

Discussion Paper No. 02-64

**Almost Consistent Estimation
of Panel Probit Models
with ‘Small’ Fixed Effects**

François Laisney and Michael Lechner

ZEW

Zentrum für Europäische
Wirtschaftsforschung GmbH

Centre for European
Economic Research

Discussion Paper No. 02-64

**Almost Consistent Estimation
of Panel Probit Models
with ‘Small’ Fixed Effects**

François Laisney and Michael Lechner

Download this ZEW Discussion Paper from our ftp server:

<ftp://ftp.zew.de/pub/zew-docs/dp/dp0264.pdf>

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von
neueren Forschungsarbeiten des ZEW. Die Beiträge liegen in alleiniger Verantwortung
der Autoren und stellen nicht notwendigerweise die Meinung des ZEW dar.

Discussion Papers are intended to make results of ZEW research promptly available to other
economists in order to encourage discussion and suggestions for revisions. The authors are solely
responsible for the contents which do not necessarily represent the opinion of the ZEW.

Non Technical Summary

The panel probit model with individual effects plays an important role in applied econometrics despite its shortcomings, the most important being the inconsistency of the maximum likelihood estimator for the panel probit model with fixed individual effects when the number of time periods, T , is small, and only the number of individuals, N , is large. Several alternative models and estimators have been proposed to tackle this inconsistency problem due to the lack of independence between individual effects and regressors, but all present some drawbacks. For instance, the conditional logit model requires strict exogeneity of the regressors, and stationarity over time. Semi-parametric estimators require strong assumptions concerning the regressors and are difficult to implement with the relatively large numbers of regressors typically used in empirical applications.

Here we propose estimators for the panel probit model with fixed individual effects which are based on the conditional moment estimation framework. These estimators are asymptotically normal and converge to their limit at the speed \sqrt{N} . They achieve asymptotic efficiency in a well-defined sense, and they are fairly easy to implement using standard software. The moments used are derived from a first order approximation of the mean of the dependent variable conditional on explanatory variables and on the fixed effect, $E(y_{it} | x_i, c_i)$, around a value \tilde{c} . We consider two different choices for the instruments and also introduce trimmed estimators designed to reduce the bias. In a Monte Carlo study we compare the performance of these estimators with those of the pooled probit estimator and of the conditional logit estimator for nine different data generation processes. The results are encouraging.

Almost Consistent Estimation of Panel Probit Models with ‘Small’ Fixed Effects

François Laisney⁺ and Michael Lechner^{++ *}

⁺*BETA-THEME, Université Louis Pasteur, Strasbourg, and Zentrum für Europäische
Wirtschaftsforschung (ZEW), Mannheim*

⁺⁺*Swiss Institute for International Economics and Applied Economic Research (SIAW),
Universität St. Gallen*

June 2002

Abstract

We propose four different GMM estimators that allow almost consistent estimation of the structural parameters of panel probit models with fixed effects for the case of small T and large N . The moments used are derived for each period from a first order approximation of the mean of the dependent variable conditional on explanatory variables and on the fixed effect. The estimators differ w.r.t. the choice of instruments and whether they use trimming to reduce the bias or not. In a Monte Carlo study we compare these estimators with pooled probit and conditional logit estimators for different DGPs. The results show that the proposed estimators outperform these competitors in several situations.

Keywords: Panel data, binary choice model, generalised method of moments, fixed effects

JEL classification: C23, C35, C51

Addresses for correspondence:

François Laisney, ZEW, Postfach 103443, DE-68034 Mannheim, Germany, laisney@zew.de

Michael Lechner, SIAW, Universität St. Gallen, Dufourstr. 48, CH-9000 St. Gallen, Switzerland,
Michael.Lechner@unisg.ch, www.unisg.ch/lechner

* The first author is also affiliated with IFS, London, and the second author with CEPR, London, IZA, Bonn, and ZEW, Mannheim. We thank two anonymous referees for insightful comments, as well as Martin Falk, Darren Grant, Bo Honoré, Joachim Inkmann, Chen Songnian, Arthur Van Soest, Tom Wansbeek and Tiemen Woutersen, as well as audiences in ESEM'99, in the Panel Data Conference in Geneva, 2000, and at seminars in St. Gallen, Strasbourg, and at ZEW, Mannheim. The second author gratefully acknowledges financial support from the Deutsche Forschungsgemeinschaft (DFG, grant 928/2) and the Swiss National Science Foundation (SNF, grants 1214-53735, 4043-058311 and 4045-050673).

1 Introduction

The panel probit model with individual specific effects plays an important role in applied econometrics, despite its shortcomings. The first obvious shortcoming lies in the arbitrariness of the normality and homoscedasticity assumptions, the second in the inconsistency of the maximum likelihood estimator for the panel probit model with fixed individual effects when the number of time periods, T , is small, and only the number of individuals, N , is large.¹ Several estimators have been proposed to tackle this inconsistency problem due to the lack of independence between individual specific error terms and regressors. The estimator proposed by Chamberlain (1984) to circumvent this problem appears to require a very large N in order to yield satisfactory results (see Lechner and Breitung, 1996, for details). It also heavily relies on correct specification of the distribution of the individual effects given the regressors.

