Proposed Estimators for Dynamic and Static Probit Models with Panel Data

Gao, Wei and Yao, Qiwei and Bergsman, Wicher

Northeast Normal University, London School of Economics and Political Science, London School of Economics and Political Science

15 July 2013

Proposed Estimators for Dynamic and Static Probit Models with Panel Data

Wei Gao, Qiwei Yao and Wicher Bergsman
Key Laboratory for Applied Statistics of MOE, School of Mathematics and Statistics, Northeast Normal University, Changchun, Jilin 130024, China
Department of Statistics, London School of Economics, London, UK

1 Introduction

Suppose there have n independent individuals and two observations are made for each one. For i-th individual, there have observations d_{i1} and d_{i2} with binary responses, 0 or 1. Here suppose they are generated from the latent dynamic Model:

$$d_{i1} = I_{\{\tau_i + \boldsymbol{x}'_{i1}\beta + \epsilon_{i1} > 0\}}, \quad d_{i2} = I_{\{\tau_i + \gamma d_{i1} + \boldsymbol{x}'_{i2}\beta + \epsilon_{i2} > 0\}} \tag{1}$$

where I denotes the indicator function, ϵ_{i1} and ϵ_{i2} are independently and identically distributed with mean 0 and variance 1, τ_i the individual effect which demonstrates heterogeneities among individuals, $\boldsymbol{x}_{i1}$ and $\boldsymbol{x}_{i2}$ are covariates with dimension of k independent of τ_i and $(\epsilon_{i1}, \epsilon_{i2})'$, and β and γ are an interested parameters.

The Model (1) is adopted by Heckman(1978), Arellano and Honore(2001), Hisao(2005, p208). γ expresses the dynamic relationship between the previous state and the future state and is of considerable substantive interest. The state dependence for $\gamma \neq 0$ has been termed as the real(or true) state dependence by Heckman(1978,1980), which means that an individual who has experienced the event will behavior differently in the future compared with an otherwise identical individual who has not experienced, and the state dependence for $\gamma = 0$ has been termed as the spurious state dependence, in the sense that temporally persistent unobservables determine the previous and future of experience or choice which behaviors similarly. The model has quite applications in microeconomic data analysis.

When τ_i is thought as a fixed effect, and ϵ_{i1} and ϵ_{i2} are distributed by logistic distributions, Chamberlain(1980, 1985), Honore and Kyriazidou(2000), and Lancaster(2002) gives a consistent estimator of γ and shows its convergent rate. In more general cases, it is the incidental problem and a challenging one for microeconometrics and statistics. For probit models(ϵ_{i1} and ϵ_{i2} normally distributed), Heckman(1980) has shown that the maximum likelihood estimation of γ behaviors badly for the large variance of individual effect in his simulation studies given in Table 4.2.

Once treating τ_i as a random effect, one must give its prior distribution. Chamberlain(1980, 1985) also discusses the maximum likelihood estimation of β when γ is 0 and the prior distribution of τ is given. For long panel data, Arellano and Bonhomme(2009) have proved the estimated results show robust with priors. But for such short panel($T = 2$), different priors may lead to quite different estimation of γ and so it is necessary to choose a proper prior. In most cases, we do not know how to choose a suitable prior.

Manski(1987) proposes maximum score methods to estimate β when the distribution of error do not know and γ is equal to zero for Models (1). Later smoothed maximum score estimators are developed by Horowitz(1992). Arellano(2003) surveys the exiting approaches to deal with binary panel data for static models with individual effects. By introducing a quadratic exponential model, Bartolucci and Farcomeni (2009), Bartolucci and Nigro (2010) consider estimating problems for binary panel data.

In this paper, we consider estimating problems of Models (1) when ϵ_{i1} and ϵ_{i2} are normally distributed, which is the probit model conditional on covariates and individual effects. New estimating methods are proposed for γ, β, and simultaneous γ and β in Section 2. In Section 3, Simulation studies are carried out.

