Wei Gao, Wicher Bergsma and Qiwei Yao

Estimation for dynamic and static panel probit models with large individual effects

Article (Accepted version) (Refereed)

Original citation: Gao, Wei, Bergsma, Wicher and Yao, Qiwei (2016) *Estimation for dynamic and static panel probit models with large individual effects.* Journal of Time Series Analysis . ISSN 0143-9782

DOI: 10.1111/jtsa.12178

Estimation for Dynamic and Static Panel Probit Models with Large Individual Effects

Wei Gao[⋆] Wicher Bergsma[∗] Qiwei Yao[†]

[⋆]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

[†]Guanghua School of Management, Peking University, Beijing, China

Abstract

For discrete panel data, the dynamic relationship between successive observations is often of interest. We consider a dynamic probit model for short panel data. A problem with estimating the dynamic parameter of interest is that the model contains a large number of nuisance parameters, one for each individual. Heckman proposed to use maximum likelihood estimation of the dynamic parameter, which, however, does not perform well if the individual effects are large. We suggest new estimators for the dynamic parameter, based on the assumption that the individual parameters are random and possibly large. Theoretical properties of our estimators are derived and a simulation study shows they have some advantages compared to Heckman's estimator and the modified profile likelihood estimator(MPL) for fixed effects.

Key Words: Dynamic probit regression; Generalized linear models; Panel data; Probit models; Static probit regression.

1 Introduction

Short binary-valued time series in the presence of covariates are often available in panel studies for which observations are taken on a panel of individuals over a short time period. Dynamic probit regression is one of the most frequently used statistical models to analyse this type of data. To set the scene, consider a panel of n independently sampled individuals. For each individual i, binary observations, denoted by $d_{i1}, \cdots, d_{iT}$, are taken at time $1, \cdots, T$, and the observations are assumed to satisfy the latent dynamic model:

$$d_{i1} = I(\tau_i + \boldsymbol{x}_{i1}^{'}\boldsymbol{\beta} + \epsilon_{i1} > 0), \cdots, d_{it} = I(\tau_i + \gamma d_{i\,t-1} + \boldsymbol{x}_{it}^{'}\boldsymbol{\beta} + \epsilon_{it} > 0) \text{ for } 1 < t \leq T, \quad (1)$$

subject to

$$\epsilon_{it} \sim_{\text{iid}} N(0, 1) \quad (2)$$

where $I(\cdot)$ denotes the indicator function, $\{\boldsymbol{x}_{it}\}$ are $k \times 1$ covariate vectors, τ_i is an unknown intercept representing the i-th individual effect, and the autoregressive coefficient γ and the regressive coefficient $\boldsymbol{\beta}$ are unknown parameters which are assumed to be the same for all individuals. In (1), only the d_{it} and $\boldsymbol{x}_{it}$ are observable. The goal is often to estimate γ and $\boldsymbol{\beta}$ while the τ_i are treated as nuisance parameters. As with most panel data, the number of individuals n is large while the length of observed time period T is small. Therefore the asymptotic approximations are often derived with $n \to \infty$ and T fixed.

Model (1) is a dynamic panel probit regression model, as the dynamic dependence is reflected by the autoregressive parameter γ which links d_{it}, i.e. the state at time t, to the state at time $t-1$. When $\gamma = 0$, (1) reduces to a static panel probit regression, as now d_{it} is independent of $d_{i,t-1}, d_{i,t-2}, \cdots$. Model (1) has been used for various applications in microeconomics by, among others, Heckman (1978), Arellano and Honore (2001), and Hsiao (2003, Section 7.5). For example, Heckman (1978, 1980) used model (1) to reveal some interesting dynamics in unemployment data: $d_{it} = 0$ indicates that individual i is unemployed at time t, and 1 otherwise, while the covariate $\boldsymbol{x}_{it}$ stands for the factors (such as age, education, family background etc) which may affect the employment status. These studies tried to provide statistical evidence to answer questions such as: *Does current unemployment cause future unemployment?* If $\gamma > 0$ this indicates that being in employment at time t increases the chances of being in employment at time $t + 1$.

Various estimation methods have been proposed for model (1). By treating the individual effects $\tau_1, \cdots, \tau_n$ as nuisance parameters or incidental parameters (Neyman and Scott, 1948), Heckman (1980) adopted the maximum likelihood estimator of γ as well as $\boldsymbol{\beta}$ when ϵ_{it} are normally distributed. Chamberlain (1980, 1985), Honore and Kyriazidou (2000), and Lancaster (2002) considered the models with logistic distributed ϵ_{it}. They proposed a consistent estimator of γ and derived its convergence rate. Bartolucci and Farcomeni (2009) and Bartolucci and Nigro (2010) considered some extended versions of dynamic logit models with heterogeneity beyond those reflected by the covariates in the models. A standard method to deal with incidental parameter problems is to use a conditional likelihood to eliminate the incidental parameters by conditioning on sufficient statistics for those parameters; see, e.g. Chamberlain (1980), Bartolucci and Nigro (2010), and also Lancaster (2000).

An attractive alternative is to treat individual effects τ_i as random effects with prespecified priors. But as far as we are aware, the literature on panel probit regression taking this approach only deals with the static model (i.e. $\gamma = 0$ in (1)). For example, Chamberlain (1980, 1985) discussed the maximum likelihood estimator for $\boldsymbol{\beta}$ with a given prior distribution for τ_i. Arellano and Bonhomme (2009) showed that this estimator is robust with respect to the choice of prior when T is large. Manski (1987) proposes maximum score methods to estimate $\boldsymbol{\beta}$ when the distribution of the errors is unknown and γ is equal to zero for model (1). Smoothed maximum score estimators were developed by Horowitz (1992). See also Arellano (2003) for a survey of static probit models.

In this paper, we propose new estimators of γ and $\boldsymbol{\beta}$ in model (1) subject to (2) based on essentially a flat prior for the τ_i. This gives numerically tractable estimators which we show perform well in terms of mean squared error. Our methodology is designed for the cases when the individual effects $\tau_1, \cdots, \tau_n$ are large while T is small. Note that when the τ_i are large, there is an innate difficulty in estimating γ and $\boldsymbol{\beta}$ as the outcome of the random event $\{\tau_i + \gamma d_{i,t-1} + \boldsymbol{x}_{it}'\boldsymbol{\beta} + \epsilon_{it} > 0\}$ may be dominated by the value of τ_i. Heckman (1980) reported that the maximum likelihood estimator for γ behaved poorly when the variance of the τ_i is large; see Table 4.2 in Heckman (1980). Our simulation results indicate that our methods work as well as Heckman's (1980) method when the variance of the τ_i is small, for example, equal to 1 and 4.

