KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE
OF
ECONOMIC RESEARCH

Discussion Paper No.886

"Deriving the Information Bounds for Nonlinear Panel Data Models with Fixed Effects"

Haruo Iwakura

January 2014

KYOTO UNIVERSITY

KYOTO, JAPAN

Deriving the Information Bounds for Nonlinear Panel Data Models with Fixed Effects

Haruo Iwakura*

January 27, 2014

Abstract

This paper studies the asymptotic efficiency of estimates in nonlinear panel data models with fixed effects when both the cross-sectional sample size and the length of time series tend to infinity. The efficiency bounds for regular estimators are derived using the infinite-dimensional convolution theorem by van der Varrt and Wellner (1996). It should be noted that the number of fixed effects increases with the sample size, so they constitute an infinite-dimensional nuisance parameter. The presence of fixed effects makes our derivation of the efficiency bounds non-trivial, and the techniques to overcome the difficulties caused by fixed effects will be discussed in detail. Our results include the efficiency bounds for models containing unknown functions (for instance, a distribution function of error terms). We apply our results to show that the bias-corrected fixed effects estimator of Hahn and Newey (2004) is asymptotically efficient.

Keywords: asymptotic efficiency; convolution theorem; double asymptotics; nonlinear panel data model; fixed effects; interactive effects; factor structure; incidental parameters.

JEL classification: C13; C23.

1 introduction

Some of a recent literature on nonlinear panel data analysis focuses on the so-called large N and large T asymptotics (N is a cross-sectional dimension and T is a time dimension). For example, Hahn and Newey (2004) study the asymptotic properties of the fixed effect estimator in a general nonlinear panel data model with individual effects. They show that the fixed effects estimator has a limiting normal distribution with a bias in the mean when N and T grow at the same rate. The non-central mean captures an incidental parameter bias of order $1/T$ and their proposed bias correction methods are shown to substantially reduce the bias without increasing a variance, even when T is only moderately large. Hahn and Kuersteiner (2011) extend the results of Hahn and Newey to a dynamic case and show that their bias-corrected estimators are asymptotically normal with mean zero. For other studies in this line of research, see Woutersen (2002), Carro (2007), Fernandez-Val (2009), Bester and Hansen (2009),

*Graduate School of Economics, Kyoto University. Yoshida-Hommachi, Sakyo, Kyoto, Kyoto, 606-8501, Japan. Email: iwakura.haruo@ht8.ecs.kyoto-u.ac.jp

and Fernandez-Val and Vella (2011), among others. The large N and large T asymptotic framework is also employed in a recent literature on panel quantile regression models. See Koenker (2004), Canay (2008) and Kato et al. (2012), among others. When we consider estimation, efficiency is an important issue. The efficiency bound for estimation is useful when we quantify the efficiency loss of particular estimators we use in applications. Moreover, the bound can give us a valuable guidance for constructing efficient estimators and calculating the limiting distributions (see, e.g., Newey (1990) and Bickel et al. (1993)). However, to our knowledge, there are no efficiency results available that can be applied to general nonlinear panel data models under large N and large T asymptotics. Hahn and Kuersteiner (2002) and Bai (2012) analyze the efficiency of panel data estimators but only in the context of linear panel autoregressive models.

The objective of the present paper is to establish asymptotic optimality theory of estimation in a general nonlinear panel data model with fixed effects when both N and T tend to infinity. We consider a general semiparametric model where the law of the data is characterized by a finite-dimensional parameter of interest, individual specific effects and an unknown nuisance function. Our specification is general enough to accommodate many important nonlinear panel data models. The simplest case is a parametric conditional density model such as logit, probit and Tobit models. Our model also includes many important semiparamtric models as special cases. Semiparametric binary choice and censored regression models are a few of the examples.

In this general setting, the present paper contributes to the panel data literature by providing a general formula of the information bound for the model parameters. Furthermore, to illustrate the usefulness of our bound, we apply our efficiency results to assess the efficiency of estimators existing in the literature. In particular, we consider the setting of Hahn and Newey (2004) and show that their bias-corrected fixed effect estimators are asymptotically efficient.

In order to derive the efficiency bound, we use the infinite-dimensional convolution theorem by van der Vaart and Wellner (1996). Notice that since there are as many individual effects as the cross-sctional sample size, the number of the fixed effects tend to infinity. Thus our model has two infinite-dimensional nuisance parameters: the first one is a sequence of individual fixed effects and the second one is an unknown function. The second component can be dealt with by similar arguments to semiprametric efficiency for models with i.i.d. observations. On the other hand, the presence of the fixed effects makes the derivation nontrivial and somewhat non-standard, in particular, in the verification of local asymptotic normality (LAN).

Showing the LAN property is an inevitable step in obtaining the efficiency bound. The main difficulty is how to perturb the sequence of the fixed effects. Hahn and Kuersteiner (2002) examine efficiency of estimates in panel autoregressive models with fixed effects. They perturb the fixed effects by another sequence and establish the local asymptotic normality under their perturbation scheme. However, their method cannot be easily applied to a general nonlinear model, as we will discuss in detail in Section 4. In order to overcome the technical difficulties caused by the presence of fixed effects, we take a different approach. We use the idea originated in the statistical literature on "functional model" (see, e.g., Phanzagl (1993)).[1] More specifically, we adapt the approach by Strasser (1998) to our nonlinear panel data setting. Strasser's idea is to perturb individual specific parameters using a continuous function, rather than a real sequence. We extend his method and establish the local asymptotic normality of our nonlinear panel data model with fixed effects. The difference from Strasser (1996) is that we consider a model with a nuisance function, while Strasser (1996) only considers a model without a nuisance function. Another important difference is that Strasser' (1998) considers the fixed T asymptotics, whereas we consider the large N and large T asymptotics.

The rest of the paper is organized as follows. The next section formally introduces the model we examine in this paper. Section 3 presents the general formula of the efficiency bound, which is the main result of the present paper. This section also provides an application of our efficiency results. In particular, we show that the bias-corrected fixed effect estimator of Hahn and Newey (2004) is asymptotically efficient. In Section 4, we give a detailed explanation of the derivation of the bounds. Section 5 concludes the paper. Most of the mathematical proofs are given in the Appendix.

2 The Ste-up

This section introduces the set-up that we examine in this paper and presents some of the examples our model accommodates.

Suppose that we have available a panel data set $\{\{z_{it}\}_{t=1}^{T}\}_{i=1}^{N}$ where z_{it} is a random vector with support $\mathcal{Z} \subseteq \mathbb{R}^{d_z}$. Let $\theta \in \Theta \subseteq \mathbb{R}^{d_\theta}$ be an unknown finite-dimensional parameter of interest and let $\eta_i \in \Lambda \subseteq \mathbb{R}$ be an unobserved individual effect.[2] We assume that θ and η_i are interior points of Θ and Λ. Let $q(\cdot)$ be an unknown measurable function with values in $\mathbb{R}^{d_q}$ We denote the space for q by Q. Each observation z_{it} of individual i at period t is distributed according to a probability measure $\mathbb{P}_{\theta,\eta_i,q}$. The family of probability measures $\{\mathbb{P}_{\theta,\eta,q} : \theta \in \Theta, \eta \in \Lambda, q \in Q\}$ is assumed to be dominated by a σ-finite measure μ with densities $f(z|\theta, \eta, q)$. As in Hahn and Newey (2004), we assume that the observations are independently distributed across both i and t. The goal of this paper is to present the information bound for estimating θ when both N and T are large. Deriving the efficiency bound for a more general parameter such as a functional of laws $\{\mathbb{P}_{\theta,\eta_i,q} : i \in \mathbb{N}\}$ is also possible. However, such an extension requires more complicated notation and additional technical terminology, which may blur the main theme of the present paper.

As is common in the large N and large T panel data literature, we will treat the unobserved individual effects η_i's as fixed constants, rather than random variables (see, e.g., Hahn and Newey (2004, p.1297), Fernandez-Val (2009) and Hahn and Kuersteiner (2012)). This situation can be interpreted as we conduct an inference conditional on the realization for random η_i's. In other words, we consider the situation in which the data is generated in the following scheme. First, the individual effects $\eta_1, \eta_2, \cdots$ are generated according to some distribution and, then, conditional on the realization for η_i's (i.e., treating η_i's as fixed), the observations z_{it} are sampled from the densities $f(z_{it}|\theta, \eta_i, q)$. Under this interpretation, the

[1] Roughly speaking, a functional model is a model that contains non-random individual specific parameters. This class of model is similar to a mixture model, but a mixture model treats heterogeneous components as random variables.

[2] An extension to the case where η_i is vector-valued is straightforward.

density $f(z_{it}|\theta, \eta_i, q)$ should be interpreted as the conditional density of z_{it} given η_i.

When the individual effects are treated as fixed constants, they become individual specific parameters. Note that since the number of the individual effects increases with the cross sectional sample size, there are infinitely many individual specific parameters $\eta_1, \eta_2, \cdots$. We denote the space for a sequence $\{\eta_i\}_{i=1}^{\infty}$ of individual effects by $W(\Lambda)$ and define $\xi := (\theta, \{\eta_i\}_{i=1}^{\infty}, q)$ and $\Xi := \Theta \times W(\Lambda) \times Q$. Notice that our model has two infinite-dimensional nuisance parameters: the first one is the function q and the second one is the sequence of the fixed effects, $\{\eta_i\}_{i=1}^{\infty}$. When deriving the efficiency bounds, the first nuisance can be perturbed by constructing parametric submodels as in the usual semiparametric efficiency arguments (see, e.g., Bickel et al. (1990) and van der Vaart (1998)). On the other hand, it will turn out that perturbation of the incidental parameters $\eta_1, \eta_2, \cdots$ is technically delicate and makes the derivation somewhat nonstandard, in particular, in the verification of local asymptotic normality. We will discuss the details in Section 4.

The model above includes many important panel data models as special cases. The simplest case may be the following conditional density model.

Example 1. Let $z_{it} = (y_{it}, x_{it}')'$ where y_{it} is a response variable and x_{it} is a vector of covariates. Suppose that $f(z_{it}|\theta, \eta_i, q)d\mu(z_{it})$ is factored into the conditional density $f_{Y|X}(y_{it}|x_{it}, \theta, \eta_i)d\mu_Y(y_{it})$ of y_{it} given x_{it} and the marginal density $f_X(x_{it}|\eta_i)d\mu_X(x_{it})$ of x_{it} ($\mu = \mu_Y \otimes \mu_X$). Here the conditional density $f_{Y|X}$ is known up to parameters (θ, η_i), whereas the form of f_X is completely unknown except that it may depend on η_i. In this case, f_X corresponds to q in a general model. It is obvious that this specification includes panel probit, logit and Tobit models as special cases. In the econometric analysis, allowing the marginal $f_X(x_{it}|\eta_i)$ to depend on $\eta_i$is important. If the individual effect η_i were treated as random, it would be desirable to allow the covariate x_{it} to be correlated with η_i. In the fixed effects situation where η_i's are treated as fixed constants, such a 'correlation' should be modeled as the dependence of the marginal density $f_X(x_{it}|\eta_i)$ on the individual specific parameter η_i.

Our model also includes many important semiparametric models. The following illustrates a few of the examples

Example 2. Let $z_{it} = (y_{it}, x_{it}')'$where y_{it} is a response variable and x_{it} is a vector of covariates. Let $y_{it} = m(\eta_i + \theta' x_{it} + u_{it})$ where m is a known real function and u_{it} is a disturbance term whose distribution is unknown. Assume that u_{it} is i.i.d. across i and t and is independent of $\{x_{it}\}$. Obviously, this specification fits into our present framework with the joiny density of (u_{it}, x_{it}) playing the role of q in a general setting. The case where $m(x) = 1\{x > 0\}$ corresponds to a semiparamtric binary choice model, while the case where $m(x) = \max\{x, 0\}$ corresponds to a semiparamtric censored regression model.

3 The Main Results

The objective of this section is to present the efficiency bound for estimating θ. Since the derivation of the information bound requires the LAN theory and is somewhat technical, we defer the details of the derivation to the next section. This section consists of three subsections. The first subsection gives the assumptions we need. The second subsection provides the general formula for the efficiency bound, which is the main result of the present paper. In the last subsection, to illustrate the usefulness of our formula, we apply it to the setting of Hahn and Newey (2004) and show that their bias-corrected fixed effects estimator is asymptotically efficient.