2 Proposed Estimating methods

2.1 A proposed estimator of γ when covariates are zeroes

When covariates are zeroes and τ_i has the density $f(x)$ and is independent of ϵ_{i1} and ϵ_{i2}, then

$$P\{d_{i1}=0, d_{i2}=0\} = \int \Phi(-x)\Phi(-x)f(x)dx, \quad P\{d_{i1}=0, d_{i2}=1\} = \int \Phi(-x)\Phi(x)f(x)dx$$

$$P\{d_{i1}=1, d_{i2}=0\} = \int \Phi(x)\Phi(-x-\gamma)f(x)dx, \quad P\{d_{i1}=1, d_{i2}=1\} = \int \Phi(x)\Phi(x+\gamma)f(x)dx$$

where $\Phi(x)$ is the distribution of standard normal variables. If $f(x)$ is known, the maximum likelihood estimation of γ is consistent and asymptotically normal distributed as the sample size n tends to infinity. By the comments given by Heckman(1978), we can deduce whether γ is equal to or greater or less than 0 from the ratio of $P\{d_{i1}=1, d_{i2}=0\}$ and $P\{d_{i1}=0, d_{i2}=1\}$, which can be estimated by

$$W = \frac{\sum\limits_{i=1}^{n} I_{\{d_{i1}=1,\, d_{i2}=0\}}}{\sum\limits_{i=1}^{n} I_{\{d_{i1}=0,\, d_{i2}=1\}}}. \tag{2}$$

Theorem 1. If

$$f(x) = \frac{1}{\sigma_\tau} g(\frac{x-\mu_\tau}{\sigma_\tau}) \tag{3}$$

where $g(x)$ is a density function with mean 0 and variance 1 and is continuous at 0, and $g(0)$ is finite, then

$$\lim_{\sigma_\tau \to \infty} \frac{\int \Phi(x)\Phi(-x-\gamma)f(x)dx}{\int \Phi(-x)\Phi(x)f(x)dx} = -\sqrt{\pi}\gamma\Phi(-\frac{\gamma}{\sqrt{2}}) + \exp\{-\frac{\gamma^2}{4}\}. \quad (4)$$

Proof. It is obvious by Lemma 1 and Lemma 2 given in Appendix.

Thus when σ_τ is sufficiently large, a proposed estimator of γ is given by

$$\hat{\gamma} = G^{-1}(W) \quad (5)$$

where

$$G(x) = -\sqrt{\pi}x\Phi(-x/\sqrt{2}) + \exp\{-x^2/4\}$$

and W is given by (2). Furthermore, for sufficient large σ_τ, the sample $(d_{i1} = 0, d_{i2} = 0)'$ or $(d_{i1} = 1, d_{i2} = 1)$ cannot supply more information about γ since

$$\begin{aligned}
\lim_{\sigma_\tau \to \infty} P\{d_{i1} = 0, d_{i2} = 0\} &= \lim_{\sigma_\tau \to \infty} \int \Phi(-x)\Phi(-x)f(x)dx \\
&= \lim_{\sigma_\tau \to \infty} \int \Phi(-x)\Phi(-x)\frac{1}{\sigma_\tau}g(\frac{x-\mu_\tau}{\sigma_\tau})dx \\
&= \lim_{\sigma_\tau \to \infty} \int \Phi(-\sigma_\tau t - \mu_\tau)\Phi(-\sigma_\tau t - \mu_\tau)g(t)dt \\
&= G_1(0)
\end{aligned}$$

and

$$\begin{aligned}
\lim_{\sigma_\tau \to \infty} P\{d_{i1} = 1, d_{i2} = 1\} &= \lim_{\sigma_\tau \to \infty} \int \Phi(x)\Phi(x+\gamma)f(x)dx \\
&= \lim_{\sigma_\tau \to \infty} \int \Phi(x)\Phi(x+\gamma)\frac{1}{\sigma_\tau}g(\frac{x-\mu_\tau}{\sigma_\tau})dx \\
&= \lim_{\sigma_\tau \to \infty} \int \Phi(\sigma_\tau t + \mu_\tau)\Phi(\sigma_\tau t + \mu_\tau + \gamma)g(t)dt \\
&= 1 - G_1(0)
\end{aligned}$$

when $G_1(x)$ is the distribution of $g(x)$. It may be one reason that the maximum likelihood estimator of γ shows badly in simulation studies given by Hecknan(1980) when σ_τ is larger. The variance of parameters will become larger when additional information has no direct connection with the interested parameters.

Theorem 2. For any $\epsilon > 0$,

$$\lim_{\sigma_\tau \to \infty} \lim_{n \to \infty} P\{|\hat{\gamma} - \gamma| \geq \epsilon\} = 0.$$

Proof: By the large number law, Theorem 1 and continuous properties of $G(x)$, it can be easily proved.