The rest of the paper is organized as follows: Section 2 presents the new estimation methods together with their asymptotic properties for the case $T = 2$ and Section 3 gives an outline of the general case. For simplicity of the presentation, we only describe the case $T = 2$ in detail. Simulations are reported in Section 4 and an example is analyzed in Section 5. Some technical proofs are relegated to Appendix 1. Details of the extension of the proposed methods to the scenario with $T = 3$ are presented in Appendix 2.

2 Estimation of γ and $\boldsymbol{\beta}$ when $T = 2$

We consider model (1) subject to (2) for $T = 2$, specifically,

$$d_{i1} = I(\tau_i + \boldsymbol{x}_{i1}'\boldsymbol{\beta} + \epsilon_{i1} > 0), \quad d_{i2} = I(\tau_i + \gamma d_{i1} + \boldsymbol{x}_{i2}'\boldsymbol{\beta} + \epsilon_{i2} > 0), \quad i = 1, \cdots, n, \tag{3}$$

where the τ_i are random and independent of the ϵ_{i1} and ϵ_{i2}. Furthermore, we assume that the $\{\tau_i\}$ are mutually independent with a common density function $f(.)$ in a location-scale family:

C1 The density function of τ_i admits the expression

$$f(x) = \frac{1}{\sigma_\tau} h(\frac{x - \mu_\tau}{\sigma_\tau}), \tag{4}$$

where $h(\cdot)$ is a density function with mean 0 and variance 1, $h(x)$ is continuous at $x = 0$, and μ_τ and $\sigma_\tau > 0$ are constants.

In Section 2.1 we give an estimator of the autoregressive coefficient γ for the case without covariates (i.e. $\boldsymbol{\beta} = 0$), in Section 2.2 we show how the regression coefficient vector $\boldsymbol{\beta}$ for the static model (i.e. $\gamma = 0$) can be estimated, and in Section 2.3 we give a method to simultaneously estimate γ and $\boldsymbol{\beta}$. All the methods are based on an asymptotic argument which involves the variance of the τ_i going to infinity, and therefore the methods are particularly relevant when the individual effects are large.

2.1 Estimation of γ when $\boldsymbol{\beta} = 0$

When $\boldsymbol{\beta} = 0$, model (3) reduces to

$$d_{i1} = I(\tau_i + \epsilon_{i1} > 0), \qquad d_{i2} = I(\tau_i + \gamma d_{i1} + \epsilon_{i2} > 0), \qquad i = 1, \ldots, n. \tag{5}$$

As τ_i, ϵ_{i1} and ϵ_{i2} are independent, and ϵ_{i1} and ϵ_{i2} are $N(0,1)$, it holds that

$$P\{d_{i1} = 0, d_{i2} = 0\} = \int \Phi(-x)\Phi(-x)f(x)dx, \tag{6}$$

$$P\{d_{i1} = 0, d_{i2} = 1\} = \int \Phi(-x)\Phi(x)f(x)dx, \tag{7}$$

$$P\{d_{i1} = 1, d_{i2} = 0\} = \int \Phi(x)\Phi(-x-\gamma)f(x)dx, \tag{8}$$

$$P\{d_{i1} = 1, d_{i2} = 1\} = \int \Phi(x)\Phi(x+\gamma)f(x)dx, \tag{9}$$

where Φ is the standard normal distribution function, and $f(\cdot)$ is the density function of τ_i. The integrals can be hard to evaluate, making it hard to estimate γ. However, Proposition 1 shows that the integration can be avoided by assuming a 'flat' prior for the τ_i, i.e., letting σ_τ in C1 go to infinity. As we show below, this then leads to a simple estimator for γ.

Proposition 1. Suppose C1 holds. Then

$$\lim_{\sigma_\tau \to \infty} \frac{P\{d_{i1} = 1, d_{i2} = 0\}}{P\{d_{i1} = 0, d_{i2} = 1\}} = G(\gamma), \tag{10}$$

where

$$G(\gamma) = -\sqrt{\pi}\gamma\Phi(-\frac{\gamma}{\sqrt{2}}) + \exp\{-\frac{\gamma^2}{4}\}. \tag{11}$$

Proof. By Lemmas 1 resp. 2 in Appendix 1 we have

$$\lim_{\sigma_\tau\to\infty} \frac{P\{d_{i1}=1, d_{i2}=0\}}{P\{d_{i1}=0, d_{i2}=1\}} = \lim_{\sigma_\tau\to\infty} \frac{\int \Phi(x)\Phi(-x-\gamma)f(x)dx}{\int \Phi(-x)\Phi(x)f(x)dx} = \frac{\int \Phi(x)\Phi(-x-\gamma)dx}{\int \Phi(x)\Phi(-x)dx} = G(\gamma),$$

□

Proposition 1 suggests the following estimator for γ:

$$\widehat{\gamma} = G^{-1}(\widehat{W}), \tag{12}$$

where $G(\cdot)$ is given by (11), and

$$\widehat{W} = \frac{\sum_{i=1}^{n} I(d_{i1}=1,\ d_{i2}=0)}{\sum_{i=1}^{n} I(d_{i1}=0,\ d_{i2}=1)}, \tag{13}$$

i.e. $\widehat{W}$ is a plug-in estimator for the ratio of the two probabilities on the left hand side of (10). Theorem 1 shows an asymptotic normality property of $\widehat{\gamma}$.

Theorem 1. Suppose C1 holds with $\sigma_\tau = a\sqrt{n}$ for some constant $a > 0$. Then $\hat{\gamma}$ is a consistent estimator of γ, i.e.,

(i) $\lim_{n\to\infty} P\{|\hat{\gamma}-\gamma| \geq \eta\} = 0$ for all $\eta > 0$.