3.1 Assumptions

In this subsection, we give the assumptions that are needed for the theoretical derivation of the efficiency bounds. We begin by summarizing the set-up described in the preceding section.

Assumption 1. (i) z_{it} is observed for $i = 1, 2, \cdots, N$ and $t = 1, 2, \cdots, T$. (ii) The observations are independent across both i and t. (iii) Each observation z_{it} is distributed according to the density $f(z_{it}|\theta, \eta_i, q)$ for $\theta \in \text{int}\Theta$, $\eta_i \in \text{int}\Lambda$ and $q \in Q$.

Next, we impose the following mild condition on the density $f(z|\theta, \eta, q)$.

Assumption 2. Suppose that Λ is a bounded closed interval in $\mathbb{R}$. Fix $q \in Q$. The mapping $(\theta, \eta) \mapsto \sqrt{f(z|\theta, \eta, q)}$ is continuously Frechet differentiable on $\text{int}\Theta \times \Lambda$ in $L^2(\mu)$.[3]

Assumption 2 is standard in the efficiency literature and is also called a quadratic mean differentiability assumption. To see the implications of this assumption, now fix $\theta \in \text{int}\Theta$ and $q \in Q$. The Frechet differentiability implies that, for every $\eta \in \Lambda$, there exists a $d_\theta + 1$-dimensional vector of measurable functions, $\dot{\ell}(z|\eta) := (\dot{\ell}_1(z|\eta)', \dot{\ell}_2(z|\eta))'$, such that

$$\int \Big[\sqrt{f(z|\theta + u_1, \eta + u_2, q)} - \sqrt{f(z|\theta, \eta, q)} \\ - \frac{1}{2} u' \dot{\ell}(z|\eta) \sqrt{f(z|\theta, \eta, q)}\Big]^2 d\mu(z) = o(\|u\|_E^2) \tag{3.1}$$

as $u \to 0$ where $u := (u_1', u_2)' \in \mathbb{R}^{d_\theta + 1}$ and $\|\cdot\|_E$ is the Euclidean norm. The functions $\dot{\ell}_1$ and $\dot{\ell}_2$ are considered to be the score functions for θ and η, respectively, when q is regarded to be known. The function $\dot{\ell}(z|\eta)$ may depend on θ and $q \in Q$, but we suppress such dependence for simplicity of notation. The continuity of the Frechet derivative implies, in particular, that the mapping $\eta \mapsto \mathbb{E}_\eta \dot{\ell}(z|\eta)\dot{\ell}(z|\eta)'$ is continuous and bounded on Λ where $\mathbb{E}_\eta$ denotes the expectation under $f(z|\theta, \eta, q)$.[4] Assumption 2 is a very mild smoothness condition on the density $f(z|\theta, \eta, q)$. A simple sufficient condition for Assumption 2 is that the density $f(z|\theta, \eta, q)$ is continuously differentiable (in the usual sense) with respect to (θ, η) for μ-almost every z and the Fisher information (when q is considered to be known) is continuous with respect to (θ, η) (see, e.g., Proposition 2.1.1 in Bickel et al. (1993)).

Based on this assumption, we now introduce the tangent set for the nuisance function q, a notion needed for the formulation of the efficiency bounds. To this end, let $\dot{l}(z|\eta)$ be any measurable function such that there exits a one dimensional parametrization $\{q_s : s \in (-\epsilon, \epsilon)\}$ in Q with the following properties: (i) q_s passes through q at $s = 0$, (ii) for every $\eta \in \Lambda$, the mapping $\sqrt{f(\cdot|\theta, \eta, q_s(\cdot, \eta))}$ from

[3] When η is on the boundary of Λ, the convergence $u = (u_1', u_2)'$ to 0 in (3.1) should be read as $u \to 0$ *with* $u_2 > 0$ for the left end-point case or with $u_2 < 0$ for the right-end point case.

[4] We again suppress the dependence on θ and q for notational convenience.

int$\Theta \times \Lambda \times (-\epsilon, \epsilon)$ to $L^2(\mu)$ is continuously Frechet differentiable at $(\theta, \eta, 0)$ with $(\dot{\ell}', \dot{l})'$ being its Frechet derivative. The requirements (i) and (ii) implies that

$$\int \Big[\sqrt{f(z|\theta + u_1, \eta + u_2, q_s)} - \sqrt{f(z|\theta, \eta, q)} \\ - \frac{1}{2}(u'\dot{\ell}(z|\eta) + s\dot{l}(z|\eta))\sqrt{f(z|\theta, \eta, q)}\Big]^2 d\mu(z) = o(\|u\|^2 + s^2) \tag{3.2}$$

as $u \to 0$ and $s \to 0$, and that the mapping $\eta \mapsto \mathbb{E}_\eta \dot{l}(z|\eta)^2$ is continuous and bounded on Λ. The function $\dot{l}$ is interpreted as a score function for the 'parameter' s when the density is parametrized as $f(z|\theta, \eta, q_s(\cdot, \eta))$. We denote the set of all possible functions $\dot{l}$ by $\mathcal{T}_q$. We assume the following mild assumption on $\mathcal{T}_q$.

Assumption 3. $\mathcal{T}_q$ is linear.

The second assumption that we will need is concerned with the behavior of the sequence of individual effects, $\{\eta_i\}_{i=1}^\infty$. As we shall see, the following assumption on individual effects is *both* mathematically convenient and econometrically interpretable.

Assumption 4. A real sequence of individual effects $\{\eta_i\}_{i=1}^\infty$ is *weakly convergent*[5] in the following sense: there exists a probability measure Γ on Λ such that, as $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^N \psi(\eta_i) \to \int \psi(s) d\Gamma(s) \text{ for all } \psi \in C_b(\mathbb{R}) \tag{3.3}$$

where $C_b(\Lambda)$ denotes a set of bounded and continuous real-valued functions on Λ. Consequently, the set $W(\Lambda)$ is the collection of weakly convergent sequences.

This type of assumption on individual effects is occasionally used in the statistics literature on functional models (see, e.g., Bickel and Klassen (1986), Phanzagl (1993), Strasser (1996) and Strasser (1998)). Assumption 4 is a key assumption of this paper and plays an important role in overcoming the mathematical difficulties caused by the presence of individual effects. In particular, the assumption is useful and crucial in our verification of local asymptotic normality, as we will discuss in detail in the next section. To get a glimpse of how well this assumption works in our efficiency analysis, notice that the assumption implies that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \mathbb{E}_{\eta_i}\dot{\ell}_1(z|\eta_i)\ell_1(z|\eta_i)' = \int \mathbb{E}_\eta \dot{\ell}_1(z|\eta)\ell_1(z|\eta)' d\Gamma \tag{3.4}$$

(recall that the mapping $\eta \mapsto \mathbb{E}_\eta \dot{\ell}_1(z|\eta)\ell_1(z|\eta)'$ is continuous on Λ by Assumption 2). As we shall see, this limit in (3.4) appears as part of the the efficiency bound for θ. The point here is that under Assumption 2, the Cesaro limit on the left can be written as the *integral* with respect to Γ. In efficiency analysis for models with an infinite-dimensional parameter, functional analysis is an essential tool. Notice that integrals are much more tractable in functional analysis than Cesaro limits. This is one of the mathematical advantages of Assumption 4. However, the assumption is not only mathematically convenient but also econometrically interpretable, as the following argument shows.

[5] Let $\Gamma_N = \frac{1}{N}\sum_{i=1}^N \delta_{\eta_i}$ where δ_x is a Dirac measure on Λ. Noting that Γ_N defines a probability measure on Λ, we can easily see that the 'weak convergence' of $\{\eta_i\}_{i=1}^\infty$ in the above assumption is equivalent to a weak convergence of probability measures $\{\Gamma_N\}_{N=1}^\infty$ in the usual sense.

A trivial example of weakly convergent sequences is a constant sequence, while a well-known nontrivial example may be the Wyle sequence. However, examples of weakly convergent sequences are not restricted to them. One general method to construct a weakly convergent sequence is to use the following result from the number-theoretic ergodic theory.

Lemma 3.1. (A version of Theorem in Furstenberg (1981)) Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\{X_i\}_{i=1}^{\infty}$ be an ergodic stationary sequence of random elements taking values in a metric space (E, ρ) eondowed with its Borel σ-field. Suppose that the function space $C_b(E)$ is separable under the sup metric. Then, for $\mathbb{P}$-almost every $\omega \in \Omega$, we have, as $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^{N} \psi(X_i(\omega)) \to \mathbb{E}\psi(X_i) \text{ for all } \psi \in C_b(E). \tag{3.5}$$

Note that Lemma 3.1 is slightly different from Birkhoff's ergodic theorem. For details, see Remark 3.1 below. Since Λ is assumed to be a bounded closed interval, the space $C_b(\Lambda)$ is separable (see, e.g., Corollary 11.2.5 in Dudley (2002)). Thus, it follows form Lemma 3.1 that a sequence of individual effects that satisfies Assumption 4 can be generally constructed as almost sure realizations of ergodic stationary processes. In view of this construction, Assumption 4 is seen to be a mild requirement on individual effects since the class of weakly convergent sequences includes real sequences that behave like a realization of random variables.

Remark 3.1. As a technical matter, notice that without the separability condition on $C_b(E)$, the result in Lemma 3.1 may not be valid. Without the separability of $C_b(E)$, Birkhoff's ergodic theorem only implies that, for each $\psi \in C_b(E)$, there exists an event Ω_ψ with probability one such that, for $\omega \in \Omega_\psi$, $(1/N)\sum_{i=1}^{N} \psi(X_i(\omega)) \to \mathbb{E}\psi(X_i)$. The point here is that the almost sure event Ω_ψ can be different across ψ. The key feature of Lemma 3.1 is that if $C_b(E)$ is *separable*, then we can take an almost sure event Ω_0 on which the convergence $(1/N)\sum_{i=1}^{N} \psi(X_i(\omega)) \to \mathbb{E}\psi(X_i)$ holds for *all* ψ (i.e. the almost sure event Ω_0 does not depend on ψ).

3.2 The Efficiency Bound

In this subsection, we present the lower bound of the asymptotic variances of estimators of θ. As is common in the statistics and econometrics literatures, we use a convolution theorem to derive the efficiency bound. A convolution theorem is a representation theorem that states that the asymptotic distribution of any regular estimator can be written as a convolution of a certain normal distribution and some noise factor. Because a convolution increases the variance, the convolution theorem implies that the normal distribution in the convolution is optimal in terms of asymptotic variances.

The convolution theorem restricts its attention to the class of regular estimators in order to avoid the well-known problem of superefficiency (see, e.g., van der Varrt 1998). Because of its inherently technical nature, we defer the precise definition of a regular estimator to the next section. Roughly speaking, regularity is a local uniformity requirement on the asymptotic behavior of an estimator. A regular estimator can be considered as desirable because the distribution of a regular estimator is not too sensitive to a small change in the parameter. Moreover, regularity is not very restrictive since regularity is typically much weaker than uniform convergence in distribution of an estimator over a small neighborhood of the true parameter value (see, e.g., Bickel et al. (1993)).

To state the main result, it is useful to introduce some additional notation. Let $\mathbb{P}^{\Gamma}_{\theta,q}$ be the mixture distribution of $\mathbb{P}_{\theta,\eta,q}$ with Γ. Let $\Pi_{\mathcal{T}}$ be a projection operator onto the closure of

$$\mathcal{T} := \left\{ \tilde{\eta}\dot{\ell}_2 + \dot{l} \in L^2(\mathbb{P}^{\Gamma}_{\theta,q}) : \tilde{\eta} \in C_b(\Lambda),\ \dot{l} \in \mathcal{T}_q \right\}. \tag{3.6}$$

Note that we regard the set $\mathcal{T}$ as a subspace of $L^2(\mathbb{P}^{\Gamma}_{\theta,q})$ and thus we take a closure with respect to the $L^2(\mathbb{P}^{\Gamma}_{\theta,q})$-norm. The set $\mathcal{T}$ can be interpreted as the tangent set for the nuisance parameters η_i's and q.