Theorem 3. If $\sigma_\tau = a\sqrt{n}(a > 0)$, then for all t,

$$\lim_{n \to \infty} P\left\{ \sqrt{\sum_{i=1}^{n} I_{\{d_{i1}=0,\, d_{i2}=1\}}}\, (\hat{\gamma} - \gamma) < t \right\} = \Phi(t/\sigma)$$

where

$$\sigma^2 = \frac{G(\gamma) + G^2(\gamma)}{[G'(\gamma)]^2} = \frac{G(\gamma) + G^2(\gamma)}{\pi \Phi^2(-\gamma/\sqrt{2})}.$$

Proof: By the Delta method, we can prove

$$\sqrt{n}\left(\hat{\gamma} - G^{-1}(p_{10})\right) = \sqrt{n}\left(G^{-1}(W) - G^{-1}(p_{10})\right) \Longrightarrow N(0, \sigma^{*2})$$

where

$$p_{01} = P\{d_{i1} = 0, d_{i2} = 1\}, \quad p_{10} = P\{d_{i1} = 1, d_{i2} = 0\}$$

and

$$\sigma^{*2} = \frac{1}{[G'(\gamma)]^2} \left[\frac{p_{10}}{p_{01}^2} + \frac{p_{10}^2}{p_{01}^3} \right].$$

Then

$$\sqrt{\sum_{i=1}^{n} I_{\{d_{i1}=0,\, d_{i2}=1\}}}\, \left(\hat{\gamma} - G^{-1}(p_{10})\right) \Longrightarrow N(0, \sigma^2)$$

by $\sqrt{\sum_{i=1}^{n} I_{\{d_{i1}=0,\, d_{i2}=1\}}/n} \longrightarrow p_{01}$ in probability and (4).

$$\begin{aligned} G^{-1}(p_{10}) - G^{-1}(G(\gamma)) &= \frac{p_{10} - G(\gamma)}{G'(G(\gamma))} + o(p_{10} - G(\gamma)) \\ &= c \times \sigma_\tau^{-1} + o(\sigma_\tau^{-1}) \end{aligned}$$

by Lemma 3 given in Appendix and

$$c = \frac{\sqrt{\pi}\left\{\frac{\gamma^2}{2}\Phi(-\frac{\gamma}{\sqrt{2}}) + \Phi(-\frac{\gamma}{\sqrt{2}}) - \int_{-\infty}^{-\gamma/\sqrt{2}} t^2\phi(t)dt\right\}}{G'(G(\gamma))}.$$

So

$$\begin{aligned} \sqrt{\sum_{i=1}^{n} I_{\{d_{i1}=0,\, d_{i2}=1\}}}\,(\hat{\gamma} - \gamma) &= \sqrt{\sum_{i=1}^{n} I_{\{d_{i1}=0,\, d_{i2}=1\}}}\left[\hat{\gamma} - G^{-1}(p_{10})\right] \\ &\quad + \sqrt{\sum_{i=1}^{n} I_{\{d_{i1}=0,\, d_{i2}=1\}}}\left[G^{-1}(p_{10}) - G^{-1}(G(\gamma))\right] \\ &= \sqrt{\sum_{i=1}^{n} I_{\{d_{i1}=0,\, d_{i2}=1\}}}\left[\hat{\gamma} - G^{-1}(p_{10})\right] + o_p(1), \end{aligned}$$

which implies that the Theorem holds.

Remark:

$$\frac{\sum_{i=1}^{n} I_{\{d_{i1}=0,\, d_{i2}=1\}}}{n} \longrightarrow f(\mu_\tau)\int \Phi(x)\Phi(-x)dx + o(\sigma_\tau^{-1})$$

and

$$\frac{\sum_{i=1}^{n} I_{\{d_{i1}=1,\, d_{i2}=0\}}}{n} \longrightarrow f(\mu_\tau)\int \Phi(x)\Phi(-x-\gamma)dx + o(\sigma_\tau^{-1}),$$

in probability, which mean that there needs the larger sample size for the bigger σ_τ.