Suppose $h(\cdot)$ has a continuous derivative and

$$\kappa_n = \Big\{\sum_{i=1}^{n} I(d_{i1}=0,\ d_{i2}=1)\Big\}^{1/2}, \quad \sigma^2 = \frac{G(\gamma)+G^2(\gamma)}{[G'(\gamma)]^2} = \frac{G(\gamma)+G^2(\gamma)}{\pi\Phi^2(-\gamma/\sqrt{2})}. \tag{14}$$

Then it holds that as $n \to \infty$, $\kappa_n(\hat{\gamma}-\gamma)$ converges in distribution to a normal random variable with mean zero and variance σ^2, i.e.,

(ii) $\lim_{n\to\infty} P\{\kappa_n(\widehat{\gamma}-\gamma) \leq x\} = \Phi(x/\sigma)$ for all $x \in \mathbb{R}$.

Remark 1. The convergence rate of $\hat{\gamma}$ under the conditions of the theorem is $O(n^{-1/4})$, as can be seen as follows. From the proof of Theorem 1 in Appendix 1, we have

$$P(d_{i1}=0,\ d_{i2}=1) = f(\mu_\tau)\int \Phi(u)\Phi(-u)du + o_p(1/\sigma_\tau) = \frac{h(0)}{a\sqrt{n}}\int \Phi(u)\Phi(-u)du + o_p(1/\sqrt{n}).$$

This and the law of large numbers implies $\kappa_n = O(n^{1/4})$.

Remark 2. Note that the observations with $(d_{i1}, d_{i2}) = (0, 0)$ or $(1, 1)$ do not contribute to $\widehat{\gamma}$. In fact, when σ_τ is large, these observations provide little information on γ, since

$$\lim_{\sigma_\tau \to \infty} P\{d_{i1} = 0, d_{i2} = 0\} = \lim_{\sigma_\tau \to \infty} \int \Phi(-x)\Phi(-x)\frac{1}{\sigma_\tau} h(\frac{x-\mu_\tau}{\sigma_\tau})dx$$
$$= \lim_{\sigma_\tau \to \infty} \int \Phi(-\sigma_\tau t - \mu_\tau)\Phi(-\sigma_\tau t - \mu_\tau)h(t)dt = H(0),$$

and similarly

$$\lim_{\sigma_\tau \to \infty} P\{d_{i1} = 1, d_{i2} = 1\} = 1 - H(0).$$

where $H(x)$ is cumulative distribution function of $h(x)$.

2.2 Estimation of $\boldsymbol{\beta}$ when $\gamma = 0$

Let D_n be the set of pairs (d_{i1}, d_{i2}) equal to $(0, 1)$ or $(1, 0)$, i.e.,

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

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

We find the conditional probability

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

Under (4), we can prove analogously to the proof of Proposition 1 that

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

where G is given by (11). Hence for sufficiently large σ_τ, a good approximation of the conditional likelihood of $\boldsymbol{\beta}$ given D_n is

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{m} p_i^{z_i}(1 - p_i)^{1 - z_i} \tag{15}$$

where $z_i = I(d_{i1} = 1, d_{i2} = 0)$ and $1 - z_i = I(d_{i1} = 0, d_{i2} = 1)$, and

$$p_i = \frac{G((\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\boldsymbol{\beta})}{G((\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\boldsymbol{\beta}) + G(-(\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\boldsymbol{\beta})}. \tag{16}$$

Note that $p_i = K((\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\boldsymbol{\beta})$ for the monotone function K defined as

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

Hence, (16) is a generalized linear model of the form

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

So iterative reweighted least squares methods for generalized linear models given by McCullagh and Nelder (1989) can be applied to (15) to estimate the parameter $\boldsymbol{\beta}$. Under some regularity conditions consistency of $\boldsymbol{\beta}$ can be shown by letting $\sigma_\tau \to \infty$.

2.3 Simultaneous estimation of γ and $\boldsymbol{\beta}$

As in Section 2.2, we have

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

where G is given by (11). For large σ_τ, we replace the conditional likelihood of γ and $\boldsymbol{\beta}$ given D_n by

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{m} p_i^{z_i}(1 - p_i)^{1 - z_i} \tag{17}$$

where $z_i = I(\{d_{i1} = 1, d_{i2} = 0)$ and $1 - z_i = I(d_{i1} = 0, d_{i2} = 1)$, m is the number of $d_{i1} + d_{i2}$ which are equal to 1 and

$$p_i = \frac{G(\gamma + (\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\boldsymbol{\beta})}{G(\gamma + (\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\boldsymbol{\beta}) + G(-(\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1})'\boldsymbol{\beta})}. \tag{18}$$

Let

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

Theorem 2. Let $(p_1, \ldots, p_m)$ be a given probability distribution in (18). Then (17) has a unique solution for γ and $\boldsymbol{\beta}$ if the following conditions hold:

(a) The rank of $\boldsymbol{X}^*$ is equal to k (the dimension of $\boldsymbol{x}_{2i} - \boldsymbol{x}_{1i}$);

(b) There exist j and $1 \le s_1, \cdots, s_k \le m$ such that

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

where $a_1, \cdots, a_k$ are non-positive real numbers.

The conditions in Theorem 2 are sufficient and can be satisfied with probability close to 1 for a large sample size n if the covariate $\boldsymbol{x}_{i2} - \boldsymbol{x}_{i1}$ is a continuous variable and its covariance matrix is positive definite.

Corollary. Under the conditions in Theorem 2, and with $\mathbf{1}_m$ be the $m-$dimensional vector with all components 1, the rank of $(\mathbf{1}_m, \boldsymbol{X}^{*'})$ is $k+1$.

From the Corollary, it seems that the identifiability condition relating to (18) is stronger than that of linear models since that the rank of design matrix being equal to the number of parameters is sufficient for linear models to be identified.

3 Outline of the general case: estimating γ and $\boldsymbol{\beta}$ when $T \geq 2$

The methods of Section 2 can be extended in a fairly straightforward manner to more than two time points. Below we give an outline, further technical details are given in Appendix 2 where the case $T = 3$ is described in some detail.

First let us define the following probability function p:

$$p(d_{i1}, d_{i2}, \cdots, d_{iT} | d_{i1} + \cdots + d_{iT} = 0, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT}) \equiv 1,$$

$$p(d_{i1}, d_{i2}, \cdots, d_{iT} | d_{i1} + \cdots + d_{iT} = T, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT}) \equiv 1,$$

and for $s = 1, 2, \cdots, T-1$,

$$\begin{aligned} & p(d_{i1}, d_{i2}, \cdots, d_{iT} | d_{i1} + \cdots + d_{iT} = s, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT}) \\ = c_s \int \Phi\Big((2d_{i1} - 1)(u + \boldsymbol{x}'_{i1}\boldsymbol{\beta})\Big) & \times \cdots \times \Phi\Big((2d_{iT} - 1)(u + \boldsymbol{x}'_{iT}\boldsymbol{\beta} + \gamma d_{i\,T-1})\Big)\, du \end{aligned}$$

where c_s is a normalizing constant chosen so that

$$\sum_{d_{i1} + \cdots + d_{iT} = s} p(d_{i1}, d_{i2}, \cdots, d_{iT} | d_{i1} + \cdots + d_{iT} = s, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT}) = 1.$$

is satisfied.

