# AN ALTERNATIVE ROOT-N CONSISTENT ESTIMATOR FOR PANEL DATA BINARY CHOICE MODELS

BY CHUNRONG AI[1] AND LI GAN[2]

In this paper, we present an alternative root-n consistent estimator for panel data fixed-effects binary choice models. The proposed estimator relaxes one of the key conditions that are required for the consistency of the estimator proposed in Honoré and Lewbel (2002), and is shown to be consistent and asymptotically normally distributed under some sufficient conditions. An easy to compute consistent estimator for the asymptotic covariance is provided.

KEYWORDS: Binary choice; panel data; root-n consistency.

## 1 Introduction

THIS PAPER CONSIDERS estimation of the following panel data binary choice model:

$$y_{it} = 1(v_{it}\lambda_o + x_{it}^{'}\beta_o + \alpha_i - \varepsilon_{it} > 0),\ i = 1, 2, ...n;\ t = 1, 2, ..., T, \quad (1)$$

where $1(\cdot)$ denotes the indicator function equal to one when $\cdot$ is true and to zero otherwise, $v_{it}$ is a regressor with a nonzero coefficient $\lambda_o$, $x_{it}$ is a $J \times 1$ vector of regressors with unknown coefficients

[1] Chunrong Ai, Department of Economics, University of Florida, Gainesville, FL 32611, USA, and Shanghai University of Finance and Economics.
[2] Li Gan, Department of Economics, Texas A&M University, College Station, TX 77843, USA. E-mail: gan@econmail.tamu.edu. Tel: (979) 862-1667.
Both authors thank two anonymous referees and a coeditor for helpful comments. The first author acknowledges the Shanghai University of Finance and Economics for providing partial financial support under the grant "211 project phase III".

$\beta_o$, $\alpha_i$ is the individual specific effect, and $\varepsilon_{it}$ is the error term of the model. The subscript $i$ indexes individual and $t$ indexes time. Estimation of the unknown parameter $\theta_o = (\lambda_o, \beta_o)$ of this model has been studied by several researchers under various assumptions on the error term, the individual effect, and the regressors. For instance, Rasch (1960) and Andersen (1970) show that $\theta_o$ can be estimated consistently at the rate $\sqrt{n}$ by a conditional likelihood approach if the error term has logistic distribution and is independent of the regressors and the individual effect. Their estimator does not require specifying the joint distribution of the regressors and the individual effect, and hence is a fixed-effects estimator. Manski (1987) presents a conditional maximum score approach under the assumptions: (i) the error term is stationary conditional on the regressors and the individual effect and (ii) $v_{it}$ is a continuous variable. Manski's estimator does not require specifying the error term distribution; nor does it require specifying the joint distribution of the regressors and the individual effect. Thus, Manski's estimator is a fixed-effects semiparametric estimator. However, Manski's estimator is shown to be less than $\sqrt{n}$ consistent and its asymptotic distribution is unknown. The difficulty to establish the statistical properties of Manski's estimator is often attributed to the non-smoothness of his score function. Horowitz (1992) modifies Manski's estimator by smoothing the score function and shows that the modified estimator can have a convergence rate arbitrarily close but never equal to $\sqrt{n}$. Honoré and Kyriazidou (2000) extend the conditional likelihood approach and the conditional maximum score approach to the case where the regressors include the lagged dependent variables. Their estimator, like Manski's estimator, is also less than $\sqrt{n}$ consistent.[3]

In light of these results, one may wonder if it is possible at all to obtain a $\sqrt{n}$ consistent estimator for panel data fixed-effects binary choice models besides the panel data fixed-effects logit model. Chamberlain (1993) gives a partial answer to this question. He shows that, even if the

[3] For other estimators for panel fixed-effect binary choice models, see Lee (1999,2001), Hahn (2001) and Hahn and Newey (2004).

error term is independent of the regressors and the individual effect and its distribution is known, the logit model is the only model where it is possible to achieve $\sqrt{n}$ consistency. An important implication of this result is that $\sqrt{n}$ consistency cannot be obtained through restrictions on the distribution of the error term only. If $\sqrt{n}$ consistency is achievable at all, it must be achieved through restrictions on the joint distribution of the regressors and the individual effect. Following exactly this approach, Honoré and Lewbel (2002) propose a $\sqrt{n}$ consistent estimator. Their estimator imposes a restriction on the joint distribution of the regressors and the individual effect through a "special regressor", say $v_{it}$. The "special regressor" is required to satisfy three conditions. First, the "special regressor" is a continuous variable and has a positive coefficient $\lambda_o$. Second, the "special regressor" is independent of the individual effect and the error term conditional on the other regressors (e.g., $x_{it}$) and some instrumental variables $w_{it}$. The instrumental variables here can include individual specific characteristics that do not vary over time, such as gender. Third, conditional on the regressors, the instrumental variables, the individual effect, and the error term, the "special regressor" $v_{it}$ has a support $(\underline{v}_t, \overline{v}_t)$ that is large enough to contain the support of $\varepsilon_{it} - x_{it}^{'}\beta_o - \alpha_i$, and that $\overline{v}_t$ is bounded and finite. The first condition allows us to integrate out the "special regressor" while the second condition permits weighting data through density weighting so that the weighted expectation is easily computable. Specifically, let $f_t(v_{it}|x_{it}, w_{it})$ denote the conditional density of $v_{it}$ given $(x_{it}, w_{it})$ and, without loss of generality, normalize the coefficient $\lambda_o$ to one. Then, under the first and the second condition, simple calculations give

$$\begin{aligned} E\left\{\frac{y_{it}}{f_t(v_{it}|x_{it}, w_{it})}|x_{it}, w_{it}, \alpha_i, \varepsilon_{it}\right\} &= 1\left\{\varepsilon_{it} - x_{it}^{'}\beta_o - \alpha_i \leqslant \underline{v}_t\right\} + \\ & \quad 1\left\{\underline{v}_t < \varepsilon_{it} - x_{it}^{'}\beta_o - \alpha_i \leqslant \overline{v}_t\right\} * (\overline{v}_t - \varepsilon_{it} + x_{it}^{'}\beta_o + \alpha_i). \end{aligned}$$

The third condition on the "special regressor" guarantees that $1\left\{\underline{v}_t < \varepsilon_{it} - x_{it}^{'}\beta_o - \alpha_i \leqslant \overline{v}_t\right\} = 1$

holds with probability one; this in turn implies that

$$E\left\{\frac{y_{it}}{f_t(v_{it}|x_{it},w_{it})}|x_{it},w_{it},\alpha_i,\varepsilon_{it}\right\}=\overline{v}_t-\varepsilon_{it}+x_{it}^{'}\beta_o+\alpha_i.$$

Notice that $\overline{v}_t=E\left\{\frac{1\{v_{it}>0\}}{f_t(v_{it}|x_{it},w_{it})}|x_{it},w_{it},\alpha_i,\varepsilon_{it}\right\}$. It follows that

$$E\left\{\frac{y_{it}-1\{v_{it}>0\}}{f_t(v_{it}|x_{it},w_{it})}|x_{it},w_{it},\alpha_i,\varepsilon_{it}\right\}=-\varepsilon_{it}+x_{it}^{'}\beta_o+\alpha_i.$$

Thus, the "special regressor" allows us to transform a nonlinear regression problem into a linear regression problem through integrating out the "special regressor". The preceding equation gives a linear panel data model with $\frac{y_{it}-1\{v_{it}>0\}}{f_t(v_{it}|x_{it},w_{it})}$ as the dependent variable. Replacing the unknown density $f_t(v_{it}|x_{it},w_{it})$ with a consistent nonparametric estimator and applying standard panel regression techniques after differencing out the individual effect, Honoré and Lewbel show that such procedure yields a $\sqrt{n}$ consistent and asymptotically normally distributed estimator.[4]