2.2 Estimation of β when γ is zero

Let

$$D_n = \{(d_{i1}, d_{i2})' : \ d_{i1} + d_{i2} = 1 \ \text{ for } \ i = 1, \cdots, n\}$$

and $m =^{\#} D_n$, the number of elements in D_n. Without loss of generality, suppose that $d_{i1} + d_{i2} = 1$ for $i = 1, \cdots, m$.

The conditional probability

$$P\{d_{i1} = 1, d_{i2} = 0 | d_{i1} + d_{i2} = 1, \ \boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}\}$$
$$= \frac{\int \Phi(\boldsymbol{x}_{i1}'\beta + t)\Phi(-\boldsymbol{x}_{i2}'\beta - t)f(t)dt}{\int \Phi(\boldsymbol{x}_{i1}'\beta + t)\Phi(-\boldsymbol{x}_{i2}'\beta - t)f(t)dt + \int \Phi(-\boldsymbol{x}_{i1}'\beta - t)\Phi(\boldsymbol{x}_{i2}'\beta + t)f(t)dt}.$$

Under (3), we can similarly prove

$$\lim_{\sigma_\tau \to \infty} P\{d_{i1} = 1, d_{i2} = 0 | d_{i1} + d_{i2} = 1, \ \boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}\} = \frac{G((\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\beta)}{G((\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\beta) + G(-(\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\beta)}.$$

For sufficient large σ_τ, we can replace the conditional likelihood of β given D_n by

$$L(\beta) = \prod_{i=1}^{m} p_i^{z_i}(1 - p_i)^{1 - z_i} \tag{6}$$

where $z_i = I_{\{d_{i1}=1, d_{i2}=0\}}$ and $1 - z_i = I_{\{d_{i1}=0, d_{i2}=1\}}$, and

$$p_i = \frac{G((\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\beta)}{G((\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\beta) + G(-(\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\beta)}. \tag{7}$$

If we define a function

$$K(t) = \frac{G(t)}{G(t) + G(-t)},$$

we can show that $K(t)$ is monotonic in t and then (7) can be expressed into generalized linear models with the link function $K^{-1}(t)$, that is,

$$K^{-1}(p_i) = (\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\beta.$$

So related results for generalized Models given by McCullagh and Nelder(1989) can be applied to (6). Under some regular conditions and $\sigma_\tau \longrightarrow \infty$, the consistency of β can be obtained.

2.3 Simultaneous estimation γ and β for Models (1)

As in Section 4, we have

$$\lim_{\sigma_\tau \to \infty} P\{d_{i1}=1, d_{i2}=0 | d_{i1}+d_{i2}=1,\ \boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}\} = \frac{G(\gamma + (\boldsymbol{x}_{i2}-\boldsymbol{x}_{i1})'\beta)}{G(\gamma + (\boldsymbol{x}_{i2}-\boldsymbol{x}_{i1})'\beta) + G(-(\boldsymbol{x}_{i2}-\boldsymbol{x}_{i1})'\beta)}.$$

For the large σ_τ, we replace the condition likelihood given D_n of γ and β by

$$L(\beta) = \prod_{i=1}^{m} p_i^{z_i}(1-p_i)^{1-z_i} \tag{8}$$

where $z_i = I_{\{d_{i1}=1, d_{i2}=0\}}$ and $1 - z_i = I_{\{d_{i1}=0, d_{i2}=1\}}$, and

$$p_i = \frac{G(\gamma + (\boldsymbol{x}_{i2}-\boldsymbol{x}_{i1})'\beta)}{G(\gamma + (\boldsymbol{x}_{i2}-\boldsymbol{x}_{i1})'\beta) + G(-(\boldsymbol{x}_{i2}-\boldsymbol{x}_{i1})'\beta)}. \tag{9}$$

Let

$$X^* = (\boldsymbol{x}_{12}-\boldsymbol{x}_{11}, \boldsymbol{x}_{22}-\boldsymbol{x}_{21}, \cdots, \boldsymbol{x}_{m2}-\boldsymbol{x}_{m1})$$

Theorem 4. (9) is identifiable for γ and β if the rank of X^* is equal to k(dimension of $\boldsymbol{x}_{2i}-\boldsymbol{x}_{1i}$) and at least there exits j and $1 \le s_1, \cdots, s_k \le m$ which satisfy

$$\boldsymbol{x}_{j2}-\boldsymbol{x}_{j1} = a_1(\boldsymbol{x}_{s_1 2}-\boldsymbol{x}_{s_1 1}) + a_2(\boldsymbol{x}_{s_2 2}-\boldsymbol{x}_{s_2 1}) + \cdots + a_k(\boldsymbol{x}_{s_k 2}-\boldsymbol{x}_{s_k 1})$$

where $a_1, \cdots, a_k$ is non-positive real number.

