A hyperbolic transformation for a fixed effects logit model*

Yoshitsugu Kitazawa†

June 11, 2010

Abstract
In this paper, a simple transformation is proposed for the fixed effects logit model, using which some valid moment conditions including the first-order condition for one of the conditional MLE proposed by Chamberlain (1980) can be generated. Some Monte Carlo experiments are carried out for the GMM estimator based on the transformation.

JEL classification: C23
Keywords: fixed effects logit; conditional logit estimator; hyperbolic transformation; moment conditions; GMM; Monte Carlo experiments

1. Introduction

Chamberlain (1980) proposes an useful and established estimator for the fixed effects logit model in panel data. This estimator is referred to as the conditional logit estimator, which maximizes the likelihood function composed of the probabilities of the (binary) dependent variables conditional on the fixed effects, the (real-valued) explanatory variables and the intertemporal sums of the dependent variables. The conditional logit estimator is consistent for the situation of small number of time periods and large cross-sectional size, since its conditional likelihood function rules out the fixed effects.[1]

This paper advocates another method of consistently estimating the fixed effects logit model for the situation of small number of time periods and large cross-sectional size. The procedure of the method is as follows: Firstly, a hyperbolic transformation is applied to the fixed effects logit model with the aim of eliminating the fixed effects. Next, the GMM (generalized method of moments) estimator proposed by Hansen (1982) is constructed by using the moment conditions based on the hyperbolic transformation. It will be seen that these moment conditions include one type of the first-order conditions of the likelihood for the conditional logit estimator. Then, the preferable small sample property of the GMM estimator using the moment conditions based on the hyperbolic transformation is shown by some Monte Carlo experiments.

The rest of the paper is as follows. Section 2 presents the implicit form of the fixed effects logit model, the moment conditions based on the hyperbolic transformation and the GMM estimator. Section 3 illustrates the link between the conditional maximum likelihood estimator (CMLE) in the first paragraph and the GMM estimator for the case of two periods. Section 4 reports some Monte Carlo results for the GMM estimator. Section 5 concludes.

2. Fixed effects logit model, transformation and GMM estimator

In this section, the fixed effects logit model is implicitly defined, where the error term is of additive form.[2] The hyperbolic transformation, which eliminates the fixed effects and then based on which the moment conditions is constructed for estimating the model consistently, is the product of the

* Discussion Paper Series, Faculty of Economics, Kyushu Sangyo University Discussion Paper, June 2010, No. 43
Errata information: http://www.ip.kyusan-u.ac.jp/J/kitazawa/ERRATA/errata_htd.html
† Correspondence: Faculty of Economics, Kyushu Sangyo University, Matsukadai 2-3-1, Higashiku, Fukuoka, Japan.
E-mail: kitazawa@ip.kyusan-u.ac.jp
1 Additionally, Honoré and Kyriazidou (2000) propose an estimator for the fixed effects logit model with the lagged dependent variable. As for details, see also p211-216 in Hsiao (2003).
2 The regression form defined implicitly is also used by Blundell et al. (2002) for count panel data .

model defined implicitly. The GMM estimator is defined by using the moment conditions constructed. Through the paper, the subscripts i and t denotes the individual and time period respectively, while N and T are number of individuals and number of time periods respectively. Since the short panel is supposed, it is assumed that $N \to \infty$ and T is fixed. In addition, it is assumed that the variables in the model are independent among individuals.

The fixed effects logit model is able to be written in the implicit form as follows:

$$y_{it} = p_{it} + v_{it} \quad , \qquad \text{for} \quad t = 1, \dots, T \quad , \tag{2.1}$$

$$p_{it} = \exp(\psi_i + \delta w_{it}) / (1 + \exp(\psi_i + \delta w_{it})) \quad , \qquad \text{for} \quad t = 1, \dots, T \quad , \tag{2.2}$$

where the observable variables y_{it} and w_{it} are the binary dependent variable and the real-valued explanatory variable respectively, while the unobservable variables Ψ_i and v_{it} are the individual fixed effect and the disturbance respectively.[3] Equations (2.1) say that y_{it} take one with probability p_{it} , while it is seen from equations (2.2) that the probability is the logistic cumulative distribution function of $\psi_i + \delta w_{it}$. Allowing for the serially uncorrelated disturbances, the uncorrelatedness between the disturbances and the fixed effect and the strictly exogenous explanatory variables, the assumptions on the disturbances are specified as