Theorem 3.2 (A convolution theorem)**.** Suppose that Assumptions 1 to 4 hold. Assume that the map $\eta \mapsto \mathbb{E}_\eta \Pi_{\mathcal{T}} \dot{\ell}_1(z|\eta)^2$ is continuous on Λ. Define

$$V_\xi := \int \mathbb{E}_\eta \left(\dot{\ell}_1(z|\eta) - \Pi_{\mathcal{T}}\dot{\ell}_1(z|\eta) \right) \left(\dot{\ell}_1(z|\eta) - \Pi_{\mathcal{T}}\dot{\ell}_1(z|\eta) \right)' d\Gamma(\eta) \tag{3.7}$$

and suppose that V_ξ is nonsingular.[6] Let τ_{NT} be any regular estimator of θ as $N, T \to \infty$. Then

$$\sqrt{NT}(\tau_{NT} - \theta) \xrightarrow{d} N(0, V_\xi^{-1}) * W_\xi \tag{3.10}$$

for some distribution W on $\mathbb{R}^p$ where $*$ denotes a convolution operator. Further, if the limit law of τ_{NT} has variance matrix Σ_ξ, then $\Sigma_\xi \geq V_\xi^{-1}$ in the matrix sense.

This theorem shows that the asymptotic distribution of any regular estimator of θ can be written as a convolution of the normal distribution $N(0, V_\xi^{-1})$ and some noise factor W_ξ. It implies that the law $N(0, V_\xi^{-1})$ is optimal in the sense that the asymptotic variances of any regular estimators cannot be smaller than V_ξ^{-1}. We refer to the matrix V_ξ as the efficient information matrix for θ. We say that an estimator sequence τ_{NT} is asymptotically efficient (or, simply, efficient) if it is regular and asymptotically normal with mean zero and variance matrix V_ξ^{-1}.

The functions $\dot{\ell}_1(z|\eta) - \Pi_{\mathcal{T}}\dot{\ell}_1(z|\eta)$ corresponds to an efficient score function in the usual semiparametric efficiency arguments for i.i.d. models. It can be shown that if an estimator τ_{NT} admits an expansion

$$\sqrt{NT}(\tau_{NT} - \theta) = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} V_\xi^{-1} \left(\dot{\ell}_1(z_{it}|\theta, \eta_i, q) - \Pi_{\mathcal{T}}\dot{\ell}_1(z_{it}|\theta, \eta_i, q) \right) + o_p(1) \tag{3.11}$$

as $N, T \to \infty$, then the estimator is efficient, i.e., it is regular and asymptotically normal with mean zero and variance matrix V_ξ^{-1}.

Although the projection $\Pi_{\mathcal{T}}\dot{\ell}_1$ is defined as the projection onto $\mathcal{T} \subseteq L^2(\mathbb{P}^{\Gamma}_{\theta,q})$, it can be heuristically calculated in the following way: we first fix η and then project the function $z \mapsto \dot{\ell}_1(z|\eta)$ onto the closure of

$$\mathcal{T}_\eta := \left\{ \tilde{\eta}(\eta)\dot{\ell}_2(\cdot|\eta) + \dot{l}(\cdot|\eta) \in L^2(\mathbb{P}_{\theta,\eta,q}) : \tilde{\eta} \in C_b(\Lambda),\ \dot{l} \in \mathcal{T}_q \right\} \subseteq L^2(\mathbb{P}_{\theta,\eta,q}). \tag{3.12}$$

[6]Recall that we have suppressed the dependence of $\dot{\ell}_1$ and $\mathbb{E}_\eta$ on θ and q. If we make explicit such dependence, then

$$\begin{aligned} V_\xi &= \int \mathbb{E}_{\theta,\eta,q} \left(\dot{\ell}_1(z|\theta,\eta,q) - \Pi_{\mathcal{T}}\dot{\ell}_1(z|\theta,\eta,q) \right) \left(\dot{\ell}_1(z|\theta,\eta,q) - \Pi_{\mathcal{T}}\dot{\ell}_1(z|\theta,\eta,q) \right)' d\Gamma(\eta) && (3.8) \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_{\theta,\eta_i,q} \left(\dot{\ell}_1(z_{it}|\theta,\eta_i,q) - \Pi_{\mathcal{T}}\dot{\ell}_1(z_{it}|\theta,\eta_i,q) \right) \left(\dot{\ell}_1(z_{it}|\theta,\eta_i,q) - \Pi_{\mathcal{T}}\dot{\ell}_1(z_{it}|\theta,\eta_i,q) \right)'. && (3.9) \end{aligned}$$

This makes clear the reason for the subscript ξ in V_ξ.

The sets $\mathcal{T}$ and $\mathcal{T}_\eta$ look similar but differ in that the former is the subset of $L^2(\mathbb{P}_{\theta,\Gamma,q})$, while the latter is the subset of $L^2(\mathbb{P}_{\theta,\eta,q})$ for a fixed η. When we interpret the density $f(z|\theta,\eta,q)$ as the conditional density of z given random η, the projection onto $\mathcal{T}_\eta$ can be regarded as a projection 'conditional on η'.

We now give an interpretation of the efficient information matrix V_ξ. Notice that by Assumption 4 we can write

$$V_\xi = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_{\eta_i} \left(\dot{\ell}_1(z_{it}|\eta_i) - \Pi_{\mathcal{T}} \dot{\ell}_1(z_{it}|\eta_i) \right) \left(\dot{\ell}_1(z_{it}|\eta_i) - \Pi_{\mathcal{T}} \dot{\ell}_1(z_{it}|\eta_i) \right)'. \tag{3.13}$$

In view of this, each summand in the last display can be interpreted as the efficient information for θ *when we only use the data for individual i*. Thus V_ξ can be interpreted as the 'average' efficient information across individuals.

Remark 3.2. One possible drawback of a convolution theorem is that it is only valid for the prescribed set of regular estimators. A local asymptotic minimax theorem, which is also popular in the efficiency literature (see, e.g., Chamberlain (1992)), complements this drawback. Specifically, using a local asymptotic minimax theorem (van der Vart and Wellner (1996, Theorem 3.11.5)), we can show that the limit law $N(0, V_\xi^{-1})$ is optimal among *all* estimators in terms of the maximum risk over an arbitrarily small 'neighborhood' of the true parameter value. To be more precise, define a neighborhood of the true parameter $\xi = (\theta, \{\eta_i\}_{i=1}^\infty, q)$ by

$$B_\xi(\delta) := \left\{ \check{\xi} = (\check{\theta}, \{\check{\eta}_i\}_{i=1}^\infty, \check{q}) \in \Theta \times W(\Lambda) \times Q : \sup_i \|\mathbb{P}_{\theta,\eta_i,q} - \mathbb{P}_{\check{\theta},\check{\eta}_i,\check{q}}\|_{TV} < \delta \right\} \quad (\delta > 0) \tag{3.14}$$

where $\|\cdot\|_{TV}$ is the total variation norm of finite signed measures on $\mathcal{Z}$. Let ℓ be an arbitrary subcovex loss function on $\mathbb{R}^{d_\theta}$.[7] It can be shown that for *any* estimator τ_{NT} of θ (not necessarily a regular estimator), we have

$$\inf_{\delta>0} \liminf_{N,T\to\infty} \sup_{\check{\xi}\in B_\xi(\delta)} \mathbb{E}_{\check{\xi}} \ell \left(\sqrt{NT}(\tau_{NT} - \theta) \right) \geq \int \ell dN(0, V_\xi^{-1}) \tag{3.15}$$

where $\mathbb{E}_{\check{\zeta}}$ denotes an expectation operator under $\check{\zeta} = (\check{\theta}, \{\check{\eta}_i\}_{i=1}^\infty, \check{q})$. This result says that the maximum risk of an arbitrary estimator is bounded below by the risk of the law $N(0, V_\xi^{-1})$. The local minimax result has its advantage over a convolution theorem in that it is valid for all estimators. However, it also has a drawback that it only evaluates the *maximum risk over a small neighborhood*, rather than the risk exactly at the point of the true parameter value (van der Vaart (2002, p.348)).

3.3 Application

As an application of Theorem 3.2, we will derive the efficiency bound for θ in the parametric conditional density model (Example 1).

Recall that, in Example 1, the density of $z_{it} = (y_{it}, x_{it}')'$ is factored into

$$f_{Y|X}(y_{it}|x_{it}, \theta, \eta_i) f_X(x_{it}|\eta_i) d\mu_Y(y_{it}) d\mu_X(x_{it}) \tag{3.16}$$

[7] For the definition of subcovex loss functions, see van der Vaart (2002, p.347).

where the form of $f_{Y|X}$ is known up to parameters (θ, η_i), while the shape of f_X is completely unknown except that it may depend on η_i. Note that in this setting $q = f_X$. We assume that the mapping

$$(\theta, \eta) \mapsto \sqrt{f_{Y|X}(y|x, \theta, \eta) f_X(x|\eta)} \tag{3.17}$$

is continuously Frechet differentiable in $L^2(\mu_Y \otimes \mu_X)$ and its Frechet derivative at (θ, η) is given by

$$\begin{pmatrix} \dot{\ell}_1(y|x, \eta) \\ \dot{\ell}_2(y|x, \eta) + g(x, \eta) \end{pmatrix} \times \sqrt{f_{Y|X}(y|x, \theta, \eta) f_X(x|\eta)} \tag{3.18}$$

(Assumption 2). If the density were differentiable with respect to θ and η, then we could write

$$\dot{\ell}_1(y|x, \eta) = \frac{\partial}{\partial \theta} \log f_{Y|X}(y|x, \theta, \eta), \quad \dot{\ell}_2(y|x, \eta) = \frac{\partial}{\partial \eta} \log f_{Y|X}(y|x, \theta_0, \eta), \tag{3.19}$$

$$g(x, \eta) = \frac{\partial}{\partial \eta} \log f_X(x, \eta). \tag{3.20}$$

It can be shown that $\mathcal{T}_{f_X}$ is a set of functions $\dot{l}(x|\eta)$ such that $\mathbb{E}_\eta \dot{l}(x|\eta) = 0$ for every η and the map $\eta \mapsto \dot{l}(x|\eta)\sqrt{f_X(x, \eta)}$ is continuous in $L^2(\nu)$. Then, in this model,

$$\mathcal{T} = \left\{ \tilde{\eta}\dot{\ell}_2 + \dot{l} \in L^2(\mathbb{P}^\Gamma_{\theta,q}) : \tilde{\eta} \in C_b(\Lambda),\ \dot{l} \in \mathcal{T}_{f_X} \right\}. \tag{3.21}$$

It is easy to verify that

$$\Pi_{\mathcal{T}} \dot{\ell}_1(y|x, \eta) = \left(\frac{\mathbb{E}_\eta \dot{\ell}_1(y|x, \eta)\dot{\ell}_2(y|x, \eta)}{\mathbb{E}_\eta \dot{\ell}_2^2(y|x, \eta)} \right) \dot{\ell}_2(y|x, \eta). \tag{3.22}$$

As noted in the preceding subsection, this can be calculated as the projection of $\ell_1(y|x, \eta)$ on the closure of $\mathcal{T}_\eta$. We note that the bias-corrected fixed effects estimator $\hat{\theta}$ of Hahn and Newy (2004) (an analytical bias corrected version) admits the following expansion

$$\sqrt{NT}(\hat{\theta} - \theta) = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} V_\xi^{-1} \left(\dot{\ell}_1(y_{it}|x_{it}, \eta_i) - \Pi_{\mathcal{T}} \dot{\ell}_1(y_{it}|x_{it}, \eta_i) \right) + o_p(1) \tag{3.23}$$

as $N, T \to \infty$ with $N/T^3 \to \infty$ under certain regularity conditions. Thus, their bias-corrected fixed-effects estimator is asymptotically efficient.

4 The Derivation of the Efficiency Bound

The efficiency bound for a nonlinear panel data model with fixed effects cannot be obtained by the usual semiparametric efficiency arguments for models with i.i.d. observations because of the presence of fixed effects and because we consider double asymptotics. In this section, we will give a detailed explanation of how we obtain the efficiency bound presented in Theorem 3.2. In order to derive the efficiency bound, we use the convolution theorem by van der Vaart and Wellner (1996). The most difficult part is how to set a tangent set for the whole model, which is far from straightforward because of the presence of the fixed effects. In the first subsection below, we introduce the convolution theorem by van der Vaart and Wellner (1996). In the second subsection, we then consider applying the theorem to our panel data model and derive the efficiency bound for estimating θ.