The three conditions on the "special regressor" are critical for Honoré and Lewbel's procedure. Without any one of the three conditions, their estimator might be inconsistent. Thus it is useful to discuss briefly these conditions. The first condition on the "special regressor" is less restrictive than it appears. For models of this sort, it is known that the unknown parameter cannot be identified if all regressors are discrete when the error term distribution is not parameterized. For identification purpose, at least one regressor must be continuous variable. Thus, the first condition is part of an identification condition and must be imposed. The positive coefficient is not restrictive. If the "special regressor" has a negative coefficient, all we need to do is to define $-v_{it}$ as the special regressor. In practice, we do not know what kind of coefficient that $v_{it}$ has. But we can

[4]Their procedure is actually slightly more general than the one here, allowing for the special regressor to be endogenous/predetermined. See their paper for details.

apply the technique developed in Ghosal et al. (2000) to test the significance and the sign of $\lambda_o$. The second condition requires conditional independence between the "special regressor" and the error term and between the "special regressor" and the individual effect. The first part of this condition is not unusual since the error term is often assumed to be independent of the regressors in this model. The second part of the condition, however, is usually not assumed in the fixed-effects model. Nevertheless, this part of the condition is not too difficult to satisfy since in many applications personal characteristics such as age can be argued to be independent of the error term and the individual effect once the other characteristics are controlled for. Besides, the instrumental variables help picking up some of the dependence between the "special regressor" and the individual effect. The third condition on the support of the "special regressor", however, is more difficult to justify, especially since the error term and the individual effect are not observed; and there are no reasons to expect their support to be smaller than the support of the "special regressor". In addition, this condition rules out widely used probit and logit models. Thus, it is important to find an alternative estimator that does not impose such stringent restriction on the support of the "special regressor".

The main objective of this paper is to present such an alternative estimator. Our estimator does not require the third condition on the support of the error term and the other regressors; keeps the first condition; but strengthens the second condition. The strengthened second condition requires that the error term $\varepsilon_{it}$ is independent of all regressors, the instruments, and the individual effect with unknown cumulative distribution function $F(\varepsilon)$. To explain why the third condition is not needed when the strengthened second condition is imposed, notice that, for any $s > t$, the strengthened second condition implies

$$E\{y_{it}|v_{it}, x_{it}, w_{it}\} = \int F(v_{it} + x'_{it}\beta_o + \alpha)g_t(\alpha|x_{it}, w_{it}) \times d\alpha = m_t(v_{it} + x'_{it}\beta_o, x_{it}, w_{it}),$$

$$E\{y_{is}|v_{is}, x_{is}, w_{is}\} = \int F(v_{is} + x_{is}'\beta_o + \alpha)g_s(\alpha|x_{is}, w_{is}) \times d\alpha = m_s(v_{is} + x_{is}'\beta_o, x_{is}, w_{is}),$$

where $g_j(\alpha|x_{ij}, w_{ij})$ is the conditional density of the individual effect $\alpha_i$ given the regressors $x_{ij}$ and the instruments $w_{ij}$. The preceding equation implies an index regression with $y_{ij}$ as dependent variable, $v_{ij} + x_{ij}'\beta_o$ as the index, and $m_j(.)$ as unknown positive and monotonically increasing (in its first argument) function. Unfortunately, this index regression does not identify $\beta_o$. To see this, we write

$$m_j(v_{ij} + x_{ij}'\beta_o, x_{ij}, w_{ij}) = m_j(v_{ij} + x_{ij}'\beta + x_{ij}'(\beta_o - \beta), x_{ij}, w_{ij}) = \widetilde{m}_j(v_{ij} + x_{ij}'\beta, x_{ij}, w_{ij}), j = t, s$$

where $\widetilde{m}_j(v_{ij}, x_{ij}, w_{ij}) = m_j(v_{ij} + x_{ij}'(\beta_o - \beta), x_{ij}, w_{ij})$ is clearly a positive and monotonically increasing function of the index $v_{ij} + x_{ij}'\beta$. This means that $(m_j, \beta_o)$ and $(\widetilde{m}_j, \beta)$ cannot be distinguished for any $\beta$ using only cross sectional information, and hence the true value $\beta_o$ is not identified. To identify $\beta_o$, we must use the restriction across time periods. To illustrate, let $z_{it,s}$ denote the union of $x_{it}, x_{is}, w_{it}, w_{is}$. Notice that the functional form of the conditional density of the individual effect given $z_{it,s}$, denoted by $g_{t,s}(\alpha|z_{it,s})$ does not depend on time $j = t, s$ but depends on the pair $(t, s)$. Write

$$(2) \qquad m_{t,so}(v_{ij} + x_{ij}'\beta_o, z_{it,s}) = E\{y_{ij}|v_{ij}, z_{it,s}\} = \int F(v_{ij} + x_{ij}'\beta_o + \alpha)g_{t,s}(\alpha|z_{it,s}) \times d\alpha, .j = t, s$$

$m_{t,so}$ is still a positive and monotonically increasing function of the index $v_{ij} + x_{ij}'\beta_o$. The functional form of $m_{t,so}$ does not vary with $j = t, s$. This turns out to be the key for identifying $\beta_o$ (see Section 2 for details).

The index equation (2) suggests a semiparametric nonlinear least squares regression where the unknown function $m_{t,so}$ will be either estimated nonparametrically or approximated arbitrarily by sieve method. In this paper, we adopt the method of sieve.[5] Details of the method of sieve and the proposed estimator will be described in Section 3. The asymptotic properties of the estimator will be derived in Section 4.

## 2 Identification

Before introducing our estimation procedure, we show that index restriction (2) identifies the true value $(m_{t,so}, \beta_o)$. Notice that $m_{t,so}$ is a probability function and must satisfy the shape restriction of the probability function. We write $m_{t,so} = \Phi(h_{t,so})$ for some $h_{t,so}$, where $\Phi$ is the standard cumulative distribution function. Such reparameterization is useful since $h_{t,so}$ does not have to satisfy the same shape restriction. Indeed, any known cumulative distribution function would have been used here just as well. For any pair $s > t$, let $\Delta_{t,s}$ denote the first difference operator between period $s$ and period $t$. Let $\mathcal{B}$ denote the finite dimensional parameter space that contains the true value $\beta_o$. Let $\mathcal{X}_t$ denote the support of $x_{it}$ and let $\mathcal{Z}_{t,s}$ denote the support of $z_{it,s}$. Define $\underline{v}_{t,s} = \min_{j=t,s;\beta\in\mathcal{B};x\in\mathcal{X}_j}\{\underline{v}_j + x'\beta\}$ and $\overline{v}_{t,s} = \max_{j=1,2;\beta\in\mathcal{B};x\in\mathcal{X}_j}\{\underline{v}_j + x'\beta\}$. Let $\mathcal{H}_{t,s}$ denote the space of the unknown functions mapping $(\underline{v}_{t,s}, \overline{v}_{t,s}) \times \mathcal{Z}_{t,s}$ into $R$ and suppose that $\mathcal{H}_{t,s}$ contains the true value $h_{t,so}$. Suppose that the following conditions are satisfied.

**Assumption 2.1.** *(i) For $j = 1, 2, ..., T$, $v_{ij}$ is a continuous random variable with support $(\underline{v}_j, \overline{v}_j)$ and has a positive effect on $y_{ij}$; for any pair $(t, s)$ and $j = t, s$: (ii) the error term $\varepsilon_{ij}$ is independent of $(v_{ij}, z_{it,s}, \alpha_i)$ with the monotonically increasing cumulative distribution function $F(\varepsilon)$ and (iii) conditional on $z_{it,s}$, $v_{ij}$ is independent of $\alpha_i$; and for some pair $(t, s)$: (iv)*

[5] Ai (1997) studies a maximum likelihood estimation of index model.