$$E[v_{it} \mid v_i^{t-1}, \psi_i, w_i^T] = 0 \quad , \qquad \text{for} \quad t = 1, \dots, T \quad , \tag{2.3}$$

where $v_i^{t-1} = (v_{i1}, \dots, v_{i,t-1})$ for $t = 2, \dots, T$, v_i^0 is defined as the empty set for convenience and $w_i^T = (w_{i1}, \dots, w_{iT})$. The assumptions (2.3) can be derived from the assumption underlying the fixed effects logit model, which is that y_{it} for $t = 1, \dots, T$ are mutually independent conditional on Ψ_i and w_i^T .[4]

From now on, based on the fixed effects logit model composed of (2.1) and (2.2) with (2.3), the moment conditions for estimating δ consistently are constructed by using a hyperbolic transformation, as stated below. Taking notice of the fact that

$$\tanh((\psi_i + \delta w_{it})/2) = 2 p_{it} - 1 \tag{2.4}$$

and using the formula that

$$\tanh(a - b) = (\tanh(a) - \tanh(b)) / (1 - \tanh(a) \tanh(b)) \tag{2.5}$$

with a and b being any real numbers, it follows that

3 It is generally assumed that the individual effect Ψ_i is correlated with the explanatory variables w_{it} for each i .

4 If the underlying assumption holds, $f(y_{it} | y_i^{t-1}, \psi_i, x_i^T) = f(y_{it} | \psi_i, x_i^T) = p_{it}$, where $f(\cdot|\cdot)$ is the conditional probability density function. Accordingly, $E[y_{it} | y_i^{t-1}, \psi_i, x_i^T] = E[y_{it} | \psi_i, x_i^T] = p_{it}$. As for details, see p23 in Cameron and Trivedi (2005). Taking notice of (2.1) and the fact that $v_{it} = y_{it} - p_{it}$, the assumptions (2.3) are obtained.

$$\tanh(\delta\,\Delta w_{it}/2)=(p_{it}-p_{i,t-1})/(p_{it}+p_{i,t-1}-2\,p_{it}\,p_{i,t-1})\quad, \tag{2.6}$$

where Δ is the first differencing operator, such as $\Delta w_{it}=w_{it}-w_{i,t-1}$. Since p_{it} and $p_{it}\,p_{i,t-1}$ are written as

$$p_{it}=E[y_{it}\,|\,v_i^{t-1},\psi_i,w_i^T] \tag{2.7}$$

and

$$p_{it}\,p_{i,t-1}=E[y_{it}y_{i,t-1}\,|\,v_i^{t-1},\psi_i,w_i^T]-p_{it}v_{i,t-1} \tag{2.8}$$

respectively by using (2.1) and (2.3), plugging (2.7) and (2.8) into (2.6) gives

$$\begin{gathered}(\,E[y_{it}\,|\,v_i^{t-1},\psi_i,w_i^T]+E[y_{i,t-1}\,|\,v_i^{t-2},\psi_i,w_i^T]-2(E[y_{it}y_{i,t-1}\,|\,v_i^{t-1},\psi_i,w_i^T]-p_{it}v_{i,t-1}))\\ \times\tanh(\delta\,\Delta w_{it}/2)\;=\;E[y_{it}\,|\,v_i^{t-1},\psi_i,w_i^T]-E[y_{i,t-1}\,|\,v_i^{t-2},\psi_i,w_i^T]\;.\end{gathered} \tag{2.9}$$

Equations (2.7) and (2.8) are obtained by plugging (2.1) into $E[y_{it}\,|\,v_i^{t-1},\psi_i,w_i^T]$ and $E[y_{it}y_{i,t-1}\,|\,v_i^{t-1},\psi_i,w_i^T]$ and then applying (2.3) to them. Taking the expectation conditional on (v_i^{t-2},ψ_i,w_i^T) for both sides of (2.9) and then applying law of iterated expectation and (2.3) dated $t-1$, it follows that

$$E[(\,(y_{it}-y_{i,t-1})\;-\;\tanh(\delta\,\Delta w_{it}/2)(y_{it}+y_{i,t-1}-2\,y_{it}y_{i,t-1})\,)\,|\,v_i^{t-2},\psi_i,w_i^T]\;=\;0\quad. \tag{2.10}$$