4.1 A convolution theorem of van der Vaart and Wellner (1996)

In this section, we briefly review the convolution theorem in van der Vaart and Wellner (1996).[8] We begin by introducing their notation and terminology. Let H be a linear subspace of a Hilbert space $\mathbb{H}$ with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. For each $N \in \mathbb{N}$ and $h \in H$, let $P_{N,h}$ be a probability measure on a measurable space $(\mathcal{X}_N, \mathcal{A}_N)$. Let B denote the M-dimensional Euclidean space.[9] Consider a problem of estimating a 'localized parameter' $\kappa_N(h) \in B$ given an observation with law $P_{N,h}$. Let $\{\Delta_h : h \in H\}$ be an iso-Gaussian process indexed by H. That is, the process $\{\Delta_h : h \in H\}$ is a Gaussian process with mean zero and covariance function $\mathbb{E}\Delta_{h_1}\Delta_{h_2} = \langle h_1, h_2 \rangle$. The sequence of experiments $\{\mathcal{X}_N, \mathcal{A}_N, P_{N,h} : h \in H\}$ or simply $\{P_{N,h} : h \in H\}$ is said to be *locally asymptotically normal (LAN)* if we can write

$$\log \frac{dP_{N,h}}{dP_{N,0}} = \Delta_{N,h} - \frac{1}{2}\|h\|^2 + o_p(1),$$

as $N \to \infty$ under $P_{N,0}$where $\Delta_{N,h}$ is a sequence of random variables such that as $N \to \infty$,

$$\Delta_{N,h} \overset{0}{\rightsquigarrow} \Delta_h, \quad \text{marginally.} \tag{4.1}$$

Here, $\overset{h}{\rightsquigarrow}$ denotes weak convergence under $P_{N,h}$. The more precise expression of the condition (4.1) is that for any finite subset $\{h_1, h_2, \cdots, h_d\} \subseteq H$,

$$\begin{pmatrix} \Delta_{N,h_1} \\ \Delta_{N,h_2} \\ \vdots \\ \Delta_{N,h_d} \end{pmatrix} \overset{0}{\rightsquigarrow} N(0, (\langle h_i, h_j \rangle)), \tag{4.2}$$

as $N \to \infty$ where $(\langle h_i, h_j \rangle)$ is a $d \times d$ matrix whose (i,j)-th component is $\langle h_i, h_j \rangle$. The sequence of parameters $\kappa_N(h)$ is assumed to be *regular* (or *differentiable*) with respect to a norming real sequence $\mathbf{r}_N$ in the sense that as $N \to \infty$,

$$\mathbf{r}_N(\kappa_N(h) - \kappa_N(0)) \to \dot{\kappa}, \quad \forall h \in H,$$

for some bounded linear map $\dot{\kappa} : H \to B$. A sequence of estimators τ_N is said to be *regular* with respect to $\mathbf{r}_N$ if, as $N \to \infty$,

$$\mathbf{r}_N(\tau_N - \kappa_N(h)) \overset{h}{\rightsquigarrow} L, \quad \forall h \in H.$$

[8]While we consider double asymptotics under which both N and T tend to infinity, the convolution theorem stated here considers only N as the index that tends to infinity. However, the theorem is sufficient for our purpose by the following argument. The theorem can be directly applied to the case of diagonal asymptotics under which T depends on N, say $T = T_N$, and $T_N \to \infty$ as $N \to \infty$. If a convergence result holds under any diagonal asymptotics in which T_N is nondecreasing in N, then that result also holds under double asymptotics (see Remark (a) after Definition 2 in Phillips and Moon (1999)). This condition is satisfied in our convolution theorem. When $N, T \to \infty$ with a restriction such as $N/T^3 \to 0$, we consider only nondecreasing sequences T_N that satisfy that restriction.

[9]van der Vaart and Wellner (1996) considers a more general setting where B is a Banach space, rather than merely the Euclidean space. For our purposes, however, it is sufficient to consider the special case where B is the Euclidean space.

It should be emphasized that this definition requires that the limit distribution L be the same across h.[10] as $N \to \infty$. A bounded linear map $\dot{\kappa} : H \to B$ has an adjoint map $\dot{\kappa}^* : B \to \bar{H}$, where $\bar{H}$ is the completion of H.[11] The adjoint map is determined by the relation

$$\langle \dot{\kappa}^* b, h \rangle = b' \dot{\kappa}(h),$$

for all $h \in H$ and all $b \in B$. Denote by e_j an M-dimensional vector such that the j-th component is one but all other elements are zero. Define $\dot{\kappa}_j^* := \dot{\kappa}^* e_j$. The following theorem, which is broadly known as a convolution theorem, is a special case of Theorem 3.11.2 in van der Vaart and Wellner (1996).

Theorem 4.1 (A special case of Theorem 3.11.2 of van der Vaart and Wellner (1996))**.** Assume that $(P_{N,h} : h \in H)$ is locally asymptotically normal. Further, assume that the sequence of parameters $\kappa_N(h)$ and that of estimators τ_N are regular. Then, the limit distribution L of $\mathbf{r}_N(\tau_N - \kappa_N(0))$ equals the sum $G + W$ of two independent random vectors in B such that

$$G \sim N\left(0, (\langle \dot{\kappa}_i^*, \dot{\kappa}_j^* \rangle)\right).$$

Applying the same argument as in Section 3, we see that the theorem implies that the law of G is optimal in the sense that the asymptotic variances of any regular estimators cannot be smaller than the variance of G.

4.2 The derivation of the efficiency bound

We now consider applying Theorem 4.1 to our nonlinear panel data model with fixed effects. In order to keep the developments as simple as possible and highlight the difficulties caused by fixed effects, we first consider the case without a nuisance function q. That is, we deal with the case where the observation z_{it} is generated according to the density $f(z_{it}|\theta, \eta_i)$. In this case, the totality of the parameters is $\xi = (\theta, \{\eta_i\}_{i=1}^{\infty})$. Later, we will show how to extend our arguments to the case where the nuisance function q is also present. The derivation of the efficiency bound consists of two steps. First, we establish the LAN property. Second, we apply the convolution theorem to calculate the bound. The second part is not long, while the first part requires a considerable work. Thus, most of this subsection is devoted to the first part, i.e., the verification of the LAN property of our panel data model with fixed effects.

In order to establish the LAN property, we begin by considering how to set the indexed set H. The convolution theorem requires that H is a linear subspace of some Hilbert space $\mathbb{H}$. However, it is usually sufficient to verify the following two conditions: (i) H is a linear space and (ii) we can define an inner

[10]When the statistical experiment is indexed by N and T and we consider double asymptotics, the regularity of a parameter $\kappa_{NT}(h)$ can be defined using the nondecreasing sequence argument. For example, a parameter $\kappa_{NT}(h)$ is said to be regular with respect to a norming sequence $\mathbf{r}_{NT}$ if there is some bounded linear map $\dot{\kappa} : H \to B$ such that for every nondecreasing sequence T_N we have $\mathbf{r}_{NT_N}(\kappa_{NT_N}(h) - \kappa_{NT_N}(0)) \to \dot{\kappa}$ for all $h \in H$, as $N \to \infty$. The regularity of an estimator τ_{NT} is defined in an analogous way. When $N, T \to \infty$ with a restriction such as $N/T^3 \to 0$, we consider only nondecreasing sequences T_N that satisfy that restriction.

[11]van der Vaart and Wellner (1996) considers an adjoint map from the dual space B^* of B to $\bar{H}$. In the present setting, since B is the Euclidean space, we can identify the dual space B^* with B.

product on H using the second term in the LAN expansion. The requirement that $\mathbb{H}$ be a Hilbert space is not essential because every inner product space can be completed. The condition (ii) can be further relaxed; even if H is only a semi-inner product space, there will be no problem because we can always take a quotient space. More specifically, if $\langle \cdot, \cdot \rangle$ is a semi-inner product on H, then we consider the quotient space H/K where $K := \{h \in H : \langle h, h \rangle = 0\}$ (for a definition of a quotient space, see, e.g., Rudin (1991)).

In models with i.i.d. observations, the probability measure $P_{N,h}$ in Theorem 4.1 corresponds to a law that is localized around the truth in the direction $h \in H$. Thus the choice of H is closely related to a localization scheme we will adopt. We first discuss the following seemingly natural localization scheme, which is employed in Hahn and Kuersteiner (2002) in the context of panel autoregressive models. As we shall see, the method cannot be easily applied to our *general* nonlinear panel data model.

In view of the usual LAN theory for parametric i.i.d. models (see, e.g, van der Vaart (1998, Chapter 7)), it may seem natural to localize the parameter $\xi = (\theta, \{\eta_i\}_{i=1}^{\infty})$ as follows:

$$\theta + \frac{1}{\sqrt{NT}}\tilde{\theta}, \eta_1 + \frac{1}{\sqrt{NT}}\tilde{\eta}_1, \eta_2 + \frac{1}{\sqrt{NT}}\tilde{\eta}_2, \cdots . \tag{4.3}$$

Here, $\tilde{\theta} \in \mathbb{R}^{d_\theta}$ and $\{\tilde{\eta}_i\}_{i=1}^{\infty}$ is some bounded real sequence. In this case, we regard $(\tilde{\theta}, \{\tilde{\eta}_i\}_{i=1}^{\infty})$ as h in Theorem 4.1. Let $\tilde{W} \subseteq \mathbb{R}^{\mathbb{N}}$ be a set for perturbing sequences $\{\tilde{\eta}_i\}_{i=1}^{\infty}$ and define $H = \mathbb{R}^{d_\theta} \times \tilde{W}$. Let $P_{NT,h}$ be the law of the panel data $\{\{z_{it}\}_{t=1}^{T}\}_{i=1}^{N}$ under the localized parameter (4.3) and let $P_{NT,0}$ be the law under the true parameter. The local log likelihood ratio is given by

$$\log \frac{dP_{NT,h}}{dP_{NT,0}} = \sum_{t=1}^{T}\sum_{j=1}^{N} \log \frac{f\left(z_{it}|\theta + \tilde{\theta}/\sqrt{NT}, \eta_i + \tilde{\eta}_i/\sqrt{NT}\right)}{f(z_{it}|\theta, \eta_i)}. \tag{4.4}$$

By heuristic calculations, we expect that the log likelihood ratio is expanded as

$$\begin{aligned} \log \frac{dP_{NT,h}}{dP_{NT,0}} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{\theta}'\dot{\ell}_1(z_{it}|\eta_i) + \tilde{\eta}_i\dot{\ell}_2(z_{it}|\eta_i)\right) \\ &\quad - \lim_{N\to\infty}\frac{1}{2N}\sum_{i=1}^{N}\mathbb{E}_{\eta_i}\left(\tilde{\theta}'\dot{\ell}_1(z_{it}|\eta_i) + \tilde{\eta}_i\dot{\ell}_2(z_{it}|\eta_i)\right)^2 + o_p(1) \end{aligned} \tag{4.5}$$

as $N, T \to \infty$ under $P_{NT,0}$. In order to verify the LAN condition, we have to specify the set $\tilde{W}$.The most natural candidate is the set $\tilde{W}_0$ of all the sequences $\{\tilde{\eta}_i\}_{i=1}^{\infty}$ such that the above expansion is valid and the first term satisfies some central limit theorem. However, this specification is not generally applicable because the set $\tilde{W}_0$ is not necessarily a linear space (notice that since H must be linear, $\tilde{W}$ itself must also be linear). The following example illustrates this fact.