Proof: By Lemma 4 given in Appendix, it can be proved with $r_i = p_i/(1-p_i)$ and $\boldsymbol{x}_i = \boldsymbol{x}_{i2} - \boldsymbol{x}_{i1}$.

The conditions in Theorem 4 is sufficient and it can be satisfied with probability near 1 for the large sample size n if the covariate $\boldsymbol{x}_{i2}-\boldsymbol{x}_{i1}$ is a continuous variable and its variance is positive definite.

Corollary Under the condition in Theorem 4, let $\mathbf{1}_m$ be the $m-$dimensional vector with its components 1 and then the rank of $(\mathbf{1}_m, X^{*'})$ is $k+1$.

Proof: Without loss of generality, suppose that $\boldsymbol{x}_{12}-\boldsymbol{x}_{11}, \cdots, \boldsymbol{x}_{k2}-\boldsymbol{x}_{k1}$ are linear independent and

$$\boldsymbol{x}_{k+1\,2}-\boldsymbol{x}_{k+1\,1} = a_1(\boldsymbol{x}_{12}-\boldsymbol{x}_{11}) + \cdots + a_k(\boldsymbol{x}_{k2}-\boldsymbol{x}_{k1})$$

where $a_1, \cdots, a_k$ is non-positive real number. Then the determinant

$$\begin{pmatrix} \boldsymbol{x}'_{12}-\boldsymbol{x}'_{11} & 1 \\ \boldsymbol{x}'_{22}-\boldsymbol{x}'_{21} & 1 \\ \vdots & \vdots \\ \boldsymbol{x}'_{k2}-\boldsymbol{x}'_{k1} & 1 \\ \boldsymbol{x}'_{k+1\,2}-\boldsymbol{x}'_{k+1\,1} & 1 \end{pmatrix} \tag{10}$$

is equal to

$$\begin{vmatrix} \boldsymbol{x}'_{12}-\boldsymbol{x}'_{11} \\ \boldsymbol{x}'_{22}-\boldsymbol{x}'_{21} \\ \vdots \\ \boldsymbol{x}'_{k2}-\boldsymbol{x}'_{k1} \end{vmatrix} \left[1-(\boldsymbol{x}_{k+1\,2}-\boldsymbol{x}_{k+1\,1})' \begin{pmatrix} \boldsymbol{x}'_{12}-\boldsymbol{x}'_{11} \\ \boldsymbol{x}'_{22}-\boldsymbol{x}'_{21} \\ \vdots \\ \boldsymbol{x}'_{k2}-\boldsymbol{x}'_{k1} \end{pmatrix}^{-1} \mathbf{1}_k \right]$$
$$= \left|\boldsymbol{x}_{12}-\boldsymbol{x}_{11}, \boldsymbol{x}_{22}-\boldsymbol{x}_{21}, \cdots, \boldsymbol{x}_{k2}-\boldsymbol{x}_{k1}\right| \left[1-\sum_{i=1}^{k} a_i\right] \neq 0$$

by the assumption. This implies that the rank of (10) is $k+1$.

Since the rank of $(\mathbf{1}_m, X^{*'})$ is equal to that of $(X^{*'}, \mathbf{1}_m)$, which is a $m \times (k+1)$ matrix, and (10) is a matrix obtained by the first $k+1$ rows of $(X^{*'}, \mathbf{1}_m)$, thus the rank of $(\mathbf{1}_m, X^{*'})$ is $k+1$.

From Corollary, it seems that identifiable conditions of (9) are stronger than that of linear models since that the rank of design matrices is equal to the dimension of parameters is sufficient for linear models to be identified.

3 Simulation studies

For the given sample size n, 100 simulations are repeated and estimating results are listed in the following table.