Since $(y_{it})^n=y_{it}$ for any positive integer value n due to the property of binary variable, equation (2.10) results in

$$E[h_{it}(\delta)\,|\,v_i^{t-2},\psi_i,w_i^T]\;=\;0\quad,\qquad\qquad \text{for}\quad t=2,\dots,T\quad, \tag{2.11}$$

where

$$h_{it}(\delta)=\Delta y_{it}\;-\;\tanh(\delta\,\Delta w_{it}/2)(\Delta y_{it})^2\quad. \tag{2.12}$$

The transformation (2.12) is referred to as “the hyperbolic tangent differencing transformation” for the fixed effects logit model in this paper and hereafter abbreviated to “the HTD transformation”. It should be noted that as seen from (2.11) and (2.12), observations for which $y_{it}=y_{i,t-1}=0$ and $y_{it}=y_{i,t-1}=1$ make no direct contribution to obtaining the estimates of δ based on the moment conditions (2.11), since $h_{it}(\delta)$ is invariably zero for these observations.

The conditional moment conditions (2.11) give the following $m\times 1$ vector of unconditional moment conditions:

$$E[(z_i)'h_i(\delta)]\;=\;0\quad, \tag{2.13}$$

where $h_i(\delta)=[h_{i2}(\delta)\cdots h_{iT}(\delta)]'$ is the $(T-1)\times 1$ vector and $z_i=\mathrm{diag}[(z_{i2})'\cdots(z_{iT})']$ is the $(T-1)\times m$ matrix with $m=\sum_{t=2}^{T} m_t$. The (transposed) blocks

$$z_{it}=f_t(v_i^{t-2},\psi_i,w_i^T)\ , \qquad \text{for} \quad t=2,\ldots,T \tag{2.14}$$

are the $m_t\times 1$ vector-valued functions of v_i^{t-2} , ψ_i and w_i^T at time t , where m_t is number of instruments for time t . By using the empirical counterpart of (2.13):

$$g_N(\delta)=(1/N)\sum_{i=1}^{N}(z_i)'h_i(\delta) \tag{2.15}$$

and the $m\times m$ inverse of optimal weighting matrix:

$$W_N(\hat{\delta}_1)=(1/N)\sum_{i=1}^{N}(z_i)'h_i(\hat{\delta}_1)\,(h_i(\hat{\delta}_1))'z_i\ , \tag{2.16}$$

where $\hat{\delta}_1$ is any initial consistent estimator for δ , the GMM estimator is constructed as follows:

$$\hat{\delta}_{GMM}=arg\min_{\delta}\,(g_N(\delta))'\,(W_N(\hat{\delta}_1))^{-1}\,g_N(\delta)\ , \tag{2.17}$$

where $N^{1/2}(\hat{\delta}_{GMM}-\delta_0)$ converges in distribution to the normal distribution as follows:

$$N^{1/2}(\hat{\delta}_{GMM}-\delta_0)\ \xrightarrow{d}\ \mathrm{N}(0,\ (\,(D(\delta_0))'\,(W(\delta_0))^{-1}\,D(\delta_0)\,)^{-1}\,) \tag{2.18}$$

with δ_0 being the true value of δ . Taking notice of the assumption that the variables are independent among individuals, $W(\delta_0)$, which is the (asymptotic) variance-covariance matrix of the moment conditions (2.13), can be written by using δ_0 as follows:

$$W(\delta_0)=E[(z_i)'h_i(\delta_0)\,(h_i(\delta_0))'z_i]\ , \tag{2.19}$$

where it should be noted that (2.16) is the empirical counterpart of (2.19) if $\hat{\delta}_1$ is replaced by δ_0 and $N^{(1/2)}g_N(\delta_0)\ \xrightarrow{d}\ \mathrm{N}(0,W(\delta_0))$. Further, the first derivative of (2.13) with respect to δ for δ_0 is as follows:

$$D(\delta_0)=(\partial E[(z_i)'h_i(\delta)]\,/\,\partial\delta)\big|_{\delta=\delta_0}\ . \tag{2.20}$$

