	1.62
Pooled QMLE	1.09	0.98	1.08	1.00	3.24	1.68	10.63	1.62
FE QMLE	0.33	0.29	0.39	0.25	0.07	0.39	-1.35	0.39
FE NLSE	0.34	0.32	0.44	0.26	0.07	0.43	-1.54	0.43
Corr. FE NLSE	0.46	0.46	0.48	0.45	0.43	0.48	0.18	0.48
# of dropped replications	1	8	14	2	1	6	3	8

[a] Notes: 1000 replications. Pooled estimators are estimated without fixed effects included. If the heterogeneities are drawn from a standard normal distribution, together with the original error term, a new normal distribution with variance 2 is formed. Here, for the convenience to read, all pooled estimates are normalized by multiplying root 2, except those in column 3 and 4, where multiplying by root 1.5 and root 2.5, respectively. Further, those fixed effects estimates in column 3 and 4 are normalized by multiplying root .5 and root 1.5, respectively. For consistency, the parameter estimates for in column 5 and 6 are normalized to 1, similarly, the parameter estimates for in column 5, 6, 7, and 8 are normalized to 0.5.

Table 7: Robustness check for dynamic cases with $T = 12$

Scenarios	$\sigma_y^2 = 0.25$ $\sigma_v^2 = 0.75$	$\sigma_y^2 = 0.75$ $\sigma_v^2 = 0.25$	$\sigma_y^2 = 0.25$ $\sigma_v^2 = 0.25$	$\sigma_y^2 = 0.75$ $\sigma_v^2 = 0.75$	$\theta_0 = 0.75$	$\theta_0 = 1.25$	$\rho = 0.05$	$\rho = 0.95$
Panel 1: APEs of x ($\mu_x^0 = 1$)								
Pooled NLSE	1.06	1.06	1.17	0.97	1.07	1.05	1.06	1.06
Pooled QMLE	1.05	1.06	1.17	0.97	1.07	1.04	1.06	1.05
Dynamic QMLE	1.00	0.99	1.11	0.91	0.99	1.00	1.00	1.00
FE QMLE	0.99	0.99	1.10	0.91	0.99	0.99	1.00	0.99
FE NLSE	1.02	1.04	1.14	0.94	1.02	1.02	1.03	1.02
Corr. FE NLSE	0.96	0.95	0.98	0.89	0.97	0.93	0.95	0.96
Panel 2: APEs lag y ($\mu_y^0 = 1$)								
Pooled NLSE	2.11	1.91	2.28	1.79	3.14	1.63	10.34	1.57
Pooled QMLE	2.11	1.91	2.28	1.80	3.15	1.64	10.38	1.57
Dynamic QMLE	1.43	1.35	1.58	1.25	1.9	1.23	4.53	1.23
FE QMLE	0.62	0.53	0.80	0.44	0.14	0.72	-2.54	0.73
FE NLSE	0.62	0.54	0.80	0.43	0.13	0.73	-2.74	0.74
Corr. FE NLSE	0.84	0.79	0.85	0.77	0.79	0.83	0.27	0.84
# of dropped replications	1	8	14	2	1	6	3	8

[a] Notes: 1000 replications. Pooled estimators are estimated without fixed effects included. If the heterogeneities are drawn from a standard normal distribution, together with the original error term, a new normal distribution with variance 2 is formed. Here, for the convenience to read, all pooled estimates are normalized by multiplying root 2, except those in column 3 and 4, where multiplying by root 1.5 and root 2.5, respectively. Further, those fixed effects estimates in column 3 and 4 are normalized by multiplying root .5 and root 1.5, respectively. Similarly, those estimates in column 5 and 6 are divided by the corresponding true parameter values.

Table 8: Descriptive statistics, 1995-2001, nomial dollars (501 Michgan school districts)

Year	Test Pass Rates in %					Foundation Grant				
	Mean	Median	SD	Minimum	Maximum	Mean	Median	SD	Minimum	Maximum
1992	36.86	36.00	12.79	5.90	86.20	n/a	n/a	n/a	n/a	n/a
1993	43.12	43.10	12.84	7.60	89.00	n/a	n/a	n/a	n/a	n/a
1994	49.25	48.70	13.80	8.30	87.40	n/a	n/a	n/a	n/a	n/a
1995	61.75	62.70	13.05	17.90	98.60	5130.74	4892.00	887.56	4200	10454
1996	62.19	62.70	14.55	15.20	100.00	5347.61	5127.00	844.65	4506.00	10607.00
1997	59.76	61.00	14.82	9.00	97.00	5559.54	5308.00	806.36	4816.00	10762.00
1998	74.88	76.00	12.87	26.00	100.00	5766.74	5642.00	771.29	5170.00	10916.00
1999	74.18	75.50	13.17	25.00	100.00	5766.74	5462.00	771.29	5170.00	10916.00
2000	77.45	79.30	12.26	26.60	100.00	6057.42	5700.00	712.43	5700.00	11091.00
2001	75.54	77.40	12.62	21.00	100.00	6348.50	6000.00	689.09	6000.00	11335.00