A popular alternative to the panel probit model with fixed effects is the conditional logit model (see Rasch, 1960, Andersen, 1970, and Chamberlain, 1980, and Oswald, 1998, for a recent application and justification of this model choice). A drawback of the latter is that it requires strict exogeneity of the regressors, and stationarity over time: for instance, it cannot, at least in principle, accommodate heteroscedasticity over time in the latent model.²

Recently, less restrictive models and corresponding estimators have been proposed, but they all present some drawbacks for the practitioner. The semiparametric estimator proposed by Manski (1987) is difficult to implement and its speed of convergence is only $\sqrt[3]{N}$; this can be improved somewhat by smoothing, as proposed by Horowitz (1992), Kyriazidou (1995), and Charlier, Melenberg and van Soest (1995), but Chamberlain (1992) shows the impossibility of attaining \sqrt{N} consistency in the framework adopted by all these papers. More recently, Chen (1998) and Lee (1999) have found sets of assumptions that permit \sqrt{N} : Chen uses a restriction on the nature of the dependence of the individual-specific error term on the regressors, and Lee introduces a restriction on the time path of the regressors. While undoubtedly appealing, these estimators require quite stringent regularity assumptions w.r.t.

¹ For the case of truly random effects (i.e., independent of the regressors), different estimators have been proposed: see Avery et al. (1983), Butler and Moffitt (1982), Bertschek and Lechner (1998), and Chib (1996).

² For completeness, we should also mention the promising Bayesian approach to the fixed effects panel probit model initiated by Lancaster (1999, 2000, 2001), and the conditional probit approach of Magnac (2002).

the regressors and become difficult or even impossible to implement with the relatively large numbers of regressors typically used in empirical applications.³

Here we propose estimators for the panel probit model with fixed individual effects which are based on the conditional moment estimation framework. Whether or not their limit is the true value of the parameter, these estimators are asymptotically normal and converge to their limit at the speed \sqrt{N} . They achieve asymptotic efficiency in a well-defined sense, and they are fairly easy to implement using standard GMM software. The moments used are derived from a first order approximation of the mean of the dependent variable conditional on explanatory variables and on the fixed effect, $E(y_{it} | x_{it}, c_i)$, around a value \tilde{c} . We consider two different choices for the instruments and also introduce trimmed estimators designed to reduce the bias. In a Monte Carlo study we compare the performance of these estimators with those of the pooled probit estimator and of the conditional logit estimator for nine different data generation processes. The results are encouraging.

Section 2 presents the notation and the panel probit model, Section 3 presents the GMM framework and the proposed estimators, Section 4 discusses the results of a Monte Carlo study and Section 5 concludes. Appendices A, B and C concern the treatment of heteroscedasticity over time, the choice of the expansion point for the Taylor approximation, and the precise definition of the trimmed estimators, respectively.

2 The panel probit model

In this section we introduce notation and describe the model specification and the moment conditions we will use as a starting point.

2.1 The fixed effects probit model

Equation (1) states the typical binary choice panel data model:

$$y_{it} = \mathbb{I}[(x_{it}\mathbf{b}_0 + c_i + u_{it}) \geq 0]; \quad t = 1, \dots, T; \quad i = 1, \dots, N. \quad (1)$$

y_{it} denotes the zero/one observable outcome of unit i at time t . The K -dimensional vector x_{it} denote the characteristics of unit i at time t that are observable to the econometrician, whereas c_i and u_{it} denote unobserved characteristics. \mathbf{b} is a deterministic coefficient vector. Its true

³ See also Abrevaya (2000) and different approaches to estimation with predetermined regressors: Arellano and Carrasco (2000), Honore and Kyriazidou (2000), Honore and Lewbel (2002), as well as the survey of

value is denoted by \mathbf{b}_0 . Finally, $\mathbb{1}(\cdot)$ denotes the indicator function that is one if its argument is true, and zero otherwise. It is useful to collect the data over time in the following way: $z_i = (y_i, x_i)$, $y_i = (y_{1i}, \dots, y_{Ti})$, $x_i = (x_{1i}, \dots, x_{Ti})$. The observations z_i are assumed to be independent draws from some large population described by the random variables $Z = (Y, X)$, $Y = (Y_1, \dots, Y_T)$, $X = (X_1, \dots, X_T)$, C and $U = (U_1, \dots, U_T)$. The realisations of C and U are not observable. In the following, small letters without the symbol i denote fixed values of the corresponding random variables. With respect to the distribution of the error terms conditional on the explanatory variables and the fixed effect, we assume a normal distribution with mean zero and variance matrix Σ :

$$(U|C = c, X = x) \sim N(0, \Sigma); \quad \forall c, x. \quad (2)$$

Equation (2) implies that the regressors are strictly exogenous, and we come back to that assumption later. Note that random effects are taken care of through Σ , which is unrestricted and could thus present equicorrelation, for instance. Due to the standard identification problem in binary choice models, we normalise one element of the main diagonal of Σ to be unity. To ease notation, we will proceed as if all elements of the main diagonal are unity.⁴

The distribution of the other error term C conditional on the explanatory variables (X) is left unrestricted. Therefore, it could potentially be correlated with regressors, and thus it must be treated like a *fixed effect*.