It is conceivable that inferences for the GMM estimator based on the HTD transformation could be permitted to be conducted on the basis of numbers of observations for which $(\Delta y_{it})^2=1$ instead of N , on the grounds that observations except for those for which $(\Delta y_{it})^2=1$ make no direct contribution to estimating δ . In this case,

$\bar{M}=(1/(T-1))\sum_{i=2}^{T} M_t$ is expediently used for the inferences instead of N, where M_t is number of observations for which $(\Delta y_{it})^2=1$ at time t.

3. Link between CMLE and GMM estimator

The discussion here is conducted for the case of two periods (i.e. $t-1$ and t). It is shown in this section that the GMM estimator opting for an instrument is identical to the CMLE in this case.

First, the GMM estimator is presented. With $h_i(\delta)=h_{it}(\delta)$ and $z_i=z_{it}=\Delta w_{it}/2$ (both of which are scalar), equation (2.13) turns to

$$E[(\Delta w_{it}/2)\, h_{it}(\delta)] \;=\; 0 \quad . \tag{3.1}$$

The moment condition (3.1) says that $\Delta w_{it}/2$ is used as the instrument for the HTD transformation $h_{it}(\delta)$. The GMM estimator for δ is the just-identified one when using only the moment condition (3.1) for the two periods. This is denoted by $\hat{\delta}^*_{GMM}$ hereafter.

The first derivative with respect to δ and square of $h_{it}(\delta)$ are respectively calculated as follows:

$$\partial h_{it}(\delta)/\partial\delta \;=\; -(\Delta y_{it})^2(\Delta w_{it}/2)\,\mathrm{sech}^2(\delta\Delta w_{it}/2) \tag{3.2}$$

and

$$\begin{aligned}(h_{it}(\delta))^2 &=(\Delta y_{it})^2 - 2\tanh(\delta\Delta w_{it}/2)(\Delta y_{it}) + \tanh^2(\delta\Delta w_{it}/2)(\Delta y_{it})^2 \\ &=(\Delta y_{it})^2\,\mathrm{sech}^2(\delta\Delta w_{it}/2) \;-\; 2\tanh(\delta\Delta w_{it}/2)\, h_{it}(\delta)\end{aligned} \quad , \tag{3.3}$$

where the relationship that $(\Delta y_{it})^n=(\Delta y_{it})^2$ if n is even and $(\Delta y_{it})^n=\Delta y_{it}$ if n is odd is used since y_{it} is binary. Using (2.19), (2.20), (3.2) and (3.3), $W(\delta_0)$ and $D(\delta_0)$ for (3.1) are respectively calculated as follows:

$$\begin{aligned}W^*(\delta_0)&=E[(h_{it}(\delta_0))^2(\Delta w_{it}/2)^2] \\ &=(1/4)E[(\Delta y_{it})^2(\Delta w_{it})^2\mathrm{sech}^2(\delta_0\Delta w_{it}/2)]\end{aligned} \quad , \tag{3.4}$$

where the relationship $E[\tanh(\delta\Delta w_{it}/2)\, h_{it}(\delta_0)\,(\Delta w_{it})^2]=0$ obtained from (2.11) is used and

$$\begin{aligned}D^*(\delta_0)&=E[(\Delta w_{it}/2)\,(\partial h_{it}(\delta)/\partial\delta\,|_{\delta=\delta_0})] \\ &=-(1/4)\,E[(\Delta y_{it})^2(\Delta w_{it})^2\mathrm{sech}^2(\delta_0\Delta w_{it}/2)]\end{aligned} \quad . \tag{3.5}$$

Looking at (3.4) and (3.5), it can be seen that

$$W^*(\delta_0)=-D^*(\delta_0) \quad . \tag{3.6}$$

In addition, the relationship (2.18) is also applicable to the just-identified estimator (see p486-487 in Hayashi, 2000). Therefore, it follows from (2.18) and (3.6) that the following relationship holds for $\hat{\delta}^*_{GMM}$:

$$N^{1/2}(\hat{\delta}^*_{GMM}-\delta_0) \xrightarrow{d} \mathrm{N}(0,\ -(1/D^*(\delta_0)))\ . \tag{3.7}$$

Lee (2002, p84-87) elucidates the equality conceptually identical to (3.6) in the context of the CMLE to be hereafter described.