		$n = 1000$		$n = 5000$
	γ			
	-2	-1.955(0.158)		-2.066(0.202)
	-1.5	-1.426 (0.233)		-1.506 (0.188)
	-1	-1.027 (0.230)		-1.002 (0.142)
	-0.5	-0.514 (0.198)		-0.514 (0.151)
U(-3,3)	0	-0.010 (0.154)	U(-10,10)	-0.009 (0.128)
	0.5	0.514 (0.205)		0.517 (0.312)
	1	1.067 (0.169)		0.984 (0.141)
	1.5	1.495 (0.153)		1.495 (0.153)
	2	1.997 (0.246)		2.037 (0.175)
	-2	-1.788(0.206)		-1.955(0.158)
	-1.5	-1.483(0.146)		-1.506 (0.188)
	-1	-0.970 (0.199)		-0.994 (0.127)
	-0.5	-0.496 (0.148)		-0.507 (0.117)
N(0,4)	0	-0.030 (0.146)	N(0,25)	-0.010 (0.111)
	0.5	0.509 (0.161)		0.472 (0.098)
	1	1.029 (0.172)		1.008 (0.110)
	1.5	1.507 (0.176)		1.496 (0.120)
	2	2.073 (0.219)		2.032 (0.164)

Appendix

Lemma 1. If $f(x)$ satisfies the conditions given in Theorem 1, then

$$\int \Phi(x)\Phi(-x-\gamma)f(x)dx = f(\mu_\tau)\int \Phi(x)\Phi(-x-\gamma)dx + o(\sigma_\tau^{-1})$$

and

$$\int \Phi(-x)\Phi(x)f(x)dx = f(\mu_\tau)\int \Phi(-x)\Phi(x)dx + o(\sigma_\tau^{-1}).$$

Proof.

$$\left|\sigma_\tau[\int \Phi(x)\Phi(-x-\gamma)f(x)dx - f(\mu_\tau)\int \Phi(x)\Phi(-x-\gamma)dx]\right|$$

$$= \left|\int \Phi(x)\Phi(-x-\gamma)g(\frac{x-\mu_\tau}{\sigma_\tau})dx - g(0)\int \Phi(x)\Phi(-x-\gamma)dx\right|$$

$$\leq \int_{x>M} \Phi(x)\Phi(-x-\gamma)g(\frac{x-\mu_\tau}{\sigma_\tau})dx + \int_{x<-M} \Phi(x)\Phi(-x-\gamma)g(\frac{x-\mu_\tau}{\sigma_\tau})dx$$

$$+g(0)\int_{x>M} \Phi(x)\Phi(-x-\gamma)dx + g(0)\int_{x<-M} \Phi(x)\Phi(-x-\gamma)dx$$

$$+\int_{|x|\leq M} \Phi(x)\Phi(-x-\gamma)\left|g(\frac{x-\mu_\tau}{\sigma_\tau}) - g(0)\right|dx$$

$$\leq \Phi(-M-\gamma) + \Phi(-M) + g(0)\int_{x>M} \Phi(x)\Phi(-x-\gamma)dx$$

$$+g(0)\int_{x<-M} \Phi(x)\Phi(-x-\gamma)dx + \int_{|x|\leq M} \Phi(x)\Phi(-x-\gamma)\left|g(\frac{x-\mu_\tau}{\sigma_\tau}) - g(0)\right|dx.$$

For given γ , $\Phi(-M-\gamma)$ and $\Phi(-M)$ can be arbitrary small for sufficient large M. Furthermore $\int \Phi(x)\Phi(-x-\gamma)$ is integrable, and so $\int_{x<-M} \Phi(x)\Phi(-x-\gamma)dx$ and $\int_{x>M} \Phi(x)\Phi(-x-\gamma)dx$ can also be arbitrary small for sufficient large M. For given M, $\int_{|x|\leq M} \Phi(x)\Phi(-x-\gamma)\left|g(\frac{x-\mu_\tau}{\sigma_\tau}) - g(0)\right|dx$ can also be arbitrary small for sufficient large σ_τ. So

$$\int \Phi(x)\Phi(-x-\gamma)f(x)dx = f(\mu_\tau)\int \Phi(x)\Phi(-x-\gamma)dx + o(\sigma_\tau^{-1}).$$

Similarly, the other part can be proved.