Example 3. We consider the celebrated example of Neyman and Scott (1948). Suppose that z_{it} is real-valued and can be written as $z_{it} = \eta_i + u_{it}$. Here u_{it} is an error term that is i.i.d. across i and t. Assume that $u_{it} \sim N(0, \gamma)$. In this example, $\theta = \gamma > 0$. By direct calculations, we can show that whenever $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\tilde{\eta}_i^2$ exists and is finite, then as $N, T \to \infty$

$$\log \frac{dP_{NT,h}}{dP_{NT,0}} = \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\frac{\tilde{\gamma}}{2}\left(\frac{u_{it}^2}{\gamma^2} - \frac{1}{\gamma}\right) - \frac{\tilde{\eta}_i u_{it}}{\gamma}\right\} \tag{4.6}$$

$$-\frac{1}{2}\left\{\frac{\tilde{\gamma}^2}{4}\mathbb{E}\left(\frac{u_{it}^2}{\gamma^2} + \frac{1}{\gamma}\right)^2 + \frac{1}{\gamma}\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\tilde{\eta}_i^2\right\} + o_p(1), \tag{4.7}$$

under $P_{NT,0}$ and the first term in the expansion is asymptotically normal with mean 0 and variance twice the second term.[12] Here $\tilde{\gamma}$ corresponds to $\tilde{\theta}$ in a general setting. Conversely, if this expansion is valid, then obviously the limit $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \tilde{\eta}_i^2$ exists and is finite. Thus, in this simple model,

$$\tilde{W}_0 = \left\{ \{\tilde{\eta}_i\}_{i=1}^\infty \in \mathbb{R}^{\mathbb{N}} : \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \tilde{\eta}_i^2 \quad \text{exists and is finite.} \right\}. \tag{4.8}$$

The set $\tilde{W}_0$ of Cesaro convergent sequences is *not* a linear space.To see this, set $a_i = 1$ for every i and $b_i = (-1)^k$ where k is the unique integer determined by $2^k < i \le 2^{k+1}$ ($i \in \mathbb{N}$). Then $\lim \frac{1}{N}\sum_{i=1}^N a_i^2 = \lim \frac{1}{N}\sum_{i=1}^N b_i^2 = 1$. Hence, $\{\tilde{\eta}_i\}_{i=1}^\infty$ and $\{b_i\}_{i=1}^\infty$ are both in $\tilde{W}_0$, provided that $\pm 1 \in \Lambda$. However, $\frac{1}{N}\sum_{i=1}^N a_i b_i$ oscillates and so does $\frac{1}{N}\sum_{i=1}^N (a_i+b_i)^2$. Thus, the sum $\{a_i+b_i\}_{i=1}^\infty$ does not lie in $\tilde{W}_0$.

Thus we must consider restricting $\tilde{W}_0$ in such a way that it becomes a linear space. However, it is not an easy task to find an appropriate restriction on $\tilde{W}_0$, because it can be model-specific. Further, a situation can be much more complicated when we consider a model with a nuisance function q.

One may think that just assuming the conditions required for the local asymptotic normality, instead of specifying the set $\tilde{W}$ concretely, may be sufficient for our present purpose. However, such a high-level approach is not always satisfactory for the following reason. Calculating the efficiency bounds typically involves a projection onto the space spanned by scores for nuisance parameters. Not specifying the space H can lead to not specifying the space of scores, making the calculation of the projection infeasible.

The main problem about the above perturbation method is intractability of the sequence space for $\{\tilde{\eta}_i\}_{i=1}^\infty$. In order to avoid the technical difficulties caused by the sequence space, we take a different approach. Motivated by Strasser (1998), we use a function, rather than a sequence, to perturb the parameters. To be more precise, let $\tilde{\eta}(\cdot) \in C_b(\Lambda)$ and localize the parameter $\xi = (\theta, \{\eta_i\}_{i=1}^\infty)$ as follows:

$$\theta + \frac{1}{\sqrt{NT}}\tilde{\theta}, \eta_1 + \frac{1}{\sqrt{NT}}\tilde{\eta}(\eta_1), \eta_2 + \frac{1}{\sqrt{NT}}\tilde{\eta}(\eta_2), \cdots, \tag{4.9}$$

where, as before, $\tilde{\theta} \in \mathbb{R}^{d_\theta}$. This localization looks similar to (4.3) but differs in that each η_i is now perturbed by the function $\tilde{\eta}(\eta_i)$ evaluated at η_i. Note that since the true parameter $\xi = (\theta, \{\eta_i\}_{i=1}^\infty)$ is held fixed, the pair $(\tilde{\theta}, \tilde{\eta}(\cdot))$ completely determines the localized parameter (4.9). Thus, under this localization, we can regard the pair $(\tilde{\theta}, \tilde{\eta}(\cdot))$ to be h in Theorem 4.1. We set $H := \mathbb{R}^{d_\theta} \times C_b(\Lambda)$. As before, we denote by $P_{NT,h}$ the law of the panel data under the localized parameter (4.9).

Under this localization scheme, we can now verify the LAN condition required in Theorem 4.1. First, it is obvious that the set $H = \mathbb{R}^{d_\theta} \times C_b(\Lambda)$ is a linear space. Further, under Assumptions 1, 2 and 4, we rigorously prove in the Appendix that the asymptotic expansion (4.5) is valid with $\tilde{\eta}_i$ replaced by $\tilde{\eta}(\eta_i)$. We also show in the Appendix that the first term in the LAN expansion converges in distribution to a zero-mean normal distribution whose variance is twice the second term. Based on the LAN expansion, we next consider a LAN inner product on H. Notice that by Assumption 2 and the continuity of $\tilde{\eta}(\cdot)$, the map

$$\eta \mapsto \mathbb{E}_\eta \left(\tilde{\theta}' \dot{\ell}_1(z|\eta) + \tilde{\eta}(\eta) \dot{\ell}_2(z|\eta) \right)^2 \tag{4.10}$$

[12] In this example, we do not assume that Λ is closed and bounded, simply because we do not need the assumption in order for the expansion (4.6) to hold.

is continuous on Λ. Thus, by Assumption 4, we have

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}_{\eta_i}\left(\tilde{\theta}'\dot{\ell}_1(z_{it}|\eta_i) + \tilde{\eta}(\eta_i)\dot{\ell}_2(z|\eta_i)\right)^2 = \int \mathbb{E}_{\eta}\left(\tilde{\theta}'\dot{\ell}_1(z|\eta) + \tilde{\eta}(\eta)\dot{\ell}_2(z|\eta)\right)^2. \quad (4.11)$$

That is, we can write the second term in the LAN expansion as the integral with respect to $\mathbb{P}_{\theta}^{\Gamma}$, rather than a Cesaro limit. As noted in Section 3, this greatly facilitates our use of techniques from functional analysis: integrals are much more tractable than Cesaro limits. This is the main advantage of our localization scheme and weak convergence assumption on individual effects (Assumption 4). Using this integral, the LAN inner product on H is defined as follows:

$$\langle h_a, h_b\rangle := \int \mathbb{E}_{\eta}\left(\tilde{\theta}_a'\dot{\ell}_1(z|\eta) + \tilde{\eta}_a(\eta)\dot{\ell}_2(z|\eta)\right)\left(\tilde{\theta}_b'\dot{\ell}_1(z|\eta) + \tilde{\eta}_b(\eta)\dot{\ell}_2(z|\eta)\right) d\Gamma(\eta).$$

for $h_a = (\tilde{\theta}_a, \tilde{\eta}_a)$ and $h_b = (\tilde{\theta}_b, \tilde{\eta}_b)$. Rigorously speaking, this functional $\langle\cdot,\cdot\rangle$ on $H \times H$ is a semi-inner product on H, so we consider the quotient space H/K, as noted above. However, we will simply write H for H/K for simplicity of notation. We set $\mathbb{H}$ to be the completion of H with respect to the inner product $\langle\cdot,\cdot\rangle$. Since we have checked all the conditions required for the LAN property, we have established the following result.

Theorem 4.2 (LAN)**.** Consider the situation without the nuisance function q. Suppose that Assumptions 1, 2 and 4 hold. We define $P_{NT,h}$, H, $\mathbb{H}$ and $\langle\cdot,\cdot\rangle$ as above. Then, the statistical experiment $\{P_{NT,h} : h \in H\}$ is locally asymptotically normal in the sense of van der Vaart and Wellner (1996). That is, for every $h = (\tilde{\theta}, \tilde{\eta}(\cdot)) \in H$, we have

$$\log\frac{dP_{NT,h}}{dP_{NT,0}} = \Delta_{NT,h} - \frac{1}{2}\|h\|^2 + o_p(1)$$

as $N, T \to \infty$ under $P_{NT,0}$ where

$$\Delta_{NT,h} := \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{\theta}'\dot{\ell}_1(z_{it}|\eta_i) + \tilde{\eta}(\eta_i)\dot{\ell}_2(z_{it}|\eta_i)\right) \quad (4.12)$$

converges weakly to $\Delta_h \sim N(0, \|h\|^2)$, marginally.

Next we turn to extend the LAN results to the case with a nuisance function q. For this purpose, we set $H := \mathbb{R}^{d_\theta} \times C_b(\Lambda) \times \mathcal{T}_q$. By Assumption 3, it is easily seen that H is a linear space. By definition of $\mathcal{T}_q$, for each $\dot{l} \in \mathcal{T}_q$, we can pick out a parametrization q_s such that (3.2) holds. For $h = (\tilde{\theta}, \tilde{\eta}, \dot{l}) \in H$, let $P_{NT,h}$ be the probability law whose density is

$$\prod_{i=1}^{N}\prod_{t=1}^{T} f\left(z_{it}\middle|\theta + \tilde{\theta}/\sqrt{NT}, \eta_i + \tilde{\eta}(\eta_i)/\sqrt{NT}, q_{1/\sqrt{NT}}\right). \quad (4.13)$$

By exactly the same arguments as in the case without the nuisance function q, we have the following expansion: under $P_{NT,0}$,

$$\begin{aligned}\log\frac{dP_{NT,h}}{dP_{NT,0}} &= \frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\tilde{\theta}'\dot{\ell}_1(z_{it}|\eta_i) + \tilde{\eta}(\eta_i)\dot{\ell}_2(z_{it}|\eta_i) + \dot{l}(z_{it}|\eta_i)\right) \\ &\quad -\frac{1}{2}\int \mathbb{E}_{\eta}\left(\tilde{\theta}'\dot{\ell}_1(z|\eta) + \tilde{\eta}(\eta)\dot{\ell}_2(z|\eta) + \dot{l}(z|\eta)\right)^2 d\Gamma(\eta) + o_p(1) \quad (4.14)\end{aligned}$$

as $N, T \to \infty$. Based on this expansion, we define an inner product on H by

$$\langle h_a, h_b \rangle := \int \mathbb{E}_\eta \left(\tilde{\theta}_a' \dot{\ell}_1(z|\eta) + \tilde{\eta}_a(\eta) \dot{\ell}_2(z|\eta) + \dot{l}_a(z|\eta) \right) \left(\tilde{\theta}_b' \dot{\ell}_1(z|\eta) + \tilde{\eta}_b(\eta) \dot{\ell}_2(z|\eta) + \dot{l}_b(z|\eta) \right) d\Gamma(\eta)$$

for $h_a = (\tilde{\theta}_a, \tilde{\eta}_a, \dot{l}_a)$ and $h_b = (\tilde{\theta}_b, \tilde{\eta}_b, \dot{l}_b)$ in H. Again, since this function is a semi-inner product, we consider the quotient space H/K. As before, we will simply write H for H/K. Let $\mathbb{H}$ be the completion of H with respect to the inner product $\langle \cdot, \cdot \rangle$. Since all the requirements for the LAN property are satisfied, we have thee following result.

Theorem 4.3 (LAN)**.** Consider the situation with a nuisance function q. Suppose that Assumptions 1 to 4 hold. We define $P_{NT,h}$, H, $\mathbb{H}$ and $\langle \cdot, \cdot \rangle$ as above. Then, the statistical experiment $\{P_{NT,h} : h \in H\}$ is locally asymptotically normal in the sense of van der Vaart and Wellner (1996). That is, for every $h = (\tilde{\theta}, \tilde{\eta}(\cdot)) \in H$, we have

$$\log \frac{dP_{NT,h}}{dP_{NT,0}} = \Delta_{NT,h} - \frac{1}{2} \|h\|^2 + o_p(1)$$

as $N, T \to \infty$ under $P_{NT,0}$ where

$$\Delta_{NT,h} := \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\tilde{\theta}' \dot{\ell}_1(z_{it}|\eta_i) + \tilde{\eta}(\eta_i) \dot{\ell}_2(z_{it}|\eta_i) + \dot{l}(z_{it}|\eta_i) \right) \tag{4.15}$$

converges weakly to $\Delta_h \sim N(0, \|h\|^2)$, marginally.