Next, the conventional CMLE proposed by Chamberlain (1980) is presented for the two periods as follows:

$$\hat{\delta}^*_{CML}=arg\max_{\delta} L(\delta)\ , \tag{3.8}$$

where $L(\delta)=\sum_{i=1}^{N} l_{it}(\delta)$. Referring to Wooldridge (2002, p490-492), the logarithm of probability composing the conditional log-likelihood function for the two-periods fixed effects logit model is written as follows, with $\pi_{it}(\delta)=\exp(\delta\Delta w_{it})/(1+\exp(\delta\Delta w_{it}))$:

$$l_{it}(\delta)=\vartheta_{it}(\varsigma_{it}\ln(\pi_{it}(\delta))+(1-\varsigma_{it})\ln(1-\pi_{it}(\delta)))\ , \tag{3.9}$$

where $\vartheta_{it}=1$ if $y_{i,t-1}+y_{it}=1$ and $\vartheta_{it}=0$ otherwise, while $\varsigma_{it}=1$ if $y_{i,t-1}=0$ and $y_{it}=1$ and $\varsigma_{it}=0$ if $y_{i,t-1}=1$ and $y_{it}=0$. In (3.9), $\pi_{it}(\delta)$ stands for the probability with which y_{it} takes one given $w_{i,t-1}$, w_{it} , Ψ_i and $y_{i,t-1}+y_{it}=1$, while $1-\pi_{it}(\delta)$ stands for the probability with which y_{it} takes zero given $w_{i,t-1}$, w_{it} , Ψ_i and $y_{i,t-1}+y_{it}=1$.

The first-order condition of $L(\delta)$ is

$$\partial L(\delta)/\partial\delta=\sum_{i=1}^{N}\partial l_{it}(\delta)/\partial\delta=0 \tag{3.10}$$

with

$$\partial l_{it}(\delta)/\partial\delta=\vartheta_{it}\Delta w_{it}(\varsigma_{it}(1-\pi_{it}(\delta))-(1-\varsigma_{it})\pi_{it}(\delta))\ . \tag{3.11}$$

It is corroborated from (3.10) with (3.11) that the first-order condition of $L(\delta)$ divided by N is the empirical counterpart of the moment condition (3.1) for the GMM estimator. The second-order derivative of $L(\delta)$ with respect to δ is written as

$$\partial^2 L(\delta)/\partial\delta^2=\sum_{i=1}^{N}-\vartheta_{it}(\Delta w_{it})^2\pi_{it}(\delta)(1-\pi_{it}(\delta))\ . \tag{3.12}$$

Taking notice of the fact that $\mathrm{sech}^2(\delta\Delta w_{it}/2)=4\pi_{it}(\delta)(1-\pi(\delta))$, it is evident that if δ is replaced by δ_0 , (3.12) divided by N is the empirical counterpart of (3.5) and accordingly identical to $-W^*(\delta_0)$ from (3.6). Therefore, the following relationship holds for $\hat{\delta}^*_{CML}$:

$$N^{1/2}(\hat{\delta}^*_{CML}-\delta_0) \xrightarrow{d} \mathrm{N}(0,\ -(1/D^*(\delta_0)))\ . \tag{3.13}$$

Judging from the above, it is ascertained that for the two periods the conventional CMLE for the fixed effects logit model is identical to the GMM estimator selecting $\Delta w_{it}/2$ as the instrument for the HTD transformation.