2.2 Moment conditions

Equations (1) and (2) allow to derive the expectations of the observable dependent variables Y conditional on regressors X and the fixed effects C :

$$E(Y_t|X_t = x_t, C = c) = E(Y_t|X = x, C = c) = \Phi(x_t \mathbf{b}_0 + c); \quad t = 1, \dots, T. \quad (3)$$

These moment conditions have the advantage that they are marginal in the sense that their expression depends only on the univariate cdf of the normal distribution, denoted by Φ . The latter can be easily and accurately evaluated numerically. Note that the panel probit model implies several other moment conditions, for example for cross-products of the dependent variables. However, such moment conditions are more difficult to evaluate because they

Arellano and Honore (2002).

⁴ The adjustments necessitated by the estimation of the variances are described in Appendix A. Although allowing for heteroscedasticity over time sounded promising, as several of the contenders (in particular the conditional logit) cannot accommodate that feature, simulation results (available on request) suggest that this is not much of a problem.

depend on cdf's of the multivariate normal distribution. Therefore, the latter will be ignored in the following. Incidentally, Bertschek and Lechner (1998) showed in a Monte Carlo study that for the random effects panel probit model GMM estimators using optimally weighted marginal moment conditions (similar to those of equation (3)), are almost as efficient as maximum likelihood estimators.

Still, at this stage it is important to note that, as pointed out by Wooldridge (2000), estimation of \mathbf{b} and Σ will only give limited information on the conditional distribution of y given x . Since the derivative of $P(y_t = 1 | x_t, c)$ w.r.t. a continuous regressor x_t^k depends on c through the density $\mathbf{j}(x_t, \mathbf{b} + c)$, we will only be able to infer the direction of the impact of the k -th regressor on the probability of the outcome, but not its magnitude. But the quotients of the marginal effects of continuous regressors, say x_t^k and x_t^j will be fully identified and coincide with the quotient of the corresponding coefficients.

3 GMM estimation

We now discuss alternative approaches to solving the problem posed by this unobservable individual effect C and consider in turn moment conditions leading to inconsistent estimators, our proposal for moment conditions which will be approximately satisfied provided that the individual effects are small, and estimators based on such moments.

3.1 *Moment conditions leading to inconsistent estimators*

The moment conditions obtained in equation (3) cannot readily be used for estimation because the individual effects c_i are unobserved. There are two standard approaches in the literature to circumvent this problem: The first way to deal with C is to include unit-specific dummy-variables in the model, i.e. to estimate c_i together with the parameters of interest \mathbf{b} . Heckman (1981) showed that for T fixed such an estimator leads to inconsistent estimates for c_i and \mathbf{b} .⁵ An alternative is to assume a random effects model, i.e. to assume $(U + C | X = x) \sim N(0, \tilde{\Sigma})$ for all x . Using the same normalisation for $\tilde{\Sigma}$ as before, the moment conditions $E(Y_t | X = x) = \Phi(x_t, \mathbf{b}_0)$ that do not depend on the individual effects can be used for consistent estimation of \mathbf{b} (see Avery, Hansen, Hotz, 1983, or Bertschek, Lechner, 1998). However, under the fixed effects specification such moment conditions are generally not valid, and

⁵ All asymptotic arguments here are for the case of T fixed and N increasing.

hence estimators based on them are inconsistent. The problem in this respect is the potential correlation of the individual effects with the regressors.

3.2 Alternative moment conditions

A ‘solution’ to the problems described above is to find moment conditions that allow for the existence of fixed effects, but do not depend on them. For example in the linear model this approach leads to the well-known within and difference estimators. However, using some sort of differencing in non-linear models is only possible in some special situations, like the Poisson count data model, or the conditional logit model.⁶ Unfortunately, for the probit model such transformations appear not to be available (although Magnac, 2002, shows how to estimate a two-period conditional probit model, and how this can be used in the general case of T periods). Our approach uses a similar idea but is based on a Taylor expansion of the moment condition (3).