Remark 4.1. When interpreting the efficiency bound given in Theorem 3.2, it is more useful to consider the LAN result in terms of a score function. We define a score operator[13] A from $\mathbb{R}^{d_\theta} \times C_b(\Lambda) \times \mathcal{T}_q$ to $L^2(\mathbb{P}_\theta^\Gamma)$ by

$$A(h) = A(\tilde{\theta}, \tilde{\eta}, \dot{l}) = \tilde{\theta}' \dot{\ell}_1 + \tilde{\eta} \dot{\ell}_2 + \dot{l}. \tag{4.16}$$

The function $\tilde{\theta}' \dot{\ell}_1(\cdot|\eta) + \tilde{\eta}(\eta) \dot{\ell}_2(\cdot|\eta) + \dot{l}$ can be regarded as a score function for the one-dimensional submodel $f(\cdot|\theta + s\tilde{\theta}, \eta + s\tilde{\eta}(\eta), q_s)$ where $s \in (-\epsilon, \epsilon)$ is the parameter of the submodel (this is the reason that the operator A is called a 'score' operator).

Since A is a linear operator and

$$\langle h_a, h_b \rangle = \langle A(h_a), A(h_b) \rangle_{L^2(\mathbb{P}_{\theta,q}^\Gamma)} \tag{4.17}$$

for all $(\tilde{\theta}_a, \tilde{\eta}_a, \dot{l}_a)$ and $(\tilde{\theta}_b, \tilde{\eta}_b, \dot{l})$ in H, we see that A is an isometry. It follows that the inner product space H can be identified with the range of A:

$$\mathcal{J} := \text{Range}(A) = \left\{ \tilde{\theta}' \dot{\ell}_1 + \tilde{\eta} \dot{\ell}_2 + \dot{l} \in L^2\left(\mathbb{P}_{\theta,q}^\Gamma\right) : \tilde{\theta} \in \mathbb{R}^{d_\theta}, \tilde{\eta} \in C_b(\Lambda), \dot{l} \in \mathcal{T}_q \right\}. \tag{4.18}$$

As a result, the completion $\mathbb{H}$ coincides with the closure of $\mathcal{J}$. Since the function $\tilde{\theta}' \dot{\ell}_1 + \tilde{\eta} \dot{\ell}_2 + \dot{l}$ corresponds to a score function for a parametric submodel of our model, the set $\mathcal{J}$ can be interpreted as the tangent set for our panel data model.

Based on Theorem 4.3, we now prove Theorem 3.2.

[13] See van der Vaart (1998, Chapter 25) for a score operator in the context of i.i.d. models.

Proof of Theorem 3.2. To prove this theorem, we follow the arguement in van der Vaart (1998, p369), which is a standard way to derive the efficiency bounds for semiparametric models. A proof using Theorem 4 in Hahn (2002) is also possible but in this case our direct proof is much simpler.

The 'parameter' of interest is now $k_{NT}(h) = \theta + \tilde{\theta}/\sqrt{NT}$ and so we have $\dot{k}(h) = \tilde{\theta}$. Since $B = \mathbb{R}^p$, every continuous linear finctional $b^* : B \to \mathbb{R}$ takes the form $\mathbb{R}^p \ni x \mapsto b'x$ for some $b \in \mathbb{R}^p$. Define a linear map $k^* : B \to \mathbb{H}$ by $\dot{k}^* b^* = b' V_\xi^{-1} \left(\dot{\ell}_1(z|\eta) - \Pi_{\mathcal{T}} \dot{\ell}_1(z|\eta) \right)$. Observe that for any $h = (\tilde{\theta}, \tilde{\eta}, \dot{l}) \in H$ and $b \in \mathbb{R}^p$

$$
\begin{aligned}
\langle \dot{k}^* b^*, h \rangle &= \int \mathbb{E}_\eta b' V_\xi^{-1} \left(\dot{\ell}_1(z|\eta) - \Pi_{\mathcal{T}} \dot{\ell}_1(z|\eta) \right) \left(\tilde{\theta}' \dot{\ell}_1(z|\eta) + \tilde{\eta}(\eta) \dot{\ell}_2(z|\eta) + \dot{l}(z|\eta) \right) d\Gamma(\eta) \\
&= \int \mathbb{E}_\eta b' V_\xi^{-1} \left(\dot{\ell}_1(z|\eta) - \Pi_{\mathcal{T}} \dot{\ell}_1(z|\eta) \right) \left(\tilde{\theta}' \dot{\ell}_1(z|\eta) \right) d\Gamma(\eta) \\
&= \int \mathbb{E}_\eta b' V_\xi^{-1} \left(\dot{\ell}_1(z|\eta) - \Pi_{\mathcal{T}} \dot{\ell}_1(z|\eta) \right) \left(\dot{\ell}_1(z|\eta) - \Pi_{\mathcal{T}} \dot{\ell}_1(z|\eta) \right)' \tilde{\theta} d\Gamma(\eta) \\
&= b' \tilde{\theta}.
\end{aligned}
$$

This shows that $\dot{k}^*$ is the adjoint map of $\dot{k}$ under the LAN inner product defined in Theorem 4.3. Further, it can be easily seen that $\|\dot{k}^* b^*\|^2 = b' V_\xi^{-1} b$. Since this holds for arbitrary $b \in \mathbb{R}^p$, we prove the theorem. □

5 Conclusion

This paper derives the efficiency bounds for estimates of model parameters in nonlinear panel data models with individual effects when both N and T tend to infinity. Our results verify that bias-corrected fixed effects estimators considered in Hahn and Newey (2005) are asymptoticaly efficient. To derive the efficiency bounds, we apply a convolution theorem that allows data to be non-i.i.d. and a parameter space to be infinite-dimensional. The presence of incidental parameters and our use of double asymptotics make the derivation of the lower bounds nontrivial and somewhat nonstandard. A weak convergence assumption on individual effects plays a key role in overcoming such difficulties. We conjecture that the method used in this paper can be applied to the derivation of the efficiency bounds for nonlinear dynamic panel data models as considered in Hahn and Kuersteiner (2011).

A Appendix: Proofs

A.1 Proof of Theorem 4.2

In this subsection, we prove Theorem 4.2, which establishes the LAN property of the nonlinear panel data models we consider. Under the perturbation (4.3), the local log likelihood ratio process of our nonlinear panel data model with individual effects is

$$
\log \frac{dP_{NT}, h}{dP_{NT,0}} = \sum_{t=1}^{T} \sum_{i=1}^{N} \log \frac{f\left(z_{it} | \theta + \tilde{\theta}/\sqrt{NT}, \eta_i + \psi(\eta_i)/\sqrt{NT} \right)}{f(z_{it}|\theta, \eta_i)}. \tag{A.1}
$$

To simplify notation let

$$
p_{NT}(z, \eta) := f\left(z \middle| \theta + \tilde{\theta}/\sqrt{NT}, \eta + \psi(\eta)/\sqrt{NT} \right) \tag{A.2}
$$

and let

$$p(z,\eta) := f\left(z|\theta,\eta\right). \tag{A.3}$$

We define

$$W_{NT}(z,\eta) := \begin{cases} 2\left(\sqrt{\frac{p_{NT}(z,\eta)}{p(z,\eta)}}-1\right) & \text{if } p(z,\eta)\neq 0 \\ 0 & \text{if } p(z,\eta)=0 \end{cases}$$

and note that $W_{NT}(z_{it},\eta_i)$ is well-defined with probability one. Apply a Taylor expansion to $\log(1+x)$ to obtain $\log(1+x) = x - (1/2)x^2 + x^2R(2x)$ where $R(x)\to 0$ as $x\to 0$. Using this expansion, we can write the local log likelihood ratio process as

$$\begin{aligned} \log\frac{dP_{NT},h}{dP_{NT,0}} &= \sum_{t=1}^{T}\sum_{i=1}^{N} W_{NT}(z_{it},\eta_i) - \frac{1}{4}\sum_{t=1}^{T}\sum_{i=1}^{N} W_{NT}^2(z_{it},\eta_i) \\ &\quad + \frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{N} W_{NT}^2(z_{it},\eta_i)R(W_{NT}(z_{it},\eta_i)). \end{aligned} \tag{A.4}$$

The local asymptptic normality of our nonlinear panel data model will be established through investigating the asymptotic behavior of each term appeared in the right hand side of (A.4) and verifying that the asymptotic variance of the central sequence defines an inner product on $H = \Theta \times C_b(\mathbb{R})$ and that the inner-product space is a subspace of a Hilbert space.

Before proceeding to the proof, we will introduce some additional notation. Define

$$g(z,\eta) := \tilde{\theta}'\dot{\ell}_1(z|\theta,\eta) + \psi(\eta)\dot{\ell}_2(z|\theta,\eta) \tag{A.5}$$

and

$$A_{NT}(\eta) := \sqrt{p_{NT}(z,\eta)} - \sqrt{p(z,\eta)}. \tag{A.6}$$

Further, write

$$r_{NT}(\eta) := A_{NT}(\eta) - \frac{1}{2\sqrt{NT}}g(z,\eta)\sqrt{p(z,\eta)}. \tag{A.7}$$

We begin the proof of the LAN property by verifying the following useful lemma, which is a consequence of Assumption 2 and 4.

Lemma A.1. (i) For every compact subset K of $\mathbb{R}$, we have

$$\limsup_{N,T\to\infty}\sup_{\eta\in K} NT\int r_{NT}^2(\eta)d\mu = 0. \tag{A.8}$$

Furthermore,

$$\limsup_{N,T\to\infty}\sup_{\eta\in K} NT\left|\int A_{NT}^2(z,\eta)d\mu - \frac{1}{4NT}\mathbb{E}_{\theta,\eta}g^2(z,\eta)\right| = 0. \tag{A.9}$$

(ii) We have

$$\limsup_{N,T\to\infty}\sup_{\eta\in\Lambda} NT\int r_{NT}^2(\eta)d\mu \le C_1 < \infty. \tag{A.10}$$

Moreover,

$$\limsup_{N,T\to\infty}\sup_{\eta\in\Lambda} NT\left|\int A_{NT}^2(z,\eta)d\mu - \frac{1}{4NT}\mathbb{E}_{\theta,\eta}g^2(z,\eta)\right| \le C_2 < \infty. \tag{A.11}$$

Proof. To show the first statement of (i), we introduce the following notation:

$$
\begin{aligned}
&s_{NT}((\theta', \eta)', (u_1', u_2)') \\
&:= \sqrt{f\left(z\middle|(\theta', \eta)' + \frac{1}{\sqrt{NT}}(u_1', u_2)'\right)} - \sqrt{f(z|\theta, \eta)} - \frac{1}{2\sqrt{NT}}(u_1', u_2)\dot{\ell}(z|\theta, \eta)\sqrt{f(z|\theta, \eta)}.
\end{aligned}
$$

From Lemma.5 of Appendix 9 in Bickel et. (1993, p509), it holds that for every compact sets K_1 and K_2 of $\mathbb{R}^{p+1}$,

$$
NT \int s_{NT}^2((\theta', \eta)', (u_1', u_2)')d\mu \tag{A.12}
$$

converges to 0 *uniformly* for $(\theta', \eta)' \in K_1$ *and* $u \in K_2$. Now fix θ, $\tilde{\theta}$ and ψ. Note that since ψ is assumed to be bounded, there exists some compact subset $A \subseteq \mathbb{R}$ such that $\psi(\eta) \in A$ for all $\eta \in \Lambda$. From the uniform convergence of (A.12) it follows that for every compact subset $K \subseteq \mathbb{R}$,

$$
\begin{aligned}
&\sup_{\eta \in K} NT \int r_{NT}^2(\eta)d\mu \\
&\qquad \leq \sup_{\eta \in K} \sup_{u_2 \in A} NT \int s_{NT}^2((\theta', \eta)', (\tilde{\theta}', u_2)')^2 d\mu \to 0.
\end{aligned}
$$

Thus the first statement of (i) follows.