To make doubly sure, the integration of $(\Delta w_{it}/2)\, h_{it}(\delta)$ with respect to δ is conducted:

$$\int (\Delta w_{it}/2) h_{it}(\delta) d\delta \;=\; \delta(\Delta w_{it}/2)\Delta y_{it} - (\Delta y_{it})^2 \ln(\cosh(\delta \Delta w_{it}/2)) + C \quad , \tag{3.14}$$

where C is the constant of integration. With $C=-(\Delta y_{it})^2 \ln(2)$ for (3.14), the logarithm of probability (3.9) (which composes the conditional log-likelihood function for the two-periods fixed effects logit model) is compactly rewritten as

$$l_{it}(\delta) \;=\; \delta(\Delta w_{it}/2)\Delta y_{it} - (\Delta y_{it})^2 \ln(2\cosh(\delta \Delta w_{it}/2)) \quad . \tag{3.15}$$

The exponential of $l_{it}(\delta)$ in (3.15) (which is equivalent to (3.9)) represents the probability density when the restriction $(\Delta y_{it})^2=1$ is imposed. In this case, number of observations for which $(\Delta y_{it})^2=1$ is used instead of N in this section and therefore $\hat{\delta}^*_{CML}$ (which is equivalent to $\hat{\delta}^*_{GMM}$) could be interpreted as being the asymptotically efficient estimator. This is because the Cramér-Rao inequality is applicable in this case.

Incidentally, Abrevaya (1997) shows that for the fixed effects logit model a scale-adjusted ordinary maximum likelihood estimator is equivalent to the CMLE for the case of two periods.

4. Monte Carlo

In this section, some Monte Carlo experiments are conducted to investigate the small sample performance of the GMM estimator for the fixed effects logit model described in section 2. The experiments are implemented by using an econometric software TSP version 4.5 (see Hall and Cummins, 2006).

The data generating process (DGP) is as follows:

$$y_{it}=\begin{cases} 1 & \text{if } p_{it}>u_{it} \\ 0 & \text{otherwise} \end{cases} \quad ,$$

$$p_{it}=\exp(\psi_i+\delta w_{it})/(1+\exp(\psi_i+\delta w_{it})) \quad ,$$

$$u_{it}\sim \mathrm{U}(0,1) \quad ,$$

$$w_{it}=\alpha w_{i,t-1}+\iota\psi_i+\zeta_{it} \quad ,$$

$$w_{i1}=(1/(1-\alpha))\iota\psi_i+(1/(1-\alpha^2)^{(1/2)})\zeta_{i1} \quad ,$$

$$\psi_i\sim \mathrm{N}(0,\sigma^2_\psi) \quad ; \quad \zeta_{it}\sim \mathrm{N}(0,\sigma^2_\zeta) \quad .$$

In the DGP, values are set to the parameters δ , α , ι , σ^2_ψ and σ^2_ζ . The experiments are carried out with the cross-sectional sizes $N=100$, 500 and 1000, the numbers of time periods $T=4$ and 8 and the number of replications $TR=1000$.

In the experiments, the GMM estimator based on the HTD transformation selects Δw_{it} as the instruments for the transformation $h_{it}(\delta)$. That is, the GMM(HTD) estimator uses the vector of moment conditions (2.13) with $z_{it}=\Delta w_{it}$, which is able to be written piecewise as follows:

$$E[\Delta w_{it}\, h_{it}(\delta)] = 0 \ , \qquad \text{for} \quad t=2,\dots,T \ . \tag{4.1}$$

As a control, another GMM estimator is used, which employs the following moment conditions disregarding the unobservable heterogeneity:

$$E[w_{it}\, \varpi_{it}(\delta)] = 0 \ , \qquad \text{for} \quad t=1,\dots,T \ , \tag{4.2}$$

where $\varpi_{it}(\delta)=\exp(\delta w_{it})/(1+\exp(\delta w_{it}))$. The GMM(LgtLev) estimator (i.e. the level GMM estimator for the logit model) for δ is inconsistent due to the ignorance of the fixed effects.