Equation (4) gives a linearized version of the moment conditions presented in equation (3). This linearization is obtained by a first order Taylor expansion around $c = \tilde{c}$:

$$\Phi(x_t, \mathbf{b} + c) \approx \Phi(x_t, \mathbf{b} + \tilde{c}) + (c - \tilde{c})\mathbf{f}(x_t, \mathbf{b} + \tilde{c}); \quad t = 1, \dots, T. \quad (4)$$

When the individual effect is close enough to the value of \tilde{c} so that the first order Taylor approximation in (4) is exact, we can write:⁷

$$C - \tilde{c} = \frac{E[Y_t | X, C] - \Phi(X_t, \mathbf{b} + \tilde{c})}{\mathbf{f}(X_t, \mathbf{b} + \tilde{c})} = \frac{E[Y_s | X, C] - \Phi(X_s, \mathbf{b} + \tilde{c})}{\mathbf{f}(X_s, \mathbf{b} + \tilde{c})}, \quad s, t = 1, \dots, T.$$

Thus, the following function,

$$m_{ts}(Z; \mathbf{b}) = \frac{Y_t - \Phi(X_t, \mathbf{b} + \tilde{c})}{\mathbf{f}(X_t, \mathbf{b} + \tilde{c})} - \frac{Y_s - \Phi(X_s, \mathbf{b} + \tilde{c})}{\mathbf{f}(X_s, \mathbf{b} + \tilde{c})}, \quad s, t = 1, \dots, T; \quad s \neq t, \quad (5)$$

has a conditional mean of zero at the true value of \mathbf{b} , given $X = x$:

$$E[m_{ts}(Z; \mathbf{b}_o) | X = x] = E\{E[m_{ts}(Z; \mathbf{b}_o) | X = x, C = c] | X = x\} = 0. \quad (6)$$

Hence, $m_{ts}(Z; \mathbf{b})$ can be used as the basis for (almost) consistent estimation of the panel probit model with fixed effects close to \tilde{c} .⁸ Under standard regularity conditions, a GMM estimator

⁶ And in the latter case, homoscedasticity over time in the latent model is assumed.

⁷ We should use the symbol \approx instead of $=$ to be precise, but for notational convenience we will keep using the $=$ sign.

of the coefficients for the time varying regressors of the panel probit model based on the moment functions given in equation (6) is consistent (almost, given the Taylor approximation) and \sqrt{N} asymptotically normal (see Newey, 1993, or Newey, McFadden, 1994).

In order to alleviate notation in the sequel, let us define $\tilde{\Phi}_t \equiv \Phi(X_t \mathbf{b} + \tilde{c})$, and $\tilde{\mathbf{f}}_t \equiv \mathbf{f}(X_t \mathbf{b} + \tilde{c})$.

Remark 1: A referee suggested that, rather than focusing on almost consistent estimation of the fixed effects probit model, one could instead focus on consistent estimation of a model which is *almost a fixed effects probit model*, i.e., treat the right hand side of (4) as the model. A drawback is of course that the latter expression does not necessarily lie between 0 and 1, but it would be interesting to see if (4) could be modified so as to be interpreted as the true model, while retaining the possibility of eliminating the fixed effect.

Remark 2: *Time-varying* (or period-specific) *coefficients* may easily be introduced in this framework, the only difference with time-constant coefficients lying in the construction of the instruments described in Subsection 3.3. The case of period-specific variances is fully described in Appendix A.

Remark 3: We could also use T moment functions that are in the spirit of *within estimation*,

$$\text{like } m_{ts}^{wt}(Z; \mathbf{b}) = \frac{Y_t - \tilde{\Phi}_t}{\tilde{\mathbf{f}}_t} - \frac{1}{T} \sum_{s=1}^T \frac{Y_s - \tilde{\Phi}_s}{\tilde{\mathbf{f}}_s}. \text{ However, we can expect that the } T(T-1)/2$$

moment functions based on all non-identical first differences are more informative, although it must be recognised that these moment conditions are not independent – a feature shared by the moments used in within estimation – and that in fact the $T-1$ moments $m_{12}(Z; \mathbf{b}), \dots, m_{T-1,T}(Z; \mathbf{b})$ form a basis.

Remark 4: To obtain some insights about *identification*, especially concerning the coefficients of time invariant regressors, which are usually not identified in fixed effects models, it is useful to consider the first derivative of the functions $m_{ts}(Z; \mathbf{b})$ w.r.t. \mathbf{b} :

⁸ Note that the path followed here is very different from the ideas of Newey (1994), who specifies a panel probit model where the conditional distribution of the sum of the individual effect and the idiosyncratic error term given the regressors is assumed normal, with a mean which is an unspecified function of the regressors. By contrast, we assume normality for the idiosyncratic error, and independence from the regressors, but make no assumption on the conditional distribution of the individual effects given the regressors. A similar approach, developed independently in a maximum likelihood framework, is presented by Grant (2000).

$$\frac{\partial m_{ts}(Z; \mathbf{b})}{\partial \mathbf{b}} = \frac{(Y_t - \tilde{\Phi}_t)}{\tilde{\mathbf{f}}_t} (X_t \mathbf{b} + \tilde{c}) X_t' - \frac{(Y_s - \tilde{\Phi}_s)}{\tilde{\mathbf{f}}_s} (X_s \mathbf{b} + \tilde{c}) X_s' + (X_s - X_t). \quad (7)$$