Using methods analogous to the ones of Section 2 and Appendix 1, we can show that

$$\begin{aligned} & \lim_{\sigma_\tau \to \infty} P\{d_{i1}, d_{i2}, \cdots, d_{iT} | d_{i1} + \cdots + d_{iT} = s, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT}\} \\ & = p(d_{i1}, d_{i2}, \cdots, d_{iT} | d_{i1} + \cdots + d_{iT} = s, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT}). \end{aligned}$$

By the asymptotic argument used in Section 2, we can estimate γ and $\boldsymbol{\beta}$ by the maximizer of

$$\prod_{i=1}^{n} p(d_{i1}, d_{i2}, \cdots, d_{iT} | d_{i1} + \cdots + d_{iT} = s, \boldsymbol{x}_{i1}, \cdots, \boldsymbol{x}_{iT}). \tag{19}$$

This maximization is computationally straightforward.

4 Simulation study

In this section, we use simulations to estimate the root mean squared errors (RMSEs) of the estimators proposed in Section 2. In Table 1, RMSEs of γ in Model (5) are given for different distributions of the individual effects. In Table 2, RMSEs of γ and β in Model (1) are given, with the x_{i1} sampled from the standard normal distribution and $x_{i2} = x_{i1} + N(0,1)$; the individual effects are normally distributed with mean 0 and variance 2. For normally distributed individual effects with mean 0 and variance σ^2 in Model (1), Heckman (1980) has proposed the maximum likelihood estimation of the dynamic parameter γ and σ^2. In Tables 3 and 4 the RMSE of our new estimator is compared with the RMSE of Heckman's estimator, in the former table for normally distributed individual effects and in the latter for individual effects with a mixture normal distribution. We see that our estimator is comparable to Heckman's for normally distributed effects with moderate variance, but greatly outperforms it when individual effects are mixed normal distributions. We also compare our proposed estimator with the modified profile likelihood estimator(MPL) for fixed effects, which is given by Bartolucci, Bellio, Salvan and Sartori (2014). Since the modified profile likelihood estimator does not exist for $T = 2$, we do simulations with $T = 3$. No covariates are assumed and the individual effects are assigned a Student t-distribution with $df = 3$. Simulation results are listed in In Table 5.

5 Real data example

We analyze the data set listed in Table 6 which has previously been considered by Heckman (1981). The dynamics of female labor supply is investigated based on panel data from the years 1968 to 1970, and 1971 to 1973. Model (1) is applied to estimate the dynamic parameter with $T = 3$ and $x_{it} \equiv 0$. Let n_{rst} be the number of observations of runs pattern

Table 1: Simulated RMSEs of the new estimator of the dynamic parameter γ in Model (5) ($T = 2$ and 100 replications)

		$n = 1000$		$n = 5000$
Distribution of the τ_i	γ	RMSE	Distribution of the τ_i	RMSE
	-2	0.16		0.21
	-1.5	0.24		0.19
	-1	0.23		0.14
	-0.5	0.20		0.15
U(-3,3)	0	0.15	U(-10,10)	0.13
	0.5	0.21		0.31
	1	0.18		0.14
	1.5	0.15		0.15
	2	0.25		0.18
	-2	0.30		0.16
	-1.5	0.15		0.19
	-1	0.20		0.13
	-0.5	0.15		0.12
N(0,4)	0	0.15	N(0,25)	0.11
	0.5	0.16		0.10
	1	0.17		0.11
	1.5	0.18		0.12
	2	0.23		0.17

Table 2: Simulated RMSEs of new estimators of γ and β for Model (1) ($T = 2$, 200 replicates and $n = 1000$)

γ	β	RMSE($\hat{\gamma}$)	RMSE($\hat{\beta}$)	γ	β	RMSE($\hat{\gamma}$)	RMSE($\hat{\beta}$)
-1	0	0.20	0.08	0	-1	0.20	0.15
-0.5	0	0.17	0.08	0	-0.5	0.18	0.10
0	0	0.14	0.08				
0.5	0	0.16	0.08	0	0.5	0.16	0.10
1	0	0.16	0.09	0	1	0.19	0.13
-1	1	0.22	0.13	1	1	0.25	0.16
-0.5	0.5	0.19	0.10	0.5	0.5	0.15	0.09
0.5	-0.5	0.16	0.10	-0.5	-0.5	0.17	0.10
1	-1	0.22	0.18	-1	-1	0.24	0.13

Table 3: Comparison of RMSE of new estimator $\hat{\gamma}_G$ and Heckman's $\hat{\gamma}_H$ for normally distributed individual effects ($T = 2$, 200 replicates for sample size $n = 1000$).

Distribution of the τ_i	γ	RMSE($\hat{\gamma}_G$)	RMSE($\hat{\gamma}_H$)	RMSE($\hat{\sigma}_H$)
	-1	0.16	0.13	0.13
	-0.5	0.14	0.11	0.12
$N(0,1)$	0	0.12	0.09	0.11
	0.5	0.13	0.10	0.11
	1	0.13	0.10	0.12
	-1	0.20	0.16	0.25
	-0.5	0.18	0.15	0.21
$N(0,4)$	0	0.15	0.12	0.18
	0.5	0.17	0.14	0.26
	1	0.17	0.15	0.20

Table 4: Comparison of RMSE of new estimator $\hat{\gamma}_G$ and Heckman's $\hat{\gamma}_H$ for individual effects distributed as $0.5N(-6, 9) + 0.5N(6, 9)$ ($T = 2$, 200 replicates with sample size $n = 3000$).