Lemma 2.

$$\int \Phi(-x)\Phi(x+\beta)dx = \beta\Phi(\frac{\beta}{\sqrt{2}}) + \frac{1}{\sqrt{\pi}}\exp\{-\frac{\beta^2}{4}\}.$$

Proof. By the fact $d(x\Phi(x) + \phi(x)) = \Phi(x)$ and integration by parts,

$$\begin{aligned}\int \Phi(-x)\Phi(x+\beta)dx &= \int \phi(x)[(x+\beta)\Phi(x+\beta) + \phi(x+\beta)]dx \\ &= \beta\int \phi(x)\Phi(x+\beta)dx + \int x\phi(x)\Phi(x+\beta)dx + \int \phi(x)\phi(x+\beta)dx \\ &= \beta\Phi(\frac{\beta}{\sqrt{2}}) + 2\int \phi(x)\phi(x+\beta)dx \\ &= \beta\Phi(\frac{\beta}{\sqrt{2}}) + \frac{1}{\sqrt{\pi}}\exp\{-\frac{\beta^2}{4}\}.\end{aligned}$$

Lemma 3. If $f(x)$ satisfies the conditions given in Theorem 1 and is derivative at μ_τ, then

$$\frac{\int \Phi(x)\Phi(-x-\gamma)f(x)dx}{\int \Phi(-x)\Phi(x)f(x)dx} - G(\gamma) = \left\{\frac{\gamma^2}{2}\Phi(-\frac{\gamma}{\sqrt{2}}) + \Phi(-\frac{\gamma}{\sqrt{2}}) - \int_{-\infty}^{-\gamma/\sqrt{2}} t^2\phi(t)dt\right\}\frac{\sqrt{\pi}}{\sigma_\tau} + o(\sigma_\tau^{-1}).$$

Proof: Expand the function

$$\frac{\int \Phi(x)\Phi(-x-\gamma)f(x)dx}{\int \Phi(-x)\Phi(x)f(x)dx}$$

at $\sigma_\tau = \infty$ and then the Lemma can be obtained.

Lemma 4. Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k, \boldsymbol{x}_{k+1} \in R^k$ satisfy: (a) $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k$ are linearly independent; (b)$\boldsymbol{x}_{k+1} = -c_1\boldsymbol{x}_1 - c_2\boldsymbol{x}_2 - \cdots - c_k\boldsymbol{x}_k$ where $c_1, \cdots, c_k$ are non-negative real number, and $r_1, \cdots, r_k, r_{k+1}$ be positive real number, then the equation

$$\begin{cases} G(\boldsymbol{x}_1^{'}\beta+\alpha)-r_1G(-\boldsymbol{x}_1^{'}\beta)=0 \\ G(\boldsymbol{x}_2^{'}\beta+\alpha)-r_2G(-\boldsymbol{x}_2^{'}\beta)=0 \\ \cdots\cdots \\ G(\boldsymbol{x}_k^{'}\beta+\alpha)-r_kG(-\boldsymbol{x}_k^{'}\beta)=0 \\ G(\boldsymbol{x}_{p+1}^{'}\beta+\alpha)-r_{k+1}G(-\boldsymbol{x}_{k+1}^{'}\beta)=0 \end{cases} \tag{11}$$

has unique solution β and α.

Proof: For fixed α, let

$$u_\alpha(z)=\frac{G(z+\alpha)}{G(-z)}$$

and

$$\begin{aligned} \frac{du_\alpha(z)}{dz} &= \frac{G^{'}(z+\alpha)G(-z)+G(z+\alpha)G^{'}(-z)}{G^2(-z)} \\ &= -\sqrt{\pi}\frac{\Phi(-(z+\alpha)/\sqrt{2})G(-z)+G(z+\alpha)\Phi(z/\sqrt{2})}{G^2(-z)} \\ &< 0. \end{aligned}$$

So $u_\alpha(z)$ is deceasing in z and $\lim_{z\to-\infty} u_\alpha(z)=\infty$ and $\lim_{z\to\infty} u_\alpha(z)=0$. Thus for fixed α, the equation

$$\begin{cases} G(\boldsymbol{x}_1^{'}\beta+\alpha)-r_1G(-\boldsymbol{x}_1^{'}\beta)=0 \\ G(\boldsymbol{x}_2^{'}\beta+\alpha)-r_2G(-\boldsymbol{x}_2^{'}\beta)=0 \\ \cdots\cdots \\ G(\boldsymbol{x}_k^{'}\beta+\alpha)-r_kG(-\boldsymbol{x}_k^{'}\beta)=0 \end{cases} \tag{12}$$

has a unique solution when $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k$ are linearly independent.