The only case where the expected values of these derivatives do not vary with a coefficient is when all regressors are constant over time. To see this consider the derivative with respect to the coefficient of variable k only:

$$\frac{\partial m_{ts}(Z; \mathbf{b})}{\partial \mathbf{b}^k} = \frac{(Y_t - \tilde{\Phi}_t)}{\tilde{\mathbf{f}}_t} (X_t \mathbf{b} + \tilde{c}) X_t^k - \frac{(Y_s - \tilde{\Phi}_s)}{\tilde{\mathbf{f}}_s} (X_s \mathbf{b} + \tilde{c}) X_s^k + X_s^k - X_t^k.$$

Suppose now that $X_s^k = X_t^k$. It follows that:

$$\frac{\partial m_{ts}(Z; \mathbf{b})}{\partial \mathbf{b}^k} \Big|_{X_t^k = X_s^k} = \left\{ \frac{(Y_t - \tilde{\Phi}_t)}{\tilde{\mathbf{f}}_t} (X_t \mathbf{b} + \tilde{c}) - \frac{(Y_s - \tilde{\Phi}_s)}{\tilde{\mathbf{f}}_s} (X_s \mathbf{b} + \tilde{c}) \right\} X_t^k.$$

Taking the expectation and evaluating it at the true values of the coefficients using the approximation (4) gives a simplified expression:

$$E \left[\frac{\partial m_{ts}(Z; \mathbf{b}_0)}{\partial \mathbf{b}^k} \Big| X = x \right] \Big|_{X_t^k = X_s^k} = E[(C - \tilde{c}) | X = x] (x_t - x_s) \mathbf{b}_0 x_t^k.$$

Thus the coefficients of the time invariant regressors are identified provided there is at least one time varying regressor, and $E[C - \tilde{c} | X = x] \neq 0$. However, since this identification hinges on the local misspecification introduced by the Taylor approximation, it seems preferable not to attempt an estimation of the coefficients of the time invariant variables, and to subsume the impact of the latter in the individual effect. Finally, note that the identification of period-specific coefficients poses no problem.

Remark 5: In order to allow for *predetermined regressors*, a referee suggested replacing assumption (2) with

$$(U_t | C, \{X_s\}_{s \leq t}, \{U_s\}_{s \leq t}) \sim N(0,1). \quad (2')$$

This would yield

$$c - \tilde{c} = E \left[\frac{Y_t - \tilde{\Phi}_t}{\tilde{\mathbf{f}}_t} \Big| x_1, \dots, x_t, y_1, \dots, y_{t-1}, c \right] = E \left[\frac{Y_s - \tilde{\Phi}_s}{\tilde{\mathbf{f}}_s} \Big| x_1, \dots, x_s, y_1, \dots, y_{s-1}, c \right],$$

and thus for $t < s$

$$E\left[\frac{Y_t - \tilde{\Phi}_t}{\tilde{\mathbf{f}}_t} \mid x_1, \dots, x_t, y_1, \dots, y_{t-1}, c\right] = E\left[\frac{Y_s - \tilde{\Phi}_s}{\tilde{\mathbf{f}}_s} \mid x_1, \dots, x_t, y_1, \dots, y_{t-1}, c\right],$$

which yields the following set of conditional moment conditions:

$$E\left[\frac{Y_t - \tilde{\Phi}_t}{\tilde{\mathbf{f}}_t} - \frac{Y_s - \tilde{\Phi}_s}{\tilde{\mathbf{f}}_s} \mid x_1, \dots, x_t, y_1, \dots, y_{t-1}\right] = 0 \quad \text{for all } t \text{ and all } s > t. \quad (6')$$

While this is undoubtedly a worthwhile extension, it must be stressed that (2') implies independence between U_t and all other variables in the model, so that in particular Σ is restricted to be diagonal.

It remains to decide how the expansion point may be chosen, and Appendix B discusses our choice of $\tilde{c} = \Phi^{-1}(\bar{y})$, where \bar{y} denotes the grand mean of the dependent variable. The Monte Carlo study will show that this approach works fine when \bar{y} is not too far from 0.5. Otherwise it appears that the approximation error due to the linearization is more severe. We then propose to trim the sample to bring back \bar{y} close to 0.5. This trimming has to be based on explanatory variables to avoid selection bias. Details concerning two versions of this trimmed sample estimator can be found in Appendix C.

3.3 An estimator based on the alternative moment conditions

We consider GMM estimators of the type typically used with samples consisting of a large number of independent observations.⁹ The GMM estimator minimizes the following quadratic form:

$$\hat{\mathbf{b}}_N^{GMM} = \arg \min_{\mathbf{b} \in B} g_N(\mathbf{b})' W_N g_N(\mathbf{b}). \quad (8)$$

where \mathbf{b} denotes the parameter vector of dimension K and B denotes the corresponding parameter space. The vector of p moment functions $g_N(\mathbf{b})$ is computed as the average of the individual moment functions $g(z_i, \mathbf{b})$.