$\Pr\left(\underline{v}_t - \overline{v}_s < \Delta_{t,s} x_i' \beta_o < \overline{v}_t - \underline{v}_s\right) > 0$ *and (v)* $E\left\{(x_{ks} - x_{it})(x_{ks} - x_{it})' | (v_{it} + x_{it}'\beta_o, z_{it,s}) = (v_{ks} + x_{ks}'\beta_o, z_{kt,$

*is nonsigular.*

Under this condition, notice that, for any $\beta \in \mathcal{B}$ and $h_{t,s} \in \mathcal{H}_{t,s}$, Assumption 2.1(ii) implies

$$\begin{aligned} E\left\{\left(y_{ij} - \Phi(h_{t,s}(v_{ij} + x_{ij}'\beta, z_{it,s}))\right)^2\right\} &= E\left\{\left(y_{ij} - m_{t,so}(v_{ij} + x_{ij}'\beta_o, z_{it,s})\right)^2\right\} + \\ & \quad E\left\{\left(m_{t,so}(v_{ij} + x_{ij}'\beta_o, z_{it,s}) - \Phi(h_{t,s}(v_{ij} + x_{ij}'\beta, z_{it,s}))\right)^2\right\} \end{aligned}$$

which is minimized when $\beta = \beta_o$ and $h_{t,s} = h_{t,so}$. Suppose that there exist other $\beta \in \mathcal{B}$ and $h_{t,s} \in \mathcal{H}_{t,s}$ that minimize the above population criterion function. Then we must have

$$h_{t,so}(v_{ij} + x_{ij}'\beta_o, z_{it,s}) = h_{t,s}(v_{ij} + x_{ij}'\beta, z_{it,s}), \text{ for } j = t, s.$$

Assumption 2.1(i)(ii) imply that $h_{t,so}$ is monotonically increasing in its first argument. The preceding equation says that $h_{t,s}$ must also be monotonically increasing in its first argument. Rewriting the above equations with $v = v_{ij} + x_{ij}'\beta_o$, we have

$$h_{t,so}(v, z_{it,s}) = h_{t,s}(v + x_{ij}'(\beta - \beta_o), z_{it,s}), \text{ for } j = t, s.$$

This means that we have two definitions for the function $h_{t,so}(v, z)$. For period $t$, we use

$$h_{t,so}(v, z) = h_{t,s}(v + x_{it}'(\beta - \beta_o), z).$$

For period $s$, we use

$$h_{t,so}(v, z) = h_{t,s}(v + x_{is}'(\beta - \beta_o), z).$$

Since the function $h_{t,so}(v,z)$ must have only one value for each $(v,z)$, individuals with the same explanatory variables must have the same function value regardless of the time periods they are in. Specifically, if individual $i$ in period $t$ and individual $k$ in period $s$ have the same values: $(v_{it} + x'_{it}\beta_o, z_{it,s}) = (v_{ks} + x'_{ks}\beta_o, z_{kt,s})$, then we must have

$$h_{t,so}(v_{it} + x'_{it}\beta_o, z_{it,s}) = h_{t,so}(v_{ks} + x'_{ks}\beta_o, z_{kt,s})$$

which in turn implies

$$h_{t,s}(v + x'_{it}(\beta - \beta_o), z_{it,s}) = h_{t,s}(v + x'_{is}(\beta - \beta_o), z_{it,s})$$

for all values of the regressors satisfying $v_{it} + x'_{it}\beta_o = v_{ks} + x'_{ks}\beta_o = v$. The monotonicity of $h_{t,s}$ implies

$$(x_{ks} - x_{it})'(\beta - \beta_o) = 0.$$

If there are enough individuals from both periods satisfying $(v_{it} + x'_{it}\beta_o, z_{it,s}) = (v_{ks} + x'_{ks}\beta_o, z_{kt,s})$ so that

$$E\left\{(x_{ks} - x_{it})(x_{ks} - x_{it})' | (v_{it} + x'_{it}\beta_o, z_{it,s}) = (v_{ks} + x'_{ks}\beta_o, z_{kt,s})\right\}$$

is nonsingular, then we have $\beta = \beta_o$, which in turn implies $h_{t,s} = h_{t,so}$ for any pair $(t,s)$.

**Lemma 2.1.** *Under Assumption 2.1,*

$$\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s} E\left\{\left(y_{ij} - \Phi(h_{t,s}(v_{ij} + x'_{ij}\beta, z_{it,s}))\right)^2\right\} \tag{3}$$

*is minimized over* $\beta \in \mathcal{B}$ *and* $h_{t,s} \in \mathcal{H}_{t,s}$ *if and only if* $\beta = \beta_o$ *and* $h_{t,s} = h_{t,so}$.

It is worth pointing out that Lemma 2.1 still holds true even if the error term depends on the individual effect and other regressors $z_{it,s}$, provided that the form of the conditional distribution function does not depend on the time. Thus, our procedure allows for some form of heteroskedasticity. In fact, Lemma 2.1 still holds true even if the error term and the individual effect depend on the "special regressor", as long as the conditional mean function is monotone in $v_{it}$ and its functional form does not vary with time.

# 3 Estimator

The population criterion function (3) suggests a least squares regression for estimation of the model parameter $\beta_o$. The problem with the least squares regression is that the unknown function is infinite dimensional and cannot be estimated consistently from finite data points. This diiffculty can be overcome with the method of sieve (see Shen (1997) and Ai and Chen (2003) for other applications of the method of sieve). The method of sieve simply replaces the unknown function with a finite dimensional series approximation, then treats the series approximation as if it is the correct specification and applies the nonlinear least squares regression. To ensure that the approximation error does not bias the parameter estimate, the method of sieve requires the dimension of the series approximation to grow with sample sizes so that the approximation error tends to zero as sample sizes go to infinity. For a sample of observations

$$\{(y_{it}, v_{it}, x_{it}, w_{it}), i = 1, 2, ...., n; t = 1, 2, ..., T\},$$

we now describe the method of sieve.

## 3.1 Sieve NL regression

Recall that $z_{is,t}$ is the union of $x_{it}, x_{is}, w_{it}, w_{is}$. Denote $\underline{v} = \min_{t,s}\{\underline{v}_{t,s}\}$, $\overline{v} = \max_{t,s}\{\overline{v}_{t,s}\}$, and $\mathcal{Z} = \cup_{t,s}\mathcal{Z}_{t,s}$. Let

$$p^k(v,z) = (p_1(v,z), ....., p_k(v,z))'$$

denote known basis functions that are defined over the support $(\underline{v}, \overline{v}) \times \mathcal{Z}$. The known basis functions can approximate any square integrable function of $(v, z)$ arbitrarily well as $k \to \infty$. Commonly used basis functions include power series, Fourier series, Hermite polynomials, splines, wavelets, and neural networks. From the asymptotics point of view, all of these series basis functions work equally well provided some sufficient conditions given below are satisfied. Thus, there is no preference given to a particular series, though in practice splines tend to work better than the other series basis functions. Once a particular class of series basis functions is chosen, we approximate any function $h_{t,s} \in \mathcal{H}_{t,s}$ with:

$$h_{t,sk}(v,z) = p^k(v,z)'\pi_{t,s},$$

where $\pi_{t,s}$ denote the unknown sieve coefficients that will be estimated jointly with $\beta_o$ and $k$ is some known integer. The integer $k$ is called the smoothing parameter; and it must be given before the method of sieve is applied. Although several approaches have been suggested to select the smoothing parameter from a given sample, we assume $k$ is known and satisfy the conditions given below. Once $k$ is chosen, the method of sieve treats $h_{t,sk}(v,z)$ as if it is the correct specification of $h_{t,s0}(v,z)$ and proceeds to apply the nonlinear least squares regression. To ensure consistency of the estimator of the model parameter, the method of sieve requires that $h_{t,sk}(v,z)$ converges to $h_{t,s}(v,z)$ under some metric $\|\cdot\|_s$ as $k \to \infty$. Examples of the metric include the sup norm $\|h(v,z)\|_s = \sup_{v,z}|h(v,z)|$ and the $L_2$ norm $\|h(v,z)\|_s = \sqrt{\int h(v,z)^2\phi(v,z)dvdz}$, where $\phi$ is a known weighting function. The sieve space is defined as

$$\mathcal{H}_{t,sk} = \left\{ h_k(v,z) = p^k(v,z)'\pi : \pi'\pi \leq C \right\}$$

for some constant $C$.