γ	RMSE($\hat{\gamma}_G$)	RMSE($\hat{\gamma}_H$)	RMSE($\hat{\sigma}_H$)
-1	0.37	0.81	3.81
-0.5	0.29	0.75	3.82
0	0.30	0.64	3.86
0.5	0.29	0.59	3.81
1	0.30	0.53	3.85

Table 5: Comparison of RMSE of new estimator $\hat{\gamma}_G$ and the MPL estimator $\hat{\gamma}_{MPL}$ for individual effects with t-distributions with $df = 3$ ($T = 3$, and 100 replicates)

	γ	RMSE($\hat{\gamma}_G$)	RMSE($\hat{\gamma}_{MPL}$)		γ	RMSE($\hat{\gamma}_G$)	RMSE($\hat{\gamma}_{MPL}$)
	-1	0.51	0.25		-1	0.16	0.15
	-0.5	0.36	0.38		-0.5	0.13	0.27
n=100	0	0.24	0.43	n=500	0	0.12	0.38
	0.5	0.25	0.44		0.5	0.11	0.36
	1	0.32	0.47		1	0.13	0.37
	-1	0.27	0.18		-1	0.11	0.13
	-0.5	0.19	0.32		-0.5	0.09	0.28
n=200	0	0.18	0.39	n=1000	0	0.09	0.36
	0.5	0.18	0.41		0.5	0.09	0.35
	1	0.19	0.42		1	0.09	0.35

(r, s, t) in Table 6 for $r, s, t = 0, 1$. The resulting estimates are listed in Table 7, where $\hat{\gamma}_G$ is the new estimator and $\hat{\gamma}_H$ and $\hat{\sigma}_H$ are Heckman's estimators.

From the analyzed results in the age group 49-59 and runs pattern from 1971 to 1973, neither Heckman's method nor the proposed method yield evidence of a dynamic relationship, and perhaps more data needs to be collected. However, the difference for the older group between the period 1968-170 and 1971-1973 is significant; the difference for the younger group between the period 1968-170 and 1971-1973 is not significant. For age group 30-44, both the proposed method and Heckman's method yield a significant dynamic relationship, with a positive estimated value of γ (here, positivity of γ implies the unsurprising result that currently holding a job increases the likelihood of holding a job in future).

Acknowledgements

We are grateful to the reviewers and the associate editor for their constructive comments that greatly improved the paper. We also thank Professor Bartolucci and Bellio for their generously supplying the related codes and programs during our simulation studies. QY's research is partially supported by EPSRC research grant EP/L01226X/1.

Table 6: Runs patterns in the data (1 corresponds to work in the year, 0 corresponds to no work)

Runs patterns 1968	1969	1970	No. of observations	Runs pattern 1971	1972	1973	No.of observations
			women aged 45-59 in 1968				
0	0	0	87	0	0	0	96
0	0	1	5	0	0	1	5
0	1	0	5	0	1	0	4
1	0	0	4	1	0	0	8
1	1	0	8	1	1	0	5
0	1	1	10	0	1	1	2
1	0	1	1	1	0	1	2
1	1	1	78	1	1	1	76
			women aged 30-44 in 1968				
0	0	0	126	0	0	0	133
0	0	1	16	0	0	1	13
0	1	0	4	0	1	0	5
1	0	0	12	1	0	0	16
1	1	0	24	1	1	0	8
0	1	1	20	0	1	1	19
1	0	1	5	1	0	1	8
1	1	1	125	1	1	1	130

Table 7: Comparison of new estimator ($\hat{\gamma}_G$) with Heckman's ($\hat{\gamma}_H$) for data in Table 6

panel data (1969-1970)			panel data (1971-1973)		
$\hat{\gamma}_G$ (s.e.)	$\hat{\gamma}_H$ (s.e.)	$\hat{\sigma}_H$ (s.e.)	$\hat{\gamma}_G$ (s.e.)	$\hat{\gamma}_H$ (s.e.)	$\hat{\sigma}_H$ (s.e.)
		women aged 45-59 in 1968			
0.62 (0.20)	0.54 (0.27)	3.24 (0.65)	−0.16 (0.26)	−0.28 (0.36)	5.59 (1.33)
		women aged 30-44 in 1968			
0.48 (0.13)	0.47 (0.17)	2.15 (0.28)	0.51 (0.14)	0.43 (0.19)	2.63 (0.37)

Appendix 1: Technical proofs

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.

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

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

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

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

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

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

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

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

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

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

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

Lemma 3. Suppose $\sigma_\tau = a\sqrt{n}(a > 0)$. Then

$$\frac{1}{n^{1/4}}\begin{pmatrix} \sum_{i=1}^{n}\left[I_{\{d_{i1}=1,d_{i2}=0\}} - EI_{\{d_{i1}=1,d_{i2}=0\}}\right] \\ \\ \sum_{i=1}^{n}\left[I_{\{d_{i1}=0,d_{i2}=1\}} - EI_{\{d_{i1}=0,d_{i2}=1\}}\right] \end{pmatrix} \xrightarrow{d} N(0,\Sigma)$$

where

$$\Sigma = \frac{h(0)}{a}\begin{pmatrix} \int \Phi(x)\Phi(-x-\gamma)dx & 0 \\ \\ 0 & \int \Phi(x)\Phi(-x)dx \end{pmatrix}.$$

Proof: For $c_1, c_2 \in R$, let

$$U_{i\,n} = c_1\left[I_{\{d_{i1}=1,d_{i2}=0\}} - EI_{\{d_{i1}=1,d_{i2}=0\}}\right] + c_2\left[I_{\{d_{i1}=0,d_{i2}=1\}} - EI_{\{d_{i1}=0,d_{i2}=1\}}\right]$$

Then

$$E(U_{i\,n}) = 0, \quad \sqrt{n}E(U_{i\,n}^2) = \frac{h(0)}{a}\left[c_1^2\int \Phi(x)\Phi(-x-\gamma)dx + c_2^2\int \Phi(x)\Phi(-x)dx\right] + o(1).$$

By simple computations,

$$\begin{aligned}
E[\exp\{U_{i\,n}t/n^{1/4}\}] &= 1 + \frac{t^2}{2\sqrt{n}}E(U_{i\,t}^2) + E[o(\frac{U_{i\,n}^2}{n^{1/2}})] \\
&= 1 + \frac{t^2}{2\sqrt{n}}E(U_{i\,t}^2) + o(n^{-1}) \\
&= 1 + \frac{h(0)\left[c_1^2\int \Phi(x)\Phi(-x-\gamma)dx + c_2^2\int \Phi(x)\Phi(-x)dx\right]t^2}{2an} + o(n^{-1}).
\end{aligned}$$

The moment generating function of $\sum_{i=1}^{n} U_{i\,n}/n^{1/4}$ is

$$\begin{aligned}
\phi_n(t) &= E[\exp\{\sum_{i=1}^{n} U_{i\,n}t/n^{1/4}\}] \\
&= [E(\exp\{U_{i\,n}t/n^{1/4}\})]^n \\
&= \left\{1 + \frac{h(0)\left[c_1^2 \int \Phi(x)\Phi(-x-\gamma)dx + c_2^2 \int \Phi(x)\Phi(-x)dx\right] t^2}{2an} + o(n^{-1})\right\}^n \\
&\longrightarrow \exp\{\frac{ah(0)\left[c_1^2 \int \Phi(x)\Phi(-x-\gamma)dx + c_2^2 \int \Phi(x)\Phi(-x)dx\right] t^2}{2a}\}
\end{aligned}$$

which implies the Lemma holds.