$$g_N(\mathbf{b}) = \frac{1}{N} \sum_{i=1}^N g(z_i; \mathbf{b}). \quad (9)$$

The $p \times p$ - dimensional weight matrix W_N is positive semi-definite and converges to some non- random matrix W . When standard regularity conditions are fulfilled – a crucial condition

for the asymptotic properties of that GMM estimator is that the individual moment functions have mean zero at the true values of the parameters ($Eg(Z; \mathbf{b}_0) = 0$) – the GMM estimator is consistent and asymptotically normal:

$$\sqrt{N}(\hat{\mathbf{b}}_N - \mathbf{b}_0) \xrightarrow{d} N(0, V); \quad V = (G'WG)^{-1}G'WPWG(G'WG)^{-1}; \quad (10)$$

$$G = E \frac{\mathbf{J}g(Z; \mathbf{b}_0)}{\mathbf{J}\mathbf{b}'}; \quad P = E[g(Z; \mathbf{b}_0)g(Z; \mathbf{b}_0)'].$$

Typically the individual moment functions can be decomposed in a conditional moment function and an instrument matrix:

$$g(z_i; \mathbf{b}) = \underbrace{A(x_i)}_{p \times q} \underbrace{m(z_i; \mathbf{b})}_{q \times 1}. \quad (11)$$

The corresponding condition for consistency and asymptotic normality in that case is:

$$E[m(Z; \mathbf{b}_0) | X = x] = 0, \quad \forall x \in \mathcal{C}, \quad \text{and this implies } E[A(X)m(Z; \mathbf{b}_0)] = 0.$$

In this case the matrix G appearing in the variance matrix V simplifies to:

$$G = E \left[A(X) \frac{\mathbf{J}m(Z; \mathbf{b}_0)}{\mathbf{J}\mathbf{b}'} \right].$$

Given the conditional moment functions, the asymptotically optimal instruments have the following form, where Q denotes some deterministic matrix of full rank:

$$A^*(x) = QD(x)\Omega(x)^{-1}; \quad (12)$$

$$D(x) = E \left[\frac{\mathbf{J}m(Z; \mathbf{b}_0)'}{\mathbf{J}\mathbf{b}} \mid X = x \right]; \quad (13)$$

$$\Omega(x) = E \{ [m(Z; \mathbf{b}_0)m(Z; \mathbf{b}_0)'] \mid X = x \}. \quad (14)$$

From a practical point of view using instruments of the type given in equations (12) to (14) has the advantage that the number of rows of the moment function $g_N(\cdot)$ equals the number of parameters (exact identification). Hence the weighting matrix W_N becomes irrelevant and

⁹ The issues relating to this type of GMM estimator are discussed extensively by Newey (1993) and Newey and McFadden (1994).

issues related to bad small sample properties of GMM due to the imprecise estimation of W_N disappear.

Let us relate the expressions in equations (11), (12), and (13) to our specification of the panel probit model. First, the conditional moment restriction has the following form:

$$m(Z, \mathbf{b}) = [\dots, m_{ts}(Z, \mathbf{b}), \dots]; \quad t = 1, \dots, T-1; s = t+1, \dots, T. \quad (15)$$

The derivative of a typical element of $m(Z, \mathbf{b})$ is given in equation (7). Taking expectations leads to the following equation:¹⁰

$$\begin{aligned} E\left[\frac{\partial m_{ts}(Z; \mathbf{b}_0)}{\partial \mathbf{b}^k} \mid X = x\right] &= d_{ts}(x; \mathbf{b}_0) \\ &= E\{(C - \tilde{c})[(x_t \mathbf{b}_0 + \tilde{c})x'_t - (x_s \mathbf{b}_0 + \tilde{c})x'_s] \mid X = x\} + (x_s - x_t)'. \end{aligned} \quad (16)$$

Without further knowledge about the joint distribution of the fixed effects and the regressors, this expression does not simplify and is difficult to estimate. Therefore, we use $\tilde{d}_{ts}(x; \mathbf{b}_0)$ instead, with

$$\tilde{d}_{ts}(x; \mathbf{b}_0) = (x_s - x_t)'. \quad (17)$$

Note that the resulting matrix has full rank if and only if the differences above are linearly independent for each pair of periods: this condition would be violated by sets of regressors moving over time in the same fashion, like for example age and potential experience. Note again also that period-specific coefficients would lead to a different expression for $d_{ts}(x; \mathbf{b}_0)$ and thus would lead to different instruments.