The sieve nonlinear least squares estimator (hereafter denoted by SNL) $(\widehat{\beta}, \widehat{\pi}) = (\widehat{\beta}, \widehat{\pi}_{t,s}, 1 \leq t \leq T-1; t < s \leq T)$ is defined as

$$(4) \qquad \begin{aligned} (\widehat{\beta}, \widehat{\pi}) &= \arg \min_{\beta \in \mathcal{B}, h_{s,t} \in \mathcal{H}_{s,tk}} \sum_{i=1}^{n} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \sum_{j=t,s} \left( y_{ij} - \Phi(h_{t,s}(v_{ij} + x'_{ij}\beta, z_{it,s})) \right)^2 \\ &= \arg \min_{\beta \in \mathcal{B}, \pi'_{t,s}\pi_{t,st} \leq C} \sum_{i=1}^{n} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \sum_{j=t,s} \left( y_{ij} - \Phi(p^k(v_{ij} + x'_{ij}\beta, z_{it,s})'\pi_{t,s}) \right)^2 . \end{aligned}$$

And the unknown function is estimated by $\widehat{h}_{t,s}(v,z) = p^k(v,z)'\widehat{\pi}_{t,s}$. Under some sufficient conditions, we show that the SNL estimator $\widehat{h}_{t,s}$ is consistent and $\widehat{\beta}$ is asymptotically normally distributed.

# 4 Asymptotics

In this and the next section, we derive the asymptotic properties of the proposed SNL estimator $(\widehat{\beta}, \widehat{h}_{t,s})$. We begin by resenting sufficient conditions under which we show that the proposed estimator is consistent, and compute their convergence rates. We then establish the asymptotic distribution of $\widehat{\beta}$, and provide an easy to compute covariance matrix. Notice that the proposed SNL estimator is a special case of the modified SMD estimator proposed in Ai and Chen (2007, equation (5) with no endogenous regressors and no unconditional moment restrictions; hereafter AC). We derive the asymptotic properties of the SNL estimator by verifying AC's conditions. The following assumption implies Assumption 3.1 in AC.

**Assumption 4.1.** *(i) For each $t$, $\{(y_{it}, v_{it}, x_{it}, w_{it}), i = 1, 2, ...., n\}$ is an independent and identi-*

*cally distributed random sample; (ii) for each pair $(t, s)$, the support of $(v_{it}, z_{it,s})$ is compact with nonempty interior; (iii) the density of $(v_{it}, z_{it,s})$ is bounded and bounded away from zero.*

Assumption 4.1 permits time series dependency but rules out cross-sectional dependency. The cross-sectional independence assumption, however, is not critical for the asymptotic results derived below. Our main results can still be proved for cross-sectionally dependent data by employing the technique developed in Chen and Shen (1998).

To verify other conditions in AC, we need a metric under which the estimator will be shown to be consistent. There are many metric (e.g., sup metric, $L_2$ metric, etc.) to choose from. We choose the one that is convenient for derivation of the asymptotic distribution. Let $\|\cdot\|_E$ denote the Euclidean norm. Denote

$$\mathcal{H} = \prod_{1 \le t < T, s > t} \mathcal{H}_{t,s} \text{ and } \mathcal{H}_k = \prod_{1 \le t < T, s > t} \mathcal{H}_{t,sk}.$$

Denote the parameter space by A$= \mathcal{B} \times \mathcal{H}$ with $\alpha_0 = (\beta_0, h_0)$ and denote the sieve approximating space by $\mathcal{A}_k = \mathcal{B} \times \mathcal{H}_k$. For any $\alpha = (\beta, h) \in \mathcal{A}$ and $\overline{\alpha} = (\overline{\beta}, \overline{h}) \in \mathcal{A}$, the following metric is adopted in this paper:

$$\begin{aligned} \|\alpha - \overline{\alpha}\|_s &= \left\|\beta - \overline{\beta}\right\|_E + \sum_{1 \le t < T, s > t} \sup_{v,z} \left|h_{t,s}(v,z) - \overline{h}_{t,s}(v,z)\right| \\ &\quad + \sum_{1 \le t < T, s > t} \sup_{v,z} \left|\frac{\partial h_{t,s}(v,z)}{\partial v} - \frac{\partial \overline{h}_{t,s}(v,z)}{\partial v}\right|. \end{aligned}$$

Let $N(\varepsilon, \mathcal{A}_k, \|\cdot\|_s)$ denote the total number of balls with radius $\varepsilon$ that cover the entire space $\mathcal{A}_k$ under $\|\cdot\|_s$. Let $\Pi_k$ denote a mapping from $\mathcal{A}$ to $\mathcal{A}_k$.

We are now ready to verify other conditions in AC. Assumption 3.2 in AC is not relevant for our application. Assumption 3.3 is an identification condition that is satisfied by Assumption 2.1. Since our model is correctly specified, Assumption 3.4(i)-(ii) in AC is satisfied by:

**Assumption 4.2.** *For all $h \in \mathcal{H}$, $h(v, z)$ has continuous derivative with respect to $v$, and $\|\alpha\|_s < \infty$ for all $\alpha \in \mathcal{A}$. For all $k \geq 1$, (i) $\mathcal{A}_k$ is compact under $\|\cdot\|_s$; (ii) $\mathcal{A}_k \subseteq \mathcal{A}_{k+1} \subseteq \mathcal{A}$, and $\|\Pi_k \alpha_0 - \alpha_0\|_s = o(1)$ as $k \to \infty$.*

Note that this condition does not require the parameter space A to be compact; it only requires the sieve space to be compact. The latter is much easier to satisfy because the approximating space is finite dimensional. Under our chosen metric, Assumption 4.2 requires the unknown function to have bounded derivative.

Denote

$$\rho_{ijt,s}(\alpha) = y_{ij} - \Phi(h_{t,s}(v_{ij} + x'_{ij}\beta, z_{it,s})).$$

For any $\alpha \in \mathcal{A}$, notice that

$$\rho_{ijt,s}(\alpha)^2 \leq y_{ij} + 1.$$

The random variable $y_{ij} + 1$ obviously has finite second and third moment.[6] Thus, the sample criterion function satisfies the envelope condition defined in AC (definition 3.1).

[6] In fact, since the dependent variable is bounded, $y_{it} + 1$ has finite moment of any order.