Lemma 4. Suppose $\sigma_\tau = a\sqrt{n}(a > 0)$ and the first derivative of $h(x)$ is continuous. Then

$$n^{1/4} \begin{pmatrix} \frac{\sum_{i=1}^{n} I_{\{d_{i1}=1, d_{i2}=0\}}}{\sqrt{n}} - \frac{h(0)}{a} \int \Phi(x)\Phi(-x-\gamma)dx \\ \\ \frac{\sum_{i=1}^{n} I_{\{d_{i1}=0, d_{i2}=1\}}}{\sqrt{n}} - \frac{h(0)}{a} \int \Phi(x)\Phi(-x)dx \end{pmatrix} \xrightarrow{d} N(0, \Sigma)$$

where

$$\Sigma = \frac{h(0)}{a} \begin{pmatrix} \int \Phi(x)\Phi(-x-\gamma)dx & 0 \\ \\ 0 & \int \Phi(x)\Phi(-x)dx \end{pmatrix}.$$

Proof: Since the first derivative of $h(x)$ is continuous and $\sigma_\tau = a\sqrt{n}$, we have

$$\begin{aligned}
\sqrt{n} \times EI_{\{d_{11}=1, d_{12}=0\}} &= \sqrt{n} \times \int \Phi(x)\Phi(-x-\gamma)f(x)dx \\
&= \sqrt{n} \times \int \Phi(x)\Phi(-x-\gamma)\frac{1}{\sigma_\tau}h(\frac{x-\mu_\tau}{\sigma_\tau})dx \\
&= \sqrt{n} \times \int \Phi(x)\Phi(-x-\gamma)\frac{1}{a\sqrt{n}}h(\frac{x-\mu_\tau}{\sigma_\tau})dx \\
&= \frac{1}{a} \int \Phi(x)\Phi(-x-\gamma)h(\frac{x-\mu_\tau}{\sigma_\tau})dx \\
&= \frac{h(0)}{a} \int \Phi(x)\Phi(-x-\gamma)dx + O(\sigma_\tau^{-1}) \\
&= \frac{h(0)}{a} \int \Phi(x)\Phi(-x-\gamma)dx + O(n^{-1/2}).
\end{aligned}$$

Similarly, we can obtain

$$\sqrt{n} \times EI_{\{d_{11}=0, d_{12}=1\}} = \frac{h(0)}{a} \int \Phi(x)\Phi(-x)dx + O(n^{-1/2}).$$

$$n^{1/4}\begin{pmatrix}\frac{\sum_{i=1}^n I_{\{d_{i1}=1,d_{i2}=0\}}}{\sqrt{n}}-\frac{h(0)}{a}\int\Phi(x)\Phi(-x-\gamma)dx\\ \frac{\sum_{i=1}^n I_{\{d_{i1}=0,d_{i2}=1\}}}{\sqrt{n}}-\frac{h(0)}{a}\int\Phi(x)\Phi(-x)dx\end{pmatrix}$$

$$=n^{1/4}\begin{pmatrix}\frac{\sum_{i=1}^n [I_{\{d_{i1}=1,d_{i2}=0\}}-EI_{\{d_{i1}=1,d_{i2}=0\}}]}{\sqrt{n}}+\sqrt{n}EI_{\{d_{11}=1,d_{12}=0\}}-\frac{h(0)}{a}\int\Phi(x)\Phi(-x-\gamma)dx\\ \frac{\sum_{i=1}^n [I_{\{d_{i1}=0,d_{i2}=1\}}-EI_{\{d_{i1}=0,d_{i2}=1\}}]}{\sqrt{n}}+\sqrt{n}EI_{\{d_{11}=0,d_{12}=1\}}-\frac{h(0)}{a}\int\Phi(x)\Phi(-x)dx\end{pmatrix}$$

$$=n^{1/4}\begin{pmatrix}\frac{\sum_{i=1}^n [I_{\{d_{i1}=1,d_{i2}=0\}}-EI_{\{d_{i1}=1,d_{i2}=0\}}]}{\sqrt{n}}\\ \frac{\sum_{i=1}^n [I_{\{d_{i1}=0,d_{i2}=1\}}-EI_{\{d_{i1}=0,d_{i2}=1\}}]}{\sqrt{n}}\end{pmatrix}+n^{1/4}\begin{pmatrix}\sqrt{n}EI_{\{d_{11}=1,d_{12}=0\}}-\frac{h(0)}{a}\int\Phi(x)\Phi(-x-\gamma)dx\\ +\sqrt{n}EI_{\{d_{11}=0,d_{12}=1\}}-\frac{h(0)}{a}\int\Phi(x)\Phi(-x)dx\end{pmatrix}$$

$$=n^{-1/4}\begin{pmatrix}\sum_{i=1}^n [I_{\{d_{i1}=1,d_{i2}=0\}}-EI_{\{d_{i1}=1,d_{i2}=0\}}]\\ \sum_{i=1}^n [I_{\{d_{i1}=0,d_{i2}=1\}}-EI_{\{d_{i1}=0,d_{i2}=1\}}]\end{pmatrix}+o(1)$$

which implies the Lemma holds by Lemma 3.