Similar computational problems appear with the expression for the conditional variance of the conditional moment functions:

$$\begin{aligned} \mathbf{w}_{ts,ts}(x) &= E\left[\left(\frac{Y_t - \tilde{\Phi}_t^0}{\tilde{\mathbf{f}}_t^0} - \frac{Y_s - \tilde{\Phi}_s^0}{\tilde{\mathbf{f}}_s^0}\right)^2 \mid X = x\right] \\ &= E\left[\left\{\left(\frac{Y_t - \tilde{\Phi}_t^0}{\tilde{\mathbf{f}}_t^0}\right)^2 + \left(\frac{Y_s - \tilde{\Phi}_s^0}{\tilde{\mathbf{f}}_s^0}\right)^2 - 2\frac{Y_t - \tilde{\Phi}_t^0}{\tilde{\mathbf{f}}_t^0}\frac{Y_s - \tilde{\Phi}_s^0}{\tilde{\mathbf{f}}_s^0}\right\} \mid X = x\right] \end{aligned}$$

¹⁰ The stochastic nature of $\tilde{c} = \Phi^{-1}(\bar{y})$ is ignored.

$$\begin{aligned}
&= E \left[\frac{Y_t + \tilde{\Phi}_t^{02} - 2Y_t \tilde{\Phi}_t^0}{\tilde{\mathbf{f}}_t^{02}} \mid X = x \right] + E \left[\frac{Y_s + \tilde{\Phi}_s^{02} - 2Y_s \tilde{\Phi}_s^0}{\tilde{\mathbf{f}}_s^{02}} \mid X = x \right] \\
&\quad - 2E \left[\frac{Y_t Y_s - \tilde{\Phi}_t^0 Y_s - Y_t \tilde{\Phi}_s^0 + \tilde{\Phi}_t^0 \tilde{\Phi}_s^0}{\tilde{\mathbf{f}}_t^0 \tilde{\mathbf{f}}_s^0} \mid X = x \right].
\end{aligned}$$

This expression cannot be estimated parametrically without further assumptions on the conditional distribution of the fixed effects given the explanatory variables.¹¹ In addition, information about the intertemporal correlation of the remaining error terms is needed. As a simplification, we assume that the errors terms u_{it} are uncorrelated over time, and – as before – that $(c_i = \tilde{c})$. Thus we have $\tilde{\mathbf{w}}_{is,ts}(x; \mathbf{b}_0) = \frac{\tilde{\Phi}_t^0 [1 - \tilde{\Phi}_t^0]}{(\tilde{\mathbf{f}}_t^0)^2} + \frac{\tilde{\Phi}_s^0 [1 - \tilde{\Phi}_s^0]}{(\tilde{\mathbf{f}}_s^0)^2}$. In addition the off-

diagonal elements are set to zero, and Ω has full rank, being a diagonal matrix with strictly positive diagonal elements.

Note that these simplifications have no influence on the consistency of the estimator, but affect its asymptotic variance. Taking everything together, we use the suboptimal – but fast and easy to compute – instruments $\tilde{A}(x, \tilde{\mathbf{b}}_N)$ given by equation (18):

$$\tilde{A}(x, \tilde{\mathbf{b}}_N) = \tilde{D}(x, \tilde{\mathbf{b}}_N) [\tilde{\Omega}(x, \tilde{\mathbf{b}}_N)]^{-1}, \quad (18)$$

where $\tilde{\mathbf{b}}_N$ denotes some initial (not even necessarily consistent) estimate for \mathbf{b}_0 . Examples for $\tilde{\mathbf{b}}_N$ are used in the Monte Carlo study, where we will refer to $\tilde{A}(x, \tilde{\mathbf{b}}_N)$ with $\tilde{\mathbf{b}}_N$ obtained from the pooled probit estimate as “pooled instruments” and to $\tilde{A}(x, \tilde{\mathbf{b}}_N)$ with $\tilde{\mathbf{b}}_N$ obtained from the GMM estimator with the pooled instruments as “consistent instruments”. The trimmed sample estimators will use $\tilde{\mathbf{b}}_N$ obtained from the GMM estimator with the pooled instruments for the first one, and the resulting estimate for the second one.

4 Monte Carlo study

This section presents a comparison of the performances of our estimators with those of the pooled probit estimator (PPE) and of the conditional logit estimator (CLE) for 9 data generating processes differing w.r.t. the correlation of the individual effects with the

¹¹ Non parametric estimation, in principle feasible, is cumbersome in the presence of numerous regressors.

regressors, the mean of the dependent variable, and the correlation over time of either the regressors or the idiosyncratic error term.

4.1 Presentation of the data generating processes

In all DGPs we consider three regressors: a regressor x_n which is standard normal, its square divided by 4, x_c , which is chi-square and without correlation with x_n (because the distribution of x_n is symmetric around 0 and has thus all odd moments 0), and a dummy, x_d , with mean 0.5, independent of the rest. The rationale for this extensive list of regressors is that we wish to mimic ‘real’ estimation problems. Except for DGP 3, the coefficients of these variables will be 1, -1, and 0.5, respectively, and 1, -1, and 2.5 for DGP 3 which also contains a constant equal to -1 . The latent variable further contains an $N(0,0.5)$ error term, independent across periods and individuals, and independent from the regressors, and an individual specific error term with variance 0.5. Unless otherwise mentioned, the observable dependent variable has unconditional expectation 0.5.