For any $\alpha \in \mathcal{A}$ and $\overline{\alpha} \in \mathcal{A}$, we have

$$
\begin{aligned}
& \left|\rho_{ijt,s}(\alpha)^2 - \rho_{ijt,s}(\overline{\alpha})^2\right| = \left|\rho_{ijt,s}(\alpha) + \rho_{ijt,s}(\overline{\alpha})\right| \times \left|\rho_{ijt,s}(\alpha) - \rho_{ijt,s}(\overline{\alpha})\right| \\
\leq\; & 2(y_{ij}+1)\left|\rho_{ijt,s}(\alpha) - \rho_{ijt,s}(\overline{\alpha})\right| \\
\leq\; & C \times 2(y_{ij}+1)\left|h_{t,s}(v_{ij}+x'_{ij}\beta, z_{it,s}) - \overline{h}_{t,s}(v_{ij}+x'_{ij}\overline{\beta}, z_{it,s})\right| \\
\leq\; & C \times 2(y_{ij}+1)\left(\begin{array}{c} \left|h_{t,s}(v_{ij}+x'_{ij}\beta, z_{it,s}) - \overline{h}_{t,s}(v_{ij}+x'_{ij}\beta, z_{it,s})\right| \\ + \left|\overline{h}_{t,s}(v_{ij}+x'_{ij}\beta, z_{it,s}) - \overline{h}_{t,s}(v_{ij}+x'_{ij}\overline{\beta}, z_{it,s})\right| \end{array}\right) \\
\leq\; & C \times 2(y_{ij}+1)\left(\sup_{v,z}\left|h_{t,s}(v,z) - \overline{h}_{t,s}(v,z)\right| + C \times \left\|\beta - \overline{\beta}\right\|_E\right) \\
\leq\; & 2(y_{ij}+1) \times C \times \|\alpha - \overline{\alpha}\|_s
\end{aligned}
$$

for some generic constant $C$, where the first inequality follows from application of the mean value theorem and the fact that the standard normal density function is bounded; and the second to the last inequality follows from $\frac{\partial \overline{h}(v,z)}{\partial v}$ and $x_{it}$ being bounded by Assumption 4.1 and 4.2. Thus, the sample criterion function is Hölder continuous, in the sense of AC (definition 3.2). This together with Lemma 2.1 and the following Assumption 4.3 imply Assumption 3.4 in AC. Notice that the conditional variance of $y_{ij}$ given $(v_{ij}, z_{it,s})$ is always bounded, Assumption 3.5(i)(ii) in AC is satisfied. Assumption 3.6(i) and 3.7(i) in AC are satisfied by the following Assumption 4.3.

**Assumption 4.3.** *(i)* $k \to \infty$ *and* $\frac{k}{n} \to 0$*; (ii)* $\ln[N(\varepsilon, \mathcal{A}_k, \|\cdot\|_s)] \div n \to 0$.

Assumption 4.3 requires the number of the approximating terms in the series approximation to grow with the sample size but not to grow too fast. This condition places an upper bound on the growth rate of $k$. The bound depends on the type of basis functions. For instance, for power series, Fourier series, and wavelet linear sieves, we have $\ln[N(\varepsilon, \mathcal{A}_k, \|\cdot\|_s)] = O(k)$. In this case, Assumption

4.3(ii) is implied by Assumption 4.3(i). For nonlinear sieves such as neural network and ridgelet, $\ln[N(\varepsilon, \mathcal{A}_k, \|\cdot\|_s)] = O(k\ln(k))$. In this case, Assumption 4.3(ii) is satisfied if $k\ln(k) \div n \to 0$. Applying Lemma 3.1 of Ai and Chen (2007), we obtain:

**Lemma 4.1.** *Under Assumption 2.1 and 4.1 - 4.3, we obtain* $\|\widehat{\alpha} - \alpha_o\|_s \to 0$ *in probability.*

The consistency result of Lemma 4.1 is useful in the sense that the estimator is shown to be in a neighborhood of the true value. It is, however, not enough for deriving the asymptotic distribution of $\widehat{\beta}$. To derive the asymptotic distribution of $\widehat{\beta}$, we need the convergence rate, particularly the convergence rate of $\widehat{h}_{t,s}$. To compute the convergence rate, it is often easier to use a metric that is locally equivalent to the criterion function. We now show that the following metric is locally equivalent to the criterion function:

$$\|\alpha - \alpha_o\|^2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s} E\left\{\begin{array}{c} \varphi(h_{t,so}(v_{ij} + x_{ij}'\beta_o, z_{it,s}))^2 \times \\ \left(\begin{array}{c} h_{t,s}(v_{ij} + x_{ij}'\beta_o, z_{it,s}) - h_{t,so}(v_{ij} + x_{ij}'\beta_o, z_{it,s}) \\ + \frac{\partial h_{t,so}(v_{ij} + x_{ij}'\beta_o, z_{it,s})}{\partial v} \times x_{ij}'(\beta - \beta_o) \end{array}\right)^2 \end{array}\right\},$$

where $\varphi$ is the standard normal density function.

It is easy to show that $\|\alpha - \alpha_o\|^2 \leq C \times \|\alpha - \alpha_o\|_s^2$ for some constant $C$ over all $\alpha \in \mathcal{A}$. Thus, the metric $\|\cdot\|$ is weaker than $\|\cdot\|_s$. Assumption 3.4(iv) in AC is satisfied by

**Assumption 4.2.** *(iii)* $\|\Pi_k\alpha_0 - \alpha_0\| = O(k^{-\mu})$ *and* $k^{-\mu}n^{1/4} \to 0$.

Assumption 3.2(iii), Assumption 3.5(iii)(iv), and Assumption 3.6(ii) in AC are not relevant for our application. Assumption 3.7(ii) in AC is satisfied by

**Assumption 4.3.** *(iii)* $\ln[N(\varepsilon, \mathcal{A}_k, \|\cdot\|_s)] \div \sqrt{n} \to 0$.

Let $\mathcal{W}$ denote the completion of $\mathcal{A} - \{\alpha_o\}$ under the metric $\|\cdot\|$. Because $\mathcal{B}$ is finite dimensional, we can write

$$\mathcal{W} = \mathcal{B} \times \mathcal{W}_h = \mathcal{B} \times \prod_{1 \leq t < T, s > t} \mathcal{W}_{t,sh}.$$

With $w_{dt,s}(\cdot) \in \mathcal{W}_{t,sh}$ for $d = 1, 2, ..., \dim(\beta)$, denote

$$w_{t,s}(\cdot) = (w_{1t,s}(\cdot), w_{2t,s}(\cdot), ..., w_{\dim(\beta)t,s}(\cdot))'.$$

Let $w^*(\cdot)$ denote the solution:

$$(5) \quad w^* = \arg\min_{w_{t,s}} \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \sum_{j=t,s} E \left\{ \begin{array}{c} \varphi(h_{t,so}(v_{ij} + x'_{ij}\beta_o, z_{it,s}))^2 \times \\ \left\| \frac{\partial h_{t,so}(v_{ij} + x'_{ij}\beta_o, z_{it,s})}{\partial v} \times x_{ij} - w_{t,s}(v_{ij} + x'_{ij}\beta_o, z_{it,s}) \right\|_E^2 \end{array} \right\}.$$

Denote

$$\begin{aligned} & D_{t,s}(v_{ij}, x_{ij}, z_{it,s}) \\ = \;& \varphi(h_{t,so}(v_{ij} + x'_{ij}\beta_o, z_{it,s})) \left( \frac{\partial h_{t,so}(v_{ij} + x'_{ij}\beta_o, z_{it,s})}{\partial v} \times x_{ij} - w^*_{t,s}(v_{ij} + x'_{ij}\beta_o, z_{it,s}) \right). \end{aligned}$$

We obtain

$$
\begin{aligned}
& \left\|\alpha-\alpha_o\right\|^2 \\
= \; & (\beta-\beta_o)' \times E\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s} D_{t,s}(v_{ij},x_{ij},z_{it,s}) \times D_{t,s}(v_{ij},x_{ij},z_{it,s})'\right\} \times (\beta-\beta_o) \\
& +E\left\{\begin{array}{c} \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s}\varphi(h_{t,so}(v_{ij}+x_{ij}'\beta_o,z_{it,s}))\times \\ \left(h_{t,s}(v_{ij}+x_{ij}'\beta_o,z_{it,s})-h_{t,so}(v_{ij}+x_{ij}'\beta_o,z_{it,s})+w_{t,s}^{*}(v_{ij}+x_{ij}'\beta_o,z_{it,s})'(\beta-\beta_o)\right)^2 \end{array}\right\} \\
\geq \; & \lambda_{\min} \times \left\|\beta-\beta_o\right\|_E^2
\end{aligned}
$$

where $\lambda_{\min}$ is the smallest eigenvalue of the matrix

$$
\Omega = E\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s} D_{t,s}(v_{ij},x_{ij},z_{it,s}) \times D_{t,s}(v_{ij},x_{ij},z_{it,s})'\right\}.
$$