To show the second statement of (i), first observe that

$$
\begin{aligned}
&NT \left| \int A_{NT}(z, \eta)^2 d\mu - \frac{1}{4NT}\mathbb{E}_{\theta,\eta} g^2(z, \eta) \right| \\
&\leq \left\{ NT \int r_{NT}^2(\eta)d\mu \right\}^{1/2} \left\{ NT \int \left(A_{NT}(\eta) + \frac{1}{2\sqrt{NT}} g(z, \eta)\sqrt{p(z, \eta)} \right)^2 d\mu \right\}^{1/2}
\end{aligned}
$$

where this inequality follows from the Cauchy-Schwarz inequality. Further, note that

$$
\begin{aligned}
&\sup_{\eta \in K} NT \int \left(A_{NT}(\eta) + \frac{1}{2\sqrt{NT}} g(z, \eta)\sqrt{p(z, \eta)} \right)^2 d\mu \\
&\leq 2 \left(\sup_{\eta \in K} NT \int r_{NT}^2(\eta)d\mu + \frac{1}{4^2} \sup_{\eta \in K} \mathbb{E}_{\theta,\eta} g^2(z, \eta) \right).
\end{aligned}
$$

By the first statement of (i), the first term in the last display is $o(1)$, and by Assumption **??** the second term is finite. It thus follows that

$$
\sup_{\eta \in K} NT \left| \int A_{NT}(z, \eta)^2 d\mu - \frac{1}{4NT}\mathbb{E}_{\theta,\eta} g^2(z, \eta) \right| = o(1),
$$

which is the required result.

The proof of (ii) is similar. □

Lemma A.2. As $N, T \to \infty$, we have

$$
\sum_{t=1}^{T} \sum_{i=1}^{N} W_{NT}(z_{it}, \eta_i) = \frac{1}{\sqrt{NT}} \sum_{t=1}^{T} \sum_{i=1}^{N} g(z_{it}, \eta_i) - \frac{1}{4} \int \mathbb{E}_{\theta,\eta} g^2(z_{it}, \eta) d\Gamma(\eta) + o_p(1). \tag{A.13}
$$

Proof. The proof proceeds through analysing the mean and variance of

$$
\sum_{t=1}^{T} \sum_{i=1}^{N} W_{NT}(z_{it}, \eta_i) - (1/\sqrt{NT}) \sum_{t=1}^{T} \sum_{i=1}^{N} g(z_{it}, \eta_i). \tag{A.14}
$$

Because we are faced with incidental parameters and use double asymptotics, the proof is more technical than in the case with i.i.d observations.

First observe that

$$\begin{aligned}
& \operatorname{Var}\left[\sum_{t=1}^{T}\sum_{i=1}^{N} W_{NT}(z_{it},\eta_i) - \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N} g(z_{it},\eta_i)\right] && \text{(A.15)}\\
= \;& T\sum_{i=1}^{N}\operatorname{Var}\left[W_{NT}(z_{it},\eta_i) - \frac{1}{\sqrt{NT}} g(z_{it},\eta_i)\right] && \text{(A.16)}\\
\leq \;& T\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i}\left[W_{NT}(z_{it},\eta_i) - \frac{1}{\sqrt{NT}} g(z_{it},\eta_i)\right]^2 && \text{(A.17)}\\
= \;& 4T\sum_{i=1}^{N}\int\left[(p_{NT}(z,\eta_i) - p(z,\eta_i)) - \frac{1}{2\sqrt{NT}} g(z,\eta_i)\sqrt{p(z,\eta)}\right]^2 d\mu(z). && \text{(A.18)}
\end{aligned}$$

The etreme tight hand side of (A.15) can be written as $4T\sum_{i=1}^{N}\int r_{NT}^2(\eta)d\mu$. By Assumption 2, we have

$$\lim_{N,T\to\infty} NT\int r_{NT}^2(\eta)d\mu = 0. \tag{A.19}$$

Now take $\epsilon > 0$. By Remark **??**, there exists a compact set $K \subseteq \Lambda$ such that

$$\frac{1}{N}\sum_{i=1}^{N} 1_K(\eta_i) > 1-\epsilon \quad \text{for all } n \in \mathbb{N}. \tag{A.20}$$

$$\limsup_{N,T\to\infty}\sup_{\eta\in K} NT\int r_{NT}^2(\eta)d\mu = 0. \tag{A.21}$$

Moreover, by Assumption **??** and the boundedness of ψ we have

$$\limsup_{N,T\to\infty}\sup_{\eta\in\Lambda} NT\int r_{NT}^2(\eta)d\mu \leq C < \infty. \tag{A.22}$$

Combining these results yields

$$\begin{aligned}
& \limsup_{N,T\to\infty} T\sum_{i=1}^{N}\int r_{NT}^2(\eta_i)d\mu && \text{(A.23)}\\
\leq \;& \limsup_{N,T\to\infty}\frac{1}{N}\sum_{\eta_i\in K} NT\int r_{NT}^2(\eta_i)d\mu + \limsup_{N,T\to\infty}\frac{1}{N}\sum_{\eta_i\notin K} NT\int r_{NT}^2(\eta_i)d\mu && \text{(A.24)}\\
\leq \;& \limsup_{N,T\to\infty}\sup_{\eta\in K} NT\int r_{NT}^2(\eta)d\mu + \epsilon\limsup_{N,T\to\infty}\sup_{\eta\in\Lambda} NT\int r_{NT}^2(\eta)d\mu && \text{(A.25)}\\
\leq \;& C\epsilon. && \text{(A.26)}
\end{aligned}$$

Since this holds for arbitrary $\epsilon > 0$, we have $\limsup_{N,T\to\infty} T\sum_{i=1}^{N}\int r_{NT}^2(\eta_i)d\mu = 0$. Thus it follows that as $N, T \to \infty$,

$$\operatorname{Var}\left[\sum_{t=1}^{T}\sum_{i=1}^{N} W_{NT}(z_{it},\eta_i) - \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N} g(z_{it},\eta_i)\right] = o(1). \tag{A.27}$$

Next we turn to the expectation of (A.14). Observe that

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{N} W_{NT}(z_{it},\eta_i) - (1/\sqrt{NT})\sum_{t=1}^{T}\sum_{i=1}^{N} g(z_{it},\eta_i)\right] \tag{A.28}$$

$$= T\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i} W_{NT}(z_{it},\eta_i) \tag{A.29}$$

$$= T\sum_{i=1}^{N}\left(2\int\sqrt{p_{NT}(z,\eta_i)p(z,\eta_i)}d\mu - 2\right) \tag{A.30}$$

$$= -T\sum_{i=1}^{N}\int\left[\sqrt{p_{NT}(z,\eta_i)} - \sqrt{p(z,\eta_i)}\right]^2 d\mu =: -T\sum_{i=1}^{N} B_{NT}(\eta_i). \tag{A.31}$$

By Minkowski's inequality, we have

$$\left|(B_{NT}(\eta))^{\frac{1}{2}} - \left(\frac{1}{4}\mathbb{E}_{\theta,\eta} g^2(z,\eta)\right)^{\frac{1}{2}}\right| \leq \int r_{NT}^2(\eta)d\mu. \tag{A.32}$$

Thus the analogous arguements with the above show that as $N, T \to \infty$,

$$-T\sum_{i=1}^{N} B_{NT}(\eta_i) = -\frac{1}{4N}\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i} g^2(z,\eta_i) + o(1). \tag{A.33}$$

By Assumption **??**, we can see that $\eta \to \mathbb{E}_{\theta,\eta} g^2(z,\eta)$ is also continuous and bounded on Λ. Therefore, Assumption 4 implies that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i} g^2(z,\eta_i) = \int \mathbb{E}_{\theta,\eta} g^2(z,\eta)d\Gamma(\eta). \tag{A.34}$$

This shows that

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{N} W_{NT}(z_{it},\eta_i) - \frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N} g(z_{it},\eta_i)\right] = -\frac{1}{4}\int \mathbb{E}_{\theta,\eta} g^2(z,\eta)d\Gamma(\eta) + o(1). \tag{A.35}$$

Combining (A.35) and (A.27), we have the desired result. □

Lemma A.3. As $N, T \to \infty$, we have

$$\sum_{t=1}^{T}\sum_{i=1}^{N} W_{NT}^2(z_{it},\eta_i) = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} g^2(z_{it},\eta_i) + o_p(1). \tag{A.36}$$

Proof. By the Cauchy-Schwarz inequality and Minkowski's inequality, we have

$$\begin{aligned}
&\mathbb{E}\left|\sum_{t=1}^{T}\sum_{i=1}^{N} W_{NT}^2(z_{it},\eta_i) - \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} g^2(z,\eta_i)\right| \\
\le\ & \sum_{t=1}^{T}\sum_{i=1}^{N}\mathbb{E}\left|W_{NT}^2(z_{it},\eta_i) - \frac{1}{NT}g^2(z,\eta_i)\right| \\
\le\ & T\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i}\left|\left(W_{NT}(z_{it},\eta_i) - \frac{1}{\sqrt{NT}}g(z,\eta_i)\right)\left(W_{NT}(z_{it},\eta_i) + \frac{1}{\sqrt{NT}}g(z,\eta_i)\right)\right| \\
\le\ & T\sum_{i=1}^{N}\left(\mathbb{E}_{\theta,\eta_i}\left(W_{NT}(z_{it},\eta_i) - \frac{1}{\sqrt{NT}}g(z,\eta_i)\right)^2\right)^{1/2} \\
&\qquad\qquad \times\left(\mathbb{E}_{\theta,\eta_i}\left(W_{NT}(z_{it},\eta_i) + \frac{1}{\sqrt{NT}}g(z,\eta_i)\right)^2\right)^{1/2} \\
\le\ & T\sum_{i=1}^{N}\left(\mathbb{E}_{\theta,\eta_i}\left(W_{NT}(z_{it},\eta_i) - \frac{1}{\sqrt{NT}}g(z,\eta_i)\right)^2\right)^{1/2} \\
&\qquad\qquad \times\left(\left(\mathbb{E}_{\theta,\eta_i}W_{NT}^2(z_{it},\eta_i)\right)^{1/2} + \left(\frac{1}{NT}\mathbb{E}_{\theta,\eta_i}g^2(z,\eta_i)\right)^{1/2}\right) \\
\le\ & \left(T\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i}\left(W_{NT}(z_{it},\eta_i) - \frac{1}{\sqrt{NT}}g(z,\eta_i)\right)^2\right)^{1/2} \times\left(T\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i}W_{NT}^2(z_{it},\eta_i)\right)^{1/2} \\
&+\left(T\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i}\left(W_{NT}(z_{it},\eta_i) - \frac{1}{\sqrt{NT}}g(z,\eta_i)\right)^2\right)^{1/2} \times\left(\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i}g^2(z,\eta_i)\right)^{1/2}.
\end{aligned}$$

We know from the proof of the previous lemma that the first factors in the two terms in the last display are $o(1)$ and further that the second factor in the second term is $O(1)$. Now observe that

$$\begin{aligned}
T\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i}W_{NT}^2(z_{it},\eta_i) &= 4T\sum_{i=1}^{N}\int\left[\sqrt{p_{NT}(z,\eta_i)} - \sqrt{p(z,\eta_i)}\right]^2 d\mu \\
&= 4T\sum_{i=1}^{N}B_{NT}(\eta_i) \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i}g^2(z,\eta_i) + o(1) \\
&= O(1).
\end{aligned}$$

where the last two equalities follows from the results in the proof of the previous lemma (see (A.33) and (A.34)). Thus, as $N,T\to\infty$, we have $\mathbb{E}\left|\sum_{t=1}^{T}\sum_{i=1}^{N}W_{NT}^2(z_{it},\eta_i) - \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}g^2(z,\eta_i)\right| = o(1)$, which implies the desired result. □

Lemma A.4. For any $\epsilon>0$, we have as $N\to\infty$, regardless of whether T is fixed or tends to infinity,

$$\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i}g^2(z_{it},\eta_i)1\left\{|g(z_{it},\eta_i)| > \sqrt{NT}\epsilon\right\}\to 0. \tag{A.37}$$

Proof. This result follows from Lemma 6.3 in Strasser (1996b). □

Lemma A.5. As $N,T\to\infty$, we have

$$\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}g^2(z_{it},\eta_i)\xrightarrow{p}\int\mathbb{E}_{\theta,\eta}g^2(z,\eta)d\Gamma(\eta). \tag{A.38}$$

Proof. This lemma is an extention of the result in p.380 in Strasser (1985). The difference is that Strasser (1985) considers the case where only the number of individuals tends to infinity, while we consider double asymptotics. Thus our proof needs more delicate calculations.