In DGP 1 (fixed effects), the individual effect c is correlated with x_n , but with none of the other regressors (although it will of course not be independent of x_c , which is a function of x_n), whereas it is independent of all regressors in DGP 2 (random effects). The exact specification of the error structure for DGP 1 is:

$$c_i = \frac{1}{\sqrt{2(T+1)}} \left[\sum_{t=1}^T \left(x_{n_{it}} + \frac{\mathbf{e}_{it}}{\sqrt{T}} \right) \right] \text{ with } \mathbf{e}_{it} \text{ iid } N(0,1).$$

As x_n is uncorrelated over time, this yields a correlation $\mathbf{r}(c, x_n) = 1/\sqrt{T+1}$. DGP1 constitutes our benchmark: with the exception of the pooled probit we expect all estimators to do well in that situation. DGP 3 and DGP 5 to DGP 9 introduce more difficult situations for our estimators, while DGP 4 should also prove difficult for the CLE. DGP 3 is as DGP 1 but with the modification mentioned above: the constant -1 is added, and the coefficient of the dummy is 2.5 instead of 0.5. DGP 4 is as DGP 1 but with the time varying error term AR(1), with coefficient 0.8. This is the DGP which is most adverse to CLE in our list. DGP 5 is as DGP 1, but with a constant of -1.2 , so that the observable dependent variable has unconditional expectation 0.22, and DGP 6 is the random effect equivalent, i.e. the corresponding modification of DGP 2. Finally, DGP 7, 8 and 9 are as DGP 1, 5 and 6, respectively, but with x_n following an AR(1) process with coefficient 0.9. This also affects

x_c , and considerably increases the correlation between c and x_n . The idea behind this choice is that our estimators and the CLE relate variation in y over time to variation in x over time, and we reduce both types of variation by introducing persistence.

For all DGPs we consider 6 combinations of N and T : $N=100, 400, 1600$, and $T=5, 10$. Table 1 summarises the statistical characteristics of the DGPs.

Table 1 around here

Two points deserve notice: firstly, we have departed substantially from the assumption of small individual effects in our set up, and thus put our estimators in situations that are a priori difficult for them; secondly our design produces somewhat smaller correlations between regressors and individual effects for the case $T=10$.

Measures of performance used in the Monte Carlo study include the bias in percentage points, the root mean square error (RMSE) and the median of absolute error (MAE). Note that we will differ from the usual conventions in referring to precision as measured by the mean square error rather than by the variance.

4.2 Discussion of Monte Carlo results

Because of the difficulty of proper scaling of the conditional logit parameters for comparison with the probit estimates and because binary choice models are only identified up to scale, we only discuss results obtained for the ratios of coefficients to the coefficient of the first regressor, x_n , and call these ratios $\mathbf{b}_c / \mathbf{b}_n$ and $\mathbf{b}_d / \mathbf{b}_n$. Note that these ratios are also the interesting identified quantities mentioned at the end of Subsection 2.2. The results discussed here can be found in Tables 2 to 5, whereby we only include the results for DGPs 1, 2, 7 and 8. The complete results are available on request. In the sequel we refer to the four estimators introduced in this paper as to the “almost consistent estimators” (ACEs).

Tables 2 to 5 around here

In order to acquaint the reader with the way the different DGPs affect the properties of the estimators, we first look at the performance of the well-known pooled probit and conditional logit estimators, which are also the most used in practice. Results concerning the latter are interesting in their own right, as many researchers tend to use the conditional logit estimator when they suspect a correlation between individual effects and regressors (see for instance Oswald, 1998).

4.2.1 Pooled probit

This simple estimator is consistent when individual effects are uncorrelated with the regressors, but is inefficient because it ignores correlations across observations for each individual. We thus expect it to give acceptable results in terms of asymptotic bias for DGPs 2, 6 and 9, which all have uncorrelated individual effects, but do not expect it to compare well to other estimators in terms of precision.

Indeed the pooled probit estimator (PPE) has a small bias for those three DGPs, even in the smaller samples. It is also fairly precise in comparison with the other estimators examined here, especially for $T=5$. For that number of periods it even dominates all other estimators in terms of bias, RMSE and MAE for DGP 9.

For the other DGPs it presents important and persistent biases (around 20% for DGPs 1, 3, 4 and 5, and 40% for DGPs 7 and 8). In terms of RMSE and MAE, its relative performance is not too bad for N large, at least as regards coefficient $\mathbf{b}_c / \mathbf{b}_n$. For the coefficient $\mathbf{b}_d / \mathbf{b}_n$ some relatively large values of RMSE appear, especially for DGPs 3 and 7.

4.2.2 Conditional logit

This estimator is based on maximum likelihood for the conditional distribution of the dichotomous dependent variable conditionally on its sum over time, as the latter is a sufficient statistic for the individual effect, under the assumption that the regressors are strictly exogenous and that the error term is i.i.d. with a logistic distribution and independent of the regressors and of the individual effect. These distributional assumptions are violated to various degrees by our DGPs, and so we cannot be assured a priori about the performance of the CLE for our samples. One clear feature is that it does not exploit observations with no change of status over time. For small T , and especially if the mean of the dependent variable substantially departs from 0.5, this will mean the loss of a large portion of the sample, and thus to