Applying the inequality $(a+b)^2 \geq a^2-b^2$, we obtain

$$
\begin{aligned}
\left\|\alpha-\alpha_o\right\|^2 \geq \; & \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s} E\left\{\begin{array}{c} \varphi(h_{t,so}(v_{ij}+x_{ij}'\beta_o,z_{it,s}))^2\times \\ \left(h_{t,s}(v_{ij}+x_{ij}'\beta_o,z_{it,s})-h_{t,so}(v_{ij}+x_{ij}'\beta_o,z_{it,s})\right)^2 \end{array}\right\} \\
& -\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s} E\left\{\begin{array}{c} \varphi(h_{t,so}(v_{ij}+x_{ij}'\beta_o,z_{it,s}))^2\times \\ \left(\frac{\partial h_{t,so}(v_{ij}+x_{ij}'\beta_o,z_{it,s})}{\partial v}\times x_{ij}'(\beta-\beta_o)\right)^2 \end{array}\right\}.
\end{aligned}
$$

Hence

$$
\begin{aligned}
\left\|h-h_o\right\|_2^2 \equiv \; & \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s} E\left\{\begin{array}{c} \varphi(h_{t,so}(v_{ij}+x_{ij}'\beta_o,z_{it,s}))^2\times \\ \left(h_{t,s}(v_{ij}+x_{ij}'\beta_o,z_{it,s})-h_{t,so}(v_{ij}+x_{ij}'\beta_o,z_{it,s})\right)^2 \end{array}\right\} \\
\leq \; & \left\|\alpha-\alpha_o\right\|^2 + C \times \left\|\beta-\beta_o\right\|_E^2.
\end{aligned}
$$

Suppose that the following condition is satisfied.

**Assumption 4.4.** *(i) $w^*$ exists and is finite, and $\Omega$ is nonsingular.*

Under this assumption, the metric $\|\alpha - \alpha_o\|$ is equivalent to the $L_2$ norm:

$$\|\alpha - \alpha_o\|_2^2 = \|\beta - \beta_o\|_E^2 + \|h - h_o\|_2^2 .$$

After some manipulations, we can show that Assumption 3.8 in AC is satisfied by the following lemma. :

**Lemma 4.2.** *Under Assumption 2.1 and 4.2 - 4.4, in the neighborhood defined by $\|\alpha - \alpha_o\|_s = o(1)$, there exist some constants $C_1, C_2$ such that*

$$\begin{aligned} C_1 \|\alpha - \alpha_o\|^2 &\leq \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s} E\left\{\rho_{ijt,s}(\alpha)^2 - \rho_{ijt,s}(\alpha_o)^2\right\} \\ &\leq C_2 \|\alpha - \alpha_o\|^2 . \end{aligned}$$

Lemma 4.2 and the derivation above imply that the population criterion function is locally equivalent to the $L_2$ norm. This result will be useful in the derivation of the asymptotic distribution of the estimator $\widehat{\beta}$. Applying Theorem 3.1 in AC, we obtain the following result.

**Theorem 4.1.** *Under Assumption 2.1 and 4.1 - 4.4, we have $\|\widehat{\alpha} - \alpha_o\| = o_p(n^{-1/4})$.*

Lemma 4.2 and Theorem 4.1 imply $\|\widehat{\alpha} - \alpha_o\|_2 = o_p(n^{-1/4})$.

#### 4.0.1 Asymptotic normality

Having computed the convergence rate of the SNL estimator, we now derive the asymptotic distribution of $\widehat{\beta}$. For any $\lambda$, denote $f(\beta) = \lambda'(\beta - \beta_o)$. Denote $v_\beta^* = \Omega^{-1}\lambda$ and $v_{t,sh}^* = -w_{t,s}^{*\prime} v_\beta^*$. Let $\langle \cdot, \cdot \rangle$ denote the inner product induced by the metric $\|\cdot\|$:

$$\langle \alpha, \overline{\alpha} \rangle = \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \sum_{j=t,s} E \left\{ \begin{array}{c} \varphi(h_{t,so}(v_{ij} + x_{ij}'\beta_o, z_{it,s}))^2 \times \\ \left( h_{t,s}(v_{ij} + x_{ij}'\beta_o, z_{it,s}) + \frac{\partial h_{t,so}(v_{ij} + x_{ij}'\beta_o, zit,s)}{\partial v} \times x_{ij}'\beta \right) \times \\ \left( \overline{h}_{t,s}(v_{ij} + x_{ij}'\beta_o, z_{it,s}) + \frac{\partial h_{t,so}(v_{ij} + x_{ij}'\beta_o, z_{it,s})}{\partial v} \times x_{ij}'\overline{\beta} \right) \end{array} \right\}.$$

Assumption 4.1 in AC is satisfied by Assumption 4.4(i) and

**Assumption 4.4.** *(ii) $\beta_o$ is an interior point of $\mathcal{B}$.*

Notice that

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \sum_{j=t,s} E \left\{ \begin{array}{c} \varphi(h_{t,so}(v_{ij} + x_{ij}'\beta_o, z_{it,s}))^2 \times \\ \left( \frac{\partial h_{t,so}(v_{ij} + x_{ij}'\beta_o, z_{it,s})}{\partial v} \times x_{ij} - w_{t,s}^*(v_{ij} + x_{ij}'\beta_o, z_{it,s}) \right) \times \\ w_{t,s}(v_{ij} + x_{ij}'\beta_o, z_{it,s}) \end{array} \right\} \\ = \ & 0 \text{ for all } w_{t,s} \in \mathcal{W}_{t,sh} \end{aligned}$$

Denote $v_h^* = vec\{v_{t,sh}^*, 1 \leq t < T, t < s \leq T\}$ and $v^* = (v_\beta^*, v_h^*)$. By substituting for $v^*$ and applying the preceding equation, we obtain:

$$
\begin{aligned}
&\langle \alpha - \alpha_o, v^* \rangle \\
= \;& \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s} E\left\{\begin{array}{c} \varphi(h_o(v_t + x_t'\beta_o, z))^2 \times \\ \left(\begin{array}{c} h_{t,s}(v_{ij} + x_{ij}'\beta_o, z_{it,s}) - h_{t,so}(v_{ij} + x_{ij}'\beta_o, z_{it,s}) \\ + \frac{\partial h_{t,so}(v_{ij}+x_{ij}'\beta_o, z_{it,s})}{\partial v} \times x_{ij}'(\beta - \beta_o) \end{array}\right) \times \\ \left(\frac{\partial h_o(v_t + x_t'\beta_o, z)}{\partial v} \times x_t - w^*(v_t + x_t'\beta_o, z)\right)' v_\beta^* \end{array}\right\} \\
= \;& \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s} (\beta - \beta_o)' E\left\{\begin{array}{c} \varphi(h_o(v_t + x_t'\beta_o, z))^2 \times \\ \left(\frac{\partial h_o(v_t + x_t'\beta_o, z)}{\partial v} \times x_t - w^*(v_t + x_t'\beta_o, z)\right) \times \\ \left(\frac{\partial h_{t,so}(v_{ij}+x_{ij}'\beta_o, z_{it,s})}{\partial v} \times x_{ij} - w_{t,s}^*(v_{ij} + x_{ij}'\beta_o, z_{it,s})\right)' v_\beta^* \end{array}\right\} \\
= \;& \lambda'(\beta - \beta_o) = f(\beta).
\end{aligned}
$$