We will show that as $N, T \to \infty$,

$$\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} g^2(z_{it}, \eta_i) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}_{\theta,\eta} g^2(z_{it}, \eta) + o_p(1), \tag{A.39}$$

which implies the required result by Assumption 4. To show this, fix $\epsilon > 0$ and $\delta > 0$. Take a positive number $\gamma > 0$ such that

$$\gamma^2 < \frac{\epsilon^2\delta}{18\sup_{N\in\mathbb{N}}(1/N)\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i} g^2(z_{it}, \eta_i)}. \tag{A.40}$$

For notational simplicity, we write $\mathbb{I}_{it,NT} := 1\left\{|g(z_{it}, \eta_i)| \leq \sqrt{NT}\gamma\right\}$. First observe taht

$$\begin{aligned} &\left|\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i} g^2(z_{it}, \eta_i) - \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i} g^2(z_{it}, \eta_i)\mathbb{I}_{it,NT}\right| \\ &\qquad \leq \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i} g^2(z_{it}, \eta_i) 1\left\{|g(z_{it}, \eta_i)| > \sqrt{NT}\gamma\right\} \to 0 \end{aligned} \tag{A.41}$$

as $N, T \to \infty$ by Lemma. This implies that there exists some $M_1 \in \mathbb{N}$ such that for all $N, T \geq M_1$ the left hand side of (A.41) is smaller than or equal to $\epsilon/3$. Thus for $N, T \geq M_1$ we have

$$\begin{aligned} &\mathbb{P}\left(\left|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} g^2(z_{it}, \eta_i) - \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i} g^2(z_{it}, \eta_i)\right| > \epsilon\right) \\ \leq\ & \mathbb{P}\left(\left|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} g^2(z_{it}, \eta_i) - \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} g^2(z_{it}, \eta_i)\mathbb{I}_{it,NT}\right| > \frac{\epsilon}{3}\right) \\ &+\mathbb{P}\left(\left|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} g^2(z_{it}, \eta_i)\mathbb{I}_{it,NT} - \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i} g^2(z_{it}, \eta_i)\mathbb{I}_{it,NT}\right| > \frac{\epsilon}{3}\right). \end{aligned} \tag{A.42}$$

Since

$$\mathbb{P}\left(\left|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} g^2(z_{it}, \eta_i) - \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} g^2(z_{it}, \eta_i)\mathbb{I}_{it,NT}\right| > \frac{\epsilon}{3}\right) \tag{A.43}$$

$$\leq \mathbb{P}\left(\max_{1\leq t\leq T}\max_{1\leq i\leq N} g^2(z_{it}, \eta_i) > \sqrt{NT}\gamma\right) \tag{A.44}$$

$$\leq \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i} g^2(z_{it}, \eta_i) 1\left\{|g(z_{it}, \eta_i)| > \sqrt{NT}\gamma\right\}, \tag{A.45}$$

there exists some $M_2 \in \mathbb{N}$ such that, for all $N, T \geq M_2$, the first term of the right hand side of (A.42) is smaller than to $\delta/2$. Further

$$\mathbb{P}\left(\left|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} g^2(z_{it}, \eta_i)\mathbb{I}_{it,NT} - \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i} g^2(z_{it}, \eta_i)\mathbb{I}_{it,NT}\right| > \epsilon\right) \tag{A.46}$$

$$\leq \frac{9}{\epsilon^2 N^2 T}\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i} g^4(z_{it}, \eta_i)\mathbb{I}_{it,NT} \leq \frac{9\gamma^2}{\epsilon^2 N}\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i} g^2(z_{it}, \eta_i) < \frac{\delta}{2}. \tag{A.47}$$

Cosequently, for all $N, T \geq M := \max\{M_1, M_2\}$, we have

$$\mathbb{P}\left(\left|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N} g^2(z_{it}, \eta_i) - \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i} g^2(z_{it}, \eta_i)\right| > \epsilon\right) < \delta. \tag{A.48}$$

This shows the lemma. □

Lemma A.6. As $N, T \to \infty$, we have

$$\sum_{t=1}^{T}\sum_{i=1}^{N} W_{NT}^2(z_{it},\eta_i)R(W_{NT}(z_{it},\eta_i)) = o_p(1) \tag{A.49}$$

Proof. First note that

$$\left|\sum_{t=1}^{T}\sum_{i=1}^{N} W_{NT}^2(z_{it},\eta_i)R(W_{NT}(z_{it},\eta_i))\right| \quad \leq \quad \max_{1\leq t\leq T}\max_{1\leq i\leq N}\left|R\left(W_{NT}(z_{it},\eta_i)\right)\right|\sum_{t=1}^{T}\sum_{i=1}^{N} W_{NT}^2(z_{it},\eta_i).$$

By Lemma A.3 and Lemma A.5, we have $\sum_{t=1}^{T}\sum_{i=1}^{N} W_{NT}^2(z_{it},\eta_i) = O_p(1)$ as $N, T \to \infty$. Thus it suffices to show that as $N, T \to \infty$,

$$\max_{1\leq t\leq T}\max_{1\leq i\leq N}|R(W_{NT}(z_{it},\eta_i))| = o_p(1). \tag{A.50}$$

Now observe that

$$\mathbb{P}\left(\max_{1\leq t\leq T}\max_{1\leq i\leq N}|W_{NT}(z_{it},\eta_i)| > \sqrt{2}\epsilon\right) \tag{A.51}$$

$$\leq \quad T\sum_{i=1}^{N}\mathbb{P}\left(W_{NT}^2(z_{it},\eta_i) > 2\epsilon^2\right) \tag{A.52}$$

$$\leq \quad T\sum_{i=1}^{N}\mathbb{P}\left(g^2(z_{it},\eta_i) > \epsilon^2 NT\right) + T\sum_{i=1}^{N}\mathbb{P}\left(|A_{it}| > \epsilon^2 NT\right) \tag{A.53}$$

where $A_{it} := NTW_{NT}^2(z_{it},\eta_i) - g^2(z,\eta_i)$. Note that from the result in the previous lemma we have $(1/N)\sum_{i=1}^{N}\mathbb{E}|A_{it}| \to 0$ as $N \to \infty$. By Markov's inequality, we have

$$\mathbb{P}\left(\max_{1\leq t\leq T}\max_{1\leq i\leq N}|W_{NT}(z_{it},\eta_i)| > \sqrt{2}\epsilon\right) \tag{A.54}$$

$$\leq \quad \frac{1}{\epsilon^2 N}\sum_{i=1}^{N}\mathbb{E}_{\theta,\eta_i}g^2(z_{it},\eta_i)1\left\{g^2(z_{it},\eta_i) > \epsilon T\right\} + \frac{1}{\epsilon^2 N}\sum_{i=1}^{N}\mathbb{E}|A_{it}|. \tag{A.55}$$

From the result in Lemma the first term of the last display tend to 0. Thus we see that as $N, T \to \infty$

$$\max_{1\leq t\leq T}\max_{1\leq i\leq N}|W_{NT}(z_{it},\eta_i)| = o_p(1). \tag{A.56}$$

Next we show that (A.56) implies (A.50). Since $R(x) \to 0$ as $x \to 0$, for $\epsilon > 0$ there exists a number $\delta(\epsilon) > 0$ such that if $|x| \leq \delta(\epsilon)$, then $|R(x)| \leq \epsilon$. Thus,

$$\left\{\max_{1\leq i\leq N}|W_{NT}(z_{it},\eta_i)| \leq \delta(\epsilon)\right\} \subseteq \left\{\max_{1\leq i\leq N}|R(W_{NT}(z_{it},\eta_i))| \leq \epsilon\right\}. \tag{A.57}$$

Since this holds for each t, we have

$$\left\{\max_{1\leq t\leq T}\max_{1\leq i\leq N}|W_{NT}(z_{it},\eta_i)| \leq \delta(\epsilon)\right\} \subseteq \left\{\max_{1\leq t\leq T}\max_{1\leq i\leq N}|R(W_{NT}(z_{it},\eta_i))| \leq \epsilon\right\}. \tag{A.58}$$

Taking complements and using the subadditivity of a measure, we have

$$\mathbb{P}\left(\max_{1\leq t\leq T}\max_{1\leq i\leq N}|R(W_{NT}(z_{it},\eta_i))| > \epsilon\right) \leq \mathbb{P}\left(\max_{1\leq t\leq T}\max_{1\leq i\leq N}|W_{NT}(z_{it},\eta_i)| > \delta(\epsilon)\right). \tag{A.59}$$

Since the right hand side tends to 0 as $N, T \to \infty$, we obtain the desired result. □

Lemma A.7. As $N, T \to \infty$, we have

$$\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N} g(z_{it},\eta_i) \xrightarrow{d} N\left(0, \int \mathbb{E}_{\theta,\eta}g^2(z_{it},\eta)d\Gamma(\eta)\right). \tag{A.60}$$

Proof. This follows by the same arguments as in Lemma 6.3 in Strasser (1998).

□

Proof of Theorem 4.2

Just combine the above results. □

References

[1] Bai, J. (2009) Panel data models with interactive fixed effects, *Econometrica* **77**(4), 1229-1279.

[2] Bickel, P. J. & C. A. L. Klaassen (1986) Empirical Bayes estimation in functional and structural models, *Advances in Applied Mathematics*, 55–69.

[3] Bickel, P. J., C. A. J. Klassen, Y. Ritov & J. A. Wellner (1993) *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press.

[4] Canay, I. A. (2011) A simple approach to quantile regression for panel data, *The Econometrics Journal* **14**(3) 368–386.

[5] arro, J. M. (2006) Estimating dynamic panel data discrete choice models with fixed effects, *Journal of Econometrics*, **140**(2) 503–528.

[6] Fernandez-Val, I. (2009) Estimation of structural parameters and marginal effects in binary choice panel data models with fixed effects, *Journal of Econometrics* 1**50**, 71–85.

[7] Fernandez-Val, I. & F. Vella (2011) Bias corrections for two-step fixed effects panel data estimators, *Journal of Econometrics*, **163**(2) 144–162.

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

[9] Hahn, J. (2002) Optimal inference with many instruments, *Econometric Theory* **18**, 140-168.

[10] Hahn, J. & G. Kuersteiner (2012) Bias reduction for dynamic nonlinear panel models with fixed effects, *Econometric Theory* **27**, 1152-1191.

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

[12] Kato, K. & A.F. Galvao, (2010) Smoothed Quantile Regression for Panel Data, Mimeo.

[13] Koenker, R. (2004) Quantile regression for longitudinal data, *Journal of Multivariate Analysis* **91**, 74–89.

[14] McNeney, B. & J. A. Wellner (2000) Application of convolution theorems in semiparametric models with non-i.i.d. data, *Journal of Statistical Planning and Inference* **91** 441–480.

[15] Newey,W. K. (1990) Semiparametric efficiency bounds, *Journal of Applied Econometrics* **5**, 99–135.

[16] Neyman, J., & E. L. Scott (1948) Consistent estimates based on partially consistent observations *Econometrica*, **16**, 1–32.

[17] Pfanzagl, J. (1993) Incidental versus random nuisance parameters, *Annals of Statistics* **21**, 1663–1691.

[18] Phillips, P. C. B. & H. R. Moon (1999) Linear regression limit theory for nonstationary panel data, *Econometrica* **67**(5), 1057–1111.

[19] Rudin, W. (1973) *Functional Analysis*, McGraw-Hill, New York

[20] trasser, H. (1985) *Mathematical Theory of Statistics*, W. de Gruyter, Berlin.

[21] Strasser, H. (1996) Asymptotic efficiency of estimates for model with incidental nuisance parameters, *Annals of Statistics* **24**(2), 879-901.

[22] Strasser, H. (1998) Perturbation invariant estimates and incidental nuisance parameters, *Mathematical Methods of Statistics* **7**, 1-26.

[23] van der Vaart, A. W. (1998) *Asymptotic Statistics*, Cambridge University Press, New York.

van der Vaart, A. W. (2002) Semiparametric statistics, in *Lectures on probability theory and statistics (Saint-Flour, 1999), Vol. 1781 of Lecture Notes in Mathematics*, Springer Verlag, 331–457.

[24] van der Vaart, A. W. & Wellner, J. A. (1996) *Weak Convergence and Empirical Processes with Applications to Statistics*, Springer Verlag, New York.

[25] Woutersen, T. M. (2002) Robustness against incidental parameters, mimemo.