Proof of Theorem 1. To demonstrate (i), by Lemma 4, we have

$$\frac{\sum_{i=1}^n I_{\{d_{i1}=1,d_{i2}=0\}}}{\sqrt{n}}-\frac{h(0)}{a}\int\Phi(x)\Phi(-x-\gamma)dx=o_p(1)$$

and

$$\frac{\sum_{i=1}^n I_{\{d_{i1}=0,d_{i2}=1\}}}{\sqrt{n}}-\frac{h(0)}{a}\int\Phi(x)\Phi(-x)dx=o_p(1).$$

Then

$$\widehat{W}=\frac{\sum_{i=1}^n I_{\{d_{i1}=1,d_{i2}=0\}}}{\sum_{i=1}^n I_{\{d_{i1}=0,d_{i2}=1\}}}=\frac{\sum_{i=1}^n I_{\{d_{i1}=1,d_{i2}=0\}}/\sqrt{n}}{\sum_{i=1}^n I_{\{d_{i1}=0,d_{i2}=1\}}/\sqrt{n}}\to_p G(\gamma).$$

(i) follows immediately from the continuity of $G^{-1}(x)$.

To prove (ii), it follows from the delta method and Lemma 4 above that

$$n^{1/4}\left(\hat{W}-G(\gamma)\right) = n^{1/4}\left(\frac{\sum_{i=1}^{n} I_{\{d_{i1}=1,d_{i2}=0\}}/\sqrt{n}}{\sum_{i=1}^{n} I_{\{d_{i1}=0,d_{i2}=1\}}/\sqrt{n}} - \frac{\frac{h(0)}{a}\int\Phi(x)\Phi(-x-\gamma)dx}{\frac{h(0)}{a}\int\Phi(x)\Phi(-x)dx}\right)$$

$$\xrightarrow{d} N(0,\sigma^{*2})$$

where

$$\sigma^{*2}=\frac{a\int\Phi(x)\Phi(-x-\gamma)dx}{h(0)[\int\Phi(x)\Phi(-x)dx]^2}+\frac{a[\int\Phi(x)\Phi(-x-\gamma)dx]^2}{h(0)[\int\Phi(x)\Phi(-x)dx]^3}.$$

Then

$$n^{1/4}\left(\hat{\gamma}-\gamma\right)=n^{1/4}\left(G^{-1}(W)-G^{-1}(G(\gamma))\right)\xrightarrow{d} N(0,\frac{\sigma^{*2}}{[G'(\gamma)]^2}).$$

So

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

by

$$\frac{\sum_{i=1}^{n} I_{\{d_{i1}=0,\ d_{i2}=1\}}}{\sqrt{n}}\xrightarrow{p}\frac{h(0)\int\Phi(x)\Phi(-x)dx}{a}.$$

Lemma 5. 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

$$\begin{cases} G(\boldsymbol{x}_1'\boldsymbol{\beta}+\alpha)-r_1G(-\boldsymbol{x}_1'\boldsymbol{\beta})=0\\ G(\boldsymbol{x}_2'\boldsymbol{\beta}+\alpha)-r_2G(-\boldsymbol{x}_2'\boldsymbol{\beta})=0\\ \cdots\cdots\\ G(\boldsymbol{x}_k'\boldsymbol{\beta}+\alpha)-r_kG(-\boldsymbol{x}_k'\boldsymbol{\beta})=0\\ G(\boldsymbol{x}_{k+1}'\boldsymbol{\beta}+\alpha)-r_{k+1}G(-\boldsymbol{x}_{k+1}'\boldsymbol{\beta})=0 \end{cases} \tag{20}$$

has a unique solution $\boldsymbol{\beta}$ 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 decreasing 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

$$\begin{cases} G(\boldsymbol{x}_1'\boldsymbol{\beta}+\alpha)-r_1G(-\boldsymbol{x}_1'\boldsymbol{\beta})=0\\ G(\boldsymbol{x}_2'\boldsymbol{\beta}+\alpha)-r_2G(-\boldsymbol{x}_2'\boldsymbol{\beta})=0\\ \qquad\qquad\cdots\cdots\\ G(\boldsymbol{x}_k'\boldsymbol{\beta}+\alpha)-r_kG(-\boldsymbol{x}_k'\boldsymbol{\beta})=0\end{cases} \tag{21}$$

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

Let $\boldsymbol{\beta}^*=(\boldsymbol{\beta}_1(\alpha),\cdots,\boldsymbol{\beta}_k(\alpha))'$ the solution of (21). Then

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

where

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

and

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

Define

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

Then

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

which implies $t(\alpha)=0$ has an unique solution and the lemma is proved.

Proof of Theorem 2. By Lemma 5 given in the above, it can be proved with $r_i = p_i/(1-p_i)$ and $\boldsymbol{x}_i = \boldsymbol{x}_{i2} - \boldsymbol{x}_{i1}$.

Proof of Corollary. Without loss of generality, suppose that $\boldsymbol{x}_{12} - \boldsymbol{x}_{11}, \cdots, \boldsymbol{x}_{k2} - \boldsymbol{x}_{k1}$ are linearly 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 a non-positive real number. Then the determinant

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

is equal to

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

by the assumption. This implies that the rank of (22) is $k+1$.

Since the rank of $(\mathbf{1}_m, \boldsymbol{X}^{*'})$ is equal to that of $(\boldsymbol{X}^{*'}, \mathbf{1}_m)$, which is a $m \times (k+1)$ matrix, and (22) is a matrix obtained by the first $k+1$ rows of $(\boldsymbol{X}^{*'}, \mathbf{1}_m)$, thus the rank of $(\mathbf{1}_m, \boldsymbol{X}^{*'})$ is $k+1$.

Appendix 2: Extension to $T = 3$

To generalize the proposed methods given in Section 2 to the case $T > 3$, we recap the main idea for $T = 2$ first. It follows from (12) that $G(\widehat{\gamma}) = \widehat{W}$. Thus

$$G(\widehat{\gamma}) = \frac{\int \Phi(x)\Phi(-x-\widehat{\gamma})dx}{\int \Phi(x)\Phi(-x)dx} = \frac{\sum_{i=1}^{n} I(d_{i1} = 1,\ d_{i2} = 0)}{\sum_{i=1}^{n} I(d_{i1} = 0,\ d_{i2} = 1)} = \widehat{W}.$$

Then

$$\frac{\int \Phi(x)\Phi(-x-\widehat{\gamma})dx}{\int \Phi(x)\Phi(-x)dx + \int \Phi(x)\Phi(-x-\widehat{\gamma})} = \frac{\sum_{i=1}^{n} I(d_{i1}=1,\ d_{i2}=0)}{\sum_{i=1}^{n} I(d_{i1}=0,\ d_{i2}=1) + \sum_{i=1}^{n} I(d_{i1}=1,\ d_{i2}=0)}.$$