Denote $v_k^* = \Pi_k v^* = (v_\beta^*, v_{hk}^*)$ and $v_{hk}^* = -w_k^{*\prime} v_\beta^*$. Given the result of Theorem 4.1, Assumption 4.2 in AC is satisfied by

**Assumption 4.5.** $\|v_k^* - v^*\| = o(n^{-1/4})$ *as* $k \to \infty$.

Notice that

$$
\frac{d\rho_{ijt,s}(\alpha_0)}{d\alpha}[v_k^*] = -\varphi(h_{t,so}(v_{ij} + x_{ij}'\beta_o, z_{it,s}))\left[v_{hk}^*(v_{ij} + x_{ij}'\beta_o, z_{it,s}) + \frac{\partial h_{t,so}(v_{ij} + x_{ij}'\beta_o, z_{it,s})}{\partial \beta'} v_\beta^*\right]
$$

is bounded for all $v_{ij}$ and $z_{it,s}$. Moreover, it is easy to show that $\varphi(h_{t,s}(v_{ij} + x_{ij}'\beta, z_{it,s}))$ and $\frac{\partial h_{t,s}(v_{ij}+x_{ij}'\beta, z_{it,s})}{\partial \beta'}$ are Hölder continuous in $h$ and $\beta$. Hence, Assumption 4.3(i) in AC is satisfied. Assumption 4.3(ii) in AC is satisfied if the following condition is satisfied.

**Assumption 4.6.** *For any $h \in \mathcal{H}$, $h$ has up to second order derivatives with respect to the first* *argument and all derivatives are bounded.*

Assumption 4.3(iii) is not relevant for our application.

After some algebraic manipulation, we also can show that Assumption 4.4 in AC holds in the neighborhood $\|\alpha - \alpha_o\|_2 = o_p(n^{-1/4})$. Since $\|\cdot\|$ is equivalent to $\|\cdot\|_2$, Assumption 4.6 in AC is satisfied trivially. Assumption 4.5 in AC is satisfied by

**Assumption 4.7.** *(iii)* $\int \sqrt{\ln[N(\varepsilon, \mathcal{A}_k, \|\cdot\|_s)]} d\varepsilon < \infty.$

Denote

$$u_{ijt,s} = \rho_{ijt,s}(\alpha_o).$$

Applying the results of AC, we have

$$(\widehat{\beta} - \beta_o) = -\Omega^{-1} \frac{1}{n} \sum_{i=1}^{n} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \sum_{j=t,s} u_{ijt,s} D_{t,s}(v_{ij}, x_{ij}, z_{it,s}) + o_p(\frac{1}{\sqrt{n}}).$$

Denote

$$\Sigma = \frac{1}{n} \sum_{i=1}^{n} E \left\{ \begin{array}{c} \left[ \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \sum_{j=t,s} u_{ijt,s} D_{t,s}(v_{ij}, x_{ij}, z_{it,s}) \right] \times \\ \left[ \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \sum_{j=t,s} u_{ijt,s} D_{t,s}(v_{ij}, x_{ij}, z_{it,s}) \right]' \end{array} \right\}.$$

Applying Theorem 4.1 in AC, we obtain:

**Theorem 4.2.** *Under Assumption 2.1 and 4.1 - 4.7, we obtain* $\sqrt{n}(\widehat{\beta} - \beta_o) \to N(0, \Omega^{-1}\Sigma\Omega^{-1})$.

It is worth pointing out the obvious discrepancy between the conditions imposed here for Theorem 4.2 and the conditions imposed in Ai and Chen (2003) for their Theorem 4.1. On the surface, it appears that Ai and Chen (2003) impose more conditions than we do here. The fact of the matter is that most conditions imposed in Ai and Chen are trivially satisfied here because (i) the dependent variable is bounded here; (ii) the regression function here is continuously differentiable so we do not have to assume that the directional derivatives exist; and (iii) our estimator converges to the true value in probability under the stronger $L_2$ norm at rate faster than $n^{-1/4}$ so we do not have to control for the higher order bias terms.

### 4.0.2 Covariance

The asymptotic distribution derived in Theorem 4.2 is useful only if a consistent covariance matrix is provided. In this section, we present a consistent covariance matrix for $\Omega^{-1}\Sigma\Omega^{-1}$. First, we present a consistent estimator for $\Omega$. To this end, we need a consistent estimator of $w^*_{t,s}$. Notice that the criterion function in (5) is globally convex. we have $w^*_{t,s} \in \overline{\mathcal{W}}_{t,sh} \subset \mathcal{W}_{t,sh}$ with $\overline{\mathcal{W}}_{t,sh}$ compact under the sup metric. Suppose that the following condition is satisfied.

**Assumption 4.8.** *For any $j$ and some constant $c > 0$, $w^*_{t,s} \in \Lambda^2_c(\mathcal{V})$.*

Then, without loss of generality, we can restrict $\overline{\mathcal{W}}_{t,sh}$ to a Hölder ball and approximate $\overline{\mathcal{W}}_{t,sh}$ with

$$\begin{aligned}
\mathcal{W}_{t,shk} &= \mathcal{W}_{1t,shk} \times \cdots \times \mathcal{W}_{\dim(\beta)t,shk} \text{ and} \\
\mathcal{W}_{jt,shk} &= \{p^k(v,z)'\pi : \pi'\pi \leq C\} \subset \Lambda^2_c(\mathcal{V})
\end{aligned}$$

for some constant $C$. We estimate $w^*_{t,s}$ by

$$\widehat{w}_{t,s}(\cdot) = \arg\min_{w_{t,s}(\cdot)} \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s} \varphi(\widehat{h}_{t,s}(v_{ij}+x'_{ij}\widehat{\beta}, z_{it,s}))^2 \times \tag{6}$$

$$\left\| \frac{\partial \widehat{h}_{t,s}(v_{ij}+x'_{ij}\widehat{\beta}, z_{it,s})}{\partial v} \times x_{ij} - w_{t,s}(\widehat{h}_{t,s}(v_{ij}+x'_{ij}\widehat{\beta}, z_{it,s})) \right\|_E^2 . \tag{7}$$

Notice that $w_{t,s}(\cdot)$ is linear in series basis functions. The above problem is a system of seemingly unrelated regression with no cross equation restrictions on the coefficients. Since the SNL estimator $\widehat{\alpha}$ is consistent under the strong metric $\|\cdot\|_s$ and the sieve approximation $\mathcal{W}_{jt,shk}$ is in the Hölder ball, it is straightforward to show

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s} \varphi(\widehat{h}_{t,s}(v_{ij}+x'_{ij}\widehat{\beta}, z_{it,s}))^2 \times \\
&\left\| \frac{\partial \widehat{h}_{t,s}(v_{ij}+x'_{ij}\widehat{\beta}, z_{it,s})}{\partial v} \times x_{ij} - w_{t,s}(\widehat{h}_{t,s}(v_{ij}+x'_{ij}\widehat{\beta}, z_{it,s})) \right\|_E^2 \\
= &\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s} \varphi(h_{t,so}(v_{ij}+x'_{ij}\beta_o, z_{it,s}))^2 \times \\
&\left\| \frac{\partial h_{t,so}(v_{ij}+x'_{ij}\beta_o, z_{it,s})}{\partial v} \times x_{ij} - w_{t,s}(h_{t,so}(v_{ij}+x'_{ij}\beta_o, z_{it,s})) \right\|_E^2 + o_p(1).
\end{aligned}$$

Moreover, since

$$\varphi(h_{t,so}(v_{ij}+x'_{ij}\beta_o, z_{it,s}))^2 \left\| \frac{\partial h_{t,so}(v_{ij}+x'_{ij}\beta_o, z_{it,s})}{\partial v} \times x_{ij} - w_{t,s}(h_{t,so}(v_{ij}+x'_{ij}\beta_o, z_{it,s})) \right\|_E^2$$

is bounded (and hence satisfies the envelope condition) and satisfies the Lipschitz condition, applying