Let $\beta^*=(\beta_1(\alpha), \cdots, \beta_k(\alpha))^{'}$ the solution of (12), and then

$$\frac{d\beta^*}{d\alpha}=-X^{'-1}\delta$$

where

$$\delta = (\delta_1, \cdots, \delta_k)', \quad \delta_i = \frac{\Phi(-(\boldsymbol{x}_i'\beta^* + \alpha)/\sqrt{2})}{\Phi(-(\boldsymbol{x}_i'\beta^* + \alpha)/\sqrt{2}) + r_i\Phi(\boldsymbol{x}_i'\beta^*/\sqrt{2})}$$

and

$$X = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k).$$

Define

$$t(\alpha) = G(\boldsymbol{x}_{k+1}'\beta^* + \alpha) - r_{k+1}G(-\boldsymbol{x}_{k+1}'\beta^*),$$

and then

$$\begin{aligned}
\frac{dt(\alpha)}{d\alpha} &= -\sqrt{\pi}\left\{\left[\Phi(-\frac{\boldsymbol{x}_{k+1}'\beta^* + \alpha}{\sqrt{2}}) + r_{k+1}\Phi(\frac{\boldsymbol{x}_{k+1}'\beta^*}{\sqrt{2}})\right]\boldsymbol{x}_{k+1}'\frac{d\beta^*}{d\alpha} + \Phi(-\frac{\boldsymbol{x}_{k+1}'\beta^* + \alpha}{\sqrt{2}})\right\} \\
&= -\sqrt{\pi}\left\{\left[\Phi(-\frac{\boldsymbol{x}_{k+1}'\beta^* + \alpha}{\sqrt{2}}) + r_{k+1}\Phi(\frac{\boldsymbol{x}_{k+1}'\beta^*}{\sqrt{2}})\right]\left(\sum_{j=1}^{k} c_j\delta_i\right) + \Phi(-\frac{\boldsymbol{x}_{k+1}'\beta^* + \alpha}{\sqrt{2}})\right\} \\
&< 0,
\end{aligned}$$

which implies $t(\alpha) = 0$ have an unique solution and the Lemma is concluded.

References

Arellano, M. (2003). Discrete choices with panel data. *Investigaciones Economicas*, 27, 423-458.

Arellano, M. and Bonhomme, S. (2009). Robust priors in nonlinear panel data models. *Econometrica*, 77, 489-536.

Arellano, M. and Honore, B. (2001). Panel data Models: some recent developments. *Handbook of Econometrics*, Vol. V, ed. by J. Heckman and E. Leamer. Amsterdam: North Holland.

Bartolucci, F. and Farcomeni, A.(2009). A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. *Journal of the American Statistical Association*, 104, 816-833.

Bartolucci, F. and Nigro, V.(2010). A dynamic model for binary panel data with unobserved heterogeneity admitting a $\sqrt{n}$ consistent conditional estimator. *Econometrica*, 78, 719-733.

Chamberlain, G.(1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, 47, 225-238.

Chamberlain, G.(1985). Heterogeneity, omitted variables bias, and duration dependence. *Longitudinal Analysis of Labor Market Data*, edited by Heckman, J. and Singer, B. Cambridge University Press.

Heckman, J.(1980). The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. *Structural Analysis of Discrete Data with Econometric Applications*, ed by C. F. Manski and D. McFadden, p179-195. Cambridge, MA: MIT Press.

Heckman, J.(1978). Simple statistical models for discrete panel data developed and applied to test the hypothesis of true state dependence against the hypothesis of spurious state dependence. *Annales de l'lNSEE* 30/31, 227-269.

Hisao, C.(2005). *Analysis of Panel Data*(Second Ed.). New York: Cambridge University Press.

Honore, B. and Kyriazidou, E.(2000). Panel data discrete choice models with lagged dependent variables. *Econometrica*, 68, 611-629.

Horowitz, J. L.(1992). A smoothed maximum score estimator for binary response model. *Econmetrica*, 60, 505-531.

Lancaster, T.(2002). Orthogonal parameters and panel data. *Review of Economic Studies*, 647-666.

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

McCullgh, P. and Nelder, J. A.(1989). *Generalized Linear Models*. London: Chapman & Hall.