Table 9: Estimation results (n=501, T=6), static model

Variable	Pooled Probit	Pooled NLSE	RE Probit	RE NLSE	FE Probit	FE NLSE	Corr. FE NLSE
Panel 1: Index Coefficients							
log(arexppp)	0.333***	0.301***	0.881***	0.905***	0.987***	0.967***	0.972***
	(0.048)	(0.050)	(0.213)	(0.220)	(0.152)	(0.158)	(0.158)
lunch	-1.190***	-1.180***	-0.219	-0.143	-0.251	-0.277	-0.269
	(0.038)	(0.039)	(0.238)	(0.243)	(0.166)	(0.171)	(0.171)
log(enroll)	-0.001	-0.003	0.089	0.123	0.080	0.073	0.071
	(0.008)	(0.008)	(0.154)	(0.158)	(0.112)	(0.118)	(0.118)
Panel 2: Marginal Effects							
log(arexppp)	0.112***	0.101***	0.297	0.305	0.321**	0.315**	0.317**
	(0.007)	(0.005)	(0.199)	(0.213)	(0.152)	(0.155)	(0.156)
lunch	-0.401***	-0.398***	-0.074	-0.048	-0.082	-0.090	-0.088
	(0.015)	(0.015)	(0.082)	(0.084)	(0.056)	(0.058)	(0.058)
log(enroll)	-0.000	-0.001	0.030	0.042	0.026	0.024	0.023
	(0.003)	(0.003)	(0.035)	(0.029)	(0.026)	(0.028)	(0.029)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District dummies	No	No	No	No	Yes	Yes	Yes

[a] Note: Pooled denotes the specification that does not control for school district unobservables; RE denotes the specification that specifies the conditional distribution of school district unobservables explicitly; while FE denotes the specification that controls for school district unobservables with fixed effects. Probit denotes that the specification applies Quasi-MLE for estimation, while NLSE denotes that the specification applies nonlinear least square estimation method. Corr. FE NLSE denotes the bias corrected estimator based on the counterpart, FE NLSE. For the marginal effects in the specification of RE Probit, here I report the asymptotic standand errors for the convenience of comparison with other specifications. Papke and Wooldridge (2008) report the bootstrapping standard errors and I replicate these standard errors when doing bootstrapping.

References

Arellano, Manuel and Jinyong Hahn, "Understanding Bias In Nonlinear Panel Models: Some Recent Developments," Working Papers, CEMFI October 2005.

Bester, C. Alan and Christian Hansen, "A Penalty Function Approach to Bias Reduction in Nonlinear Panel Models with Fixed Effects," *Journal of Business & Economic Statistics*, 2009, *27* (2), 131–148.

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

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

— , "Panel data," in Z. Griliches and M. D. Intriligator, eds., *Handbook of Econometrics*, Vol. 2 of *Handbook of Econometrics*, Elsevier, March 1984, chapter 22, pp. 1247–1318.

Fernandez-Val, Ivan, "Fixed effects estimation of structural parameters and marginal effects in panel probit models," *Journal of Econometrics*, May 2009, *150* (1), 71–85.

Gourieroux, Christian, Alain Monfort, and Alain Trognon, "Pseudo Maximum Likelihood Methods: Theory," *Econometrica*, May 1984, *52* (3), 681–700.

Greene, William, "The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects," *Econometrics Journal*, June 2004, *7* (1), 98–119.

Hahn, Jinyong and Guido Kuersteiner, "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects when Both "n" and "T" Are Large," *Econometrica*, July 2002, *70* (4), 1639–1657.

— **and** — , "Bias reduction for dynamic nonlinear panel models with fixed effects," Working Papers, UC Davis, Department of Economics November 2007.

_ **and Hyungsik Roger Moon**, "Reducing Bias Of Mle In A Dynamic Panel Model," *Econometric Theory*, June 2006, *22* (03), 499–512.

_ **and Whitney Newey**, "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica*, July 2004, *72* (4), 1295–1319.

Heckman, James J., "The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process," in Daniel L. McFadden and Charles F. Manski, eds., *Structural Analysis of Discrete Panel Data with Econometric Applications*, Cambridge: MIT Press, 1981, chapter 4, pp. 179–195.

Honor, Bo E. and Ekaterini Kyriazidou, "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, July 2000, *68* (4), 839–874.

Li, Xiaoming, AKM Rezaul Hossain, and Stephen L. Ross, "Neighborhood Information Externalities and the Provision of Mortgage Credit," Working papers 2010-10, University of Connecticut, Department of Economics May 2010.

Loudermilk, Margaret S., "Estimation of Fractional Dependent Variables in Dynamic Panel Data Models With an Application to Firm Dividend Policy," *Journal of Business & Economic Statistics*, October 2007, *25*, 462–472.

Manski, Charles F, "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, March 1987, *55* (2), 357–62.

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

Neyman, Jerzy and Elizabeth L. Scott, "Consistent Estimates Based on Partially Consistent Observations," *Econometrica*, January 1948, *16* (1), 1–32.

Nickell, Stephen J, "Biases in Dynamic Models with Fixed Effects," *Econometrica*, November 1981, *49* (6), 1417–26.

Papke, Leslie E and Jeffrey M Wooldridge, "Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates," *Journal of Applied Econometrics*, Nov.-Dec. 1996, *11* (6), 619–32.

Papke, Leslie E. and Jeffrey M. Wooldridge, "Panel data methods for fractional response variables with an application to test pass rates," *Journal of Econometrics*, July 2008, *145* (1-2), 121–133.

Wooldridge, Jeffrey M, *Econometric Analysis of Cross-Section and Panel Data*, Cambridge, MA: MIT Press, 2002.

Wooldridge, Jeffrey M., "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity," *Journal of Applied Econometrics*, 2005, *20* (1), 39–54.