If we let

$$\begin{aligned} p(\gamma) &= \frac{\int \Phi(x)\Phi(-x-\gamma)dx}{\int \Phi(x)\Phi(-x)dx + \int \Phi(x)\Phi(-x-\gamma)} \\ &= \lim_{\sigma_\tau \to \infty} \frac{P\{d_{i1}=1, d_{i2}=0\}}{P\{d_{i1}=0, d_{i2}=1\} + P\{d_{i1}=1, d_{i2}=0\}} \\ &= \lim_{\sigma_\tau \to \infty} P\{d_{i1}=1, d_{i2}=0 \big| d_{i1}+d_{i2}=1\}, \end{aligned}$$

then

$$p(\widehat{\gamma}) = \arg\max\{[p(\gamma)]^{n_{10}}[1-p(\gamma)]^{n_{01}}\}$$

where

$$n_{10} = \sum_{i=1}^{n} I(d_{i1}=1,\ d_{i2}=0), \quad n_{01} = \sum_{i=1}^{n} I(d_{i1}=0,\ d_{i2}=1).$$

Here $\widehat{\gamma}$ is the conditional maximum likelihood estimation of γ under $d_{i1}+d_{i2}=1$, based on the likelihood obtained from $p(\gamma)$. Based on the above results, we can generalize our results to the case $T > 2$. To illustrate how this is done, we consider the case of $T = 3$ without covariates. The more general case can be derived analogously but requires more complex notation. In the case $T = 3$ there are three observations for each individual

$$d_{i1} = I(\tau_i + \epsilon_{i1} > 0), \quad d_{i2} = I(\tau_i + \gamma d_{i1} + \epsilon_{i2} > 0), \quad d_{i3} = I(\tau_i + \gamma d_{i2} + \epsilon_{i3} > 0).$$

For each individual i, $d_{i1}+d_{i2}+d_{i3} = 0, 1, 2$ or 3. As for $T = 2$, units for which there is no change (i.e., for $T = 3$, if $d_{i1}+d_{i2}+d_{i3}$ is equal to 0 and 3) provide little information about γ, so we delete these cases.

For $d_{i1}+d_{i2}+d_{i3} = 1$ or 2, we can prove analogously to Lemma 1

$$p_{100}(\gamma) = \lim_{\sigma_\tau \to \infty} P\{d_{i1}=1, d_{i2}=0, d_{i3}=0 \big| d_{i1}+d_{i2}+d_{i3}=1\} = \frac{\int \Phi(t)\Phi(-t-\gamma)\Phi(-t)dt}{K_1},$$

$$p_{010}(\gamma) = \lim_{\sigma_\tau \to \infty} P\{d_{i1}=0, d_{i2}=1, d_{i3}=0 \big| d_{i1}+d_{i2}+d_{i3}=1\} = \frac{\int \Phi(-t)\Phi(t)\Phi(-t-\gamma)dt}{K_1},$$

$$p_{001}(\gamma) = \lim_{\sigma_\tau \to \infty} P\{d_{i1} = 0, d_{i2} = 0, d_{i3} = 1 \big| d_{i1} + d_{i2} + d_{i3} = 1\} = \frac{\int \Phi(-t)\Phi(-t)\Phi(t)dt}{K_1},$$

$$p_{110}(\gamma) = \lim_{\sigma_\tau \to \infty} P\{d_{i1} = 1, d_{i2} = 0, d_{i3} = 0 \big| d_{i1}+d_{i2}+d_{i3} = 2\} = \frac{\int \Phi(t)\Phi(t+\gamma)\Phi(-t-\gamma)dt}{K_2},$$

$$p_{101}(\gamma) = \lim_{\sigma_\tau \to \infty} P\{d_{i1} = 0, d_{i2} = 1, d_{i3} = 0 \big| d_{i1} + d_{i2} + d_{i3} = 2\} = \frac{\int \Phi(t)\Phi(-t-\gamma)\Phi(t)}{K_2},$$

$$p_{011}(\gamma) = \lim_{\sigma_\tau \to \infty} P\{d_{i1} = 0, d_{i2} = 0, d_{i3} = 1 \big| d_{i1} + d_{i2} + d_{i3} = 2\} = \frac{\int \Phi(-t)\Phi(t)\Phi(t+\gamma)dt}{K_2}$$

where

$$K_1 = \int \Phi(t)\Phi(-t-\gamma)\Phi(-t)dt + \int \Phi(-t)\Phi(t)\Phi(-t-\gamma)dt + \int \Phi(-t)\Phi(-t)\Phi(t)dt$$

and

$$K_2 = \int \Phi(t)\Phi(t+\gamma)\Phi(-t-\gamma)dt + \int \Phi(t)\Phi(-t-\gamma)\Phi(t) + \int \Phi(-t)\Phi(t)\Phi(t+\gamma)dt.$$

Thus the conditional maximum likelihood estimation of γ, conditioning on $d_{i1} + d_{i2} + d_{i3}$ being equal to 1 or 2, is

$$\widehat{\gamma} = \arg\max_\gamma \{[p_{100}(\gamma)]^{n_{100}} [p_{010}(\gamma)]^{n_{010}} [p_{001}(\gamma)]^{n_{001}} [p_{110}(\gamma)]^{n_{110}} [p_{101}(\gamma)]^{n_{101}} [p_{011}(\gamma)]^{n_{011}}\}$$

where

$$n_{rst} = \sum_{i=1}^{n} I(d_{i1} = r,\ d_{i2} = s, d_{i3} = t), \quad r, s, t = 0, 1.$$

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.

Bartolucci, F. ,Bellio, R., Salvan, A. and Sartori, N. (2014). Modified Profile Likelihood for Fixed-Effects Panel Data Models, Econometric Reviews, DOI: 10.1080/07474938.2014.975642.

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.(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.

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.(1981). Heterogeneity and state dependence. *Studies in Labor Markets*, ed by S. Rosen, p91-140. University of Chicago Press.

Hisao, C.(2003). *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. *Econometrica*, 60, 505-531.

Lancaster, T.(2000). The incidental parameter problem since 1984. *Journal of Econometrics*, 95, 391-413.

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.

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

Neyman, J. and Scott, E. S.(1948). Consistent estimation from partially consistent observations. *Econometrica*, 16, 1-32.