Lemma A.1 of Ai and Chen (2003), we obtain

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s}\varphi(h_{t,so}(v_{ij}+x_{ij}'\beta_o,z_{it,s}))^2\times \\
& \left\|\frac{\partial h_{t,so}(v_{ij}+x_{ij}'\beta_o,z_{it,s})}{\partial v}\times x_{ij}-w_{t,s}(h_{t,so}(v_{ij}+x_{ij}'\beta_o,z_{it,s}))\right\|_E^2 \\
= & \; E\left\{\begin{array}{c} \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s}\varphi(h_{t,so}(v_{ij}+x_{ij}'\beta_o,z_{it,s}))^2\times \\ \left\|\frac{\partial h_{t,so}(v_{ij}+x_{ij}'\beta_o,z_{it,s})}{\partial v}\times x_{ij}-w_{t,s}(h_{t,so}(v_{ij}+x_{ij}'\beta_o,z_{it,s}))\right\|_E^2 \end{array}\right\}+o_p(1)
\end{aligned}$$

uniformly over $w_{t,s}\in\overline{\mathcal{W}}_{t,sh}$. This and the compactness of $\overline{\mathcal{W}}_{t,sh}$ implies $\widehat{w}_{t,s}(v,z)=w_{t,s}^*(v,z)+o_p(1)$ uniformly over $(v,z)$. Denote

$$\begin{aligned}
\widehat{D}_{t,s}(v_{ij},x_{ij},z_{it,s}) &= \varphi(\widehat{h}_{t,s}(v_{ij}+x_{ij}'\widehat{\beta},z_{it,s}))\times \\
& \left(\frac{\partial\widehat{h}_{t,s}(v_{ij}+x_{ij}'\widehat{\beta},z_{it,s})}{\partial v}\times x_{ij}-\widehat{w}_{t,s}(\widehat{h}_{t,s}(v_{ij}+x_{ij}'\widehat{\beta},z_{it,s}))\right); \\
\widehat{\Omega} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s}\widehat{D}_{t,s}(v_{ij},x_{ij},z_{it,s})\widehat{D}_{t,s}(v_{ij},x_{ij},z_{it,s})'.
\end{aligned}$$

Then $\widehat{D}_{t,s}(v_{ij},x_{ij},z_{it,s})=D_{t,s}(v_{ij},x_{ij},z_{it,s})$ uniformly over $(v_{ij},x_{ij},z_{it,s})$ and hence $\widehat{\Omega}=\Omega+o_p(1)$.

Next, we present a consistent estimator for $\Sigma$. Denote

$$\widehat{\varepsilon}_{ijt,s}=y_{ij}-\Phi(\widehat{h}_{t,s}(v_{ij}+x_{ij}'\widehat{\beta},z_{it,s})).$$

It follows from Lemma 3.1 that $\widehat{\varepsilon}_{ijt,s}=\varepsilon_{ijt,s}+o_p(1)$ uniformly over $i$ and $j$. We estimate $\Sigma$ by

$$\widehat{\Sigma}=\frac{1}{n}\sum_{i=1}^{n}\left\{\begin{array}{c}\left[\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s}\widehat{\varepsilon}_{ijt,s}\widehat{D}_{t,s}(v_{ij},x_{ij},z_{it,s})\right]\times \\ \left[\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\sum_{j=t,s}\widehat{\varepsilon}_{ijt,s}\widehat{D}_{t,s}(v_{ij},x_{ij},z_{it,s})\right]'\end{array}\right\}.$$

The uniform convergence of $\widehat{\varepsilon}_{ijt,s}\widehat{D}_{t,s}(v_{ij}, x_{ij}, z_{it,s})$ implies $\widehat{\Sigma} = \Sigma + +o_p(1)$.

**Theorem 4.3.** *Under Assumption 2.1 and 4.1 - 3.8, we obtain* $\widehat{\Omega}^{-1}\widehat{\Sigma}\widehat{\Omega}^{-1} = \Omega^{-1}\Sigma\Omega^{-1} + o_p(1)$.

Theorem 4.3 provides standard errors for the $\beta$ estimator that can be used in statistical inference.

# 5 Conclusion

In this paper, we study estimation of the panel fixed effect binary choice model. Panel fixed-effect binary choice model is common in empirical studies; but its estimation is known to be quite difficult. Honoré and Lewbel (2003) presents a root-n consistent estimator at the cost that a "special regressor" exists. In this paper, we also present a root-n consistent estimator at the cost of a "special regressor". However, our "special regressor" is not required to satisfy a stringent support condition that their "special regressor" is required to satisfy. The key insight is that the conditional mean of the dependent variable has an index form which does not vary with time and hence satisfies a cross-equation restriction. This cross-equation restriction is the key to identify the parameter of interest. Under some sufficient conditions, we show that our estimator is root-n consistent and asymptotically normally distributed. We also provide an easy to compute and consistent covariance matrix.

Our estimator, however, has limitations. For instance, it rules out models with predetermined regressors such as lagged dependent variables. Extension of our approach to allow for predetermined regressors is conceivable and will be pursued in a future study.

# 6 Reference

Ai, C. (1997): “A Semiparametric Maximum Likelihood Estimator,” Econometrica, 65, 933-964.

Ai, C. and Chen, X. (2003). Efficient estimation of conditional moment restrictions models containing unknown functions, Econometrica, 71: 1795-1843.

Ai, C. and Chen, X. (2007). Estimation of Possibly Misspecified Semiparametric Conditional Moment Restriction Models With Different Conditioning Variables, Journal of Econometrics, in press.

Anderson, E. (1970). Asymptotic properties of conditional maximum likelihood estimators, Journal of the Royal Statistical Society, 32: 283-301.

Chamberlain, G. (1993). Feedback in panel data models, Manuscript.

Chen, X. and X. Shen (1998): “Sieve Extremum Estimates for Weakly Dependent Data,” Econometrica, 66, 289-314.

Chen, X., O. Linton, and I. van Keilegom (2002): “Estimation of Semiparametric Models when the Criterion Function is not Smooth,” Econometrica, 71, 1591-1608.

Ghosal, Subhashis, Arusharka Sen and Aad W. van der Vaart: ”Testing monotonicity of regression”, Annals of Statist. 28, no. 4 (2000), 1054 - 1082.

Hahn, J. (2001). The information bound of a dynamic panel Logit model with fixed effects. Econometric Theory 17, 913-932.

Hahn, J. and Newey, W. K. (2004). Jackknife and analytical bias reduction for nonlinear panel models, Econometrica, 72: 1295-1319.

Honoré, B. (1992). Trimmed LAD and least squares estimation of truncated and censored regression models with fixed effects, Econometrica, 60: 533-565.

Honoré, B. and Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables, Econometrica, 68: 839-874.

Honoré, B. and Lewbel, A. (2002). Semiparametric binary choice panel data models without strictly exogenous regressors, Econometrica, 70: 2053-2063.

Horowitz, J. (1992). A smoothed maximum score estimator for the binary response model, Econometrica, 60: 505-531.

Lee, M.J. (1999). A root-N consistent semiparametric estimator for fixed effects binary response panel data, Econometrica, 67: 427-433.

Lee, M.J. (2001), First-Difference estimator for panel censored-selection models, Economics Letters, 70: 43-49.

Manski, C. (1987). Semiparametric analysis of random effects linear models from binary panel data, Econometrica, 55: 357-362.

Rasch, G. (1960): Probabilistic Models for Some Intelligence and attainment tests, Denmarks Paedagogiske Institute, Copenhagen

Shen, X. (1997). On methods of sieves and penalization, Annals of Statistics, 25: 2555-2591.