The Monte Carlo results are exhibited in Table 1. The settings of values of the parameters for the explanatory variables w_{it} are the same as those used by Blundell et al. (2002) for count panel data model. The small sample property of the GMM (HTD) estimator can be said to be preferable and their bias and rmse (root mean squared error) decrease as the cross-sectional size N increases, which is the reflection of the consistency. In contrast, the sizable downward bias and rmse for the (inconsistent) GMM(LgtLev) estimator remain at virtually constant levels when N increases. As is seen from comparison between Simulations (a) and (b) for the GMM(HTD) estimator, the small sample performance of the GMM(HTD) estimator is better off for $T=8$ than for $T=4$, reflecting the substantive increase of sample size. Further, the results of Simulations (b), (c) and (d) for the GMM(HTD) estimator raise the possibility that more persistent series of the explanatory variables might bring about more deteriorated small sample performance of the GMM(HTD) estimator.[5]

5. Conclusion

This paper proposed the hyperbolic tangent differencing (HTD) transformation for the fixed effects logit model, with the intention of ruling out the fixed effects. The consistent GMM estimator was constructed by using the HTD transformation. Then, the equivalence of the GMM estimator opting for an instrument and the CMLE proposed by Chamberlain (1980) was revealed for the case of two periods. In addition, the Monte Carlo experiments indicated the desirable small sample property of the GMM estimator based on the HTD transformation.

References

Abrevaya, J. (1997). The equivalence of two estimators of the fixed effects logit model. *Economics Letters* **55,** 41-43.

Blundell, R., Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* **87**, 115-143.

Blundell, R., Griffith, R., Windmeijer, F. (2002). Individual effects and dynamics in count data models. *Journal of Econometrics* **108**, 113-131.

Cameron, A.C., Trivedi, P.K. (2005). *Microeconometrics: methods and applications*, Cambridge university press.

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

5 This possibility is also pointed out in the framework of ordinary and count panel data models. For example, see Blundell and Bond (1998) and Blundell et al. (2002).

47, 225-238.
Hall, B.H., Cummins, C. (2006). *TSP 5.0 user's guide*, TSP international.
Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica* **50**, 1029-1054.
Hayashi, F. (2000). *Econometrics*, Princeton university press.
Honoré B.E., Kyriazidou, E. (2000). Panel data discrete choice models with lagged dependent variables. *Econometrica* **68**, 839-874.
Hsiao, C. (2003). *Analysis of panel data*, *2nd Edition*, Cambridge university press.
Lee, M.J. (2002). *Panel data econometrics*, Academic press.
Wooldridge, J.M. (2002). *Econometric analysis of cross-section and panel data*, MIT press.

Table 1. Monte Carlo results for the fixed effects logit model

		N=100		N=500		N=1000	
		bias	rmse	bias	rmse	bias	rmse
Simulation (a):	T=4						
GMM(HTD)	δ	0.08	0.29	0.02	0.11	0.01	0.08
GMM(LgtLev)	δ	−0.50	0.53	−0.50	0.50	−0.50	0.50
Simulation (b):	T=8						
GMM(HTD)	δ	0.06	0.19	0.02	0.08	0.01	0.05
GMM(LgtLev)	δ	−0.50	0.52	−0.50	0.51	−0.50	0.50
Simulation (c):	T=8						
GMM(HTD)	δ	0.10	0.58	0.03	0.25	0.02	0.17
GMM(LgtLev)	δ	−1.00	1.08	−1.01	1.02	−1.01	1.01
Simulation (d):	T=8						
GMM(HTD)	δ	0.09	1.04	0.02	0.43	0.02	0.31
GMM(LgtLev)	δ	−1.00	1.16	−1.01	1.03	−1.01	1.02

Notes:
The parameter settongs in the DGP are as follows:
Simulations (a) and (b): $\delta=0.5$; $\alpha=0.5$; $\iota=0.1$; $\sigma_{\psi}^{2}=0.5$; $\sigma_{\zeta}^{2}=0.5$.
Simulation (c): $\delta=1$; $\alpha=0.9$; $\iota=0$; $\sigma_{\psi}^{2}=0.5$; $\sigma_{\zeta}^{2}=0.05$.
Simulation (d): $\delta=1$; $\alpha=0.95$; $\iota=0$; $\sigma_{\psi}^{2}=0.5$; $\sigma_{\zeta}^{2}=0.015$.
No non-convergence is found in all replications.
In each GMM estimation, the initial consistent estimate is obtained by using the inverse of cross-sectional average of the products between the instruments matrix as the non-optimal weighting matrix.
The values of the Monte Carlo statistics are obtained using the true values of δ as the starting values in the optimization for each replication. The values of the statistics obtained using the true values are not much different from those obtained using two different types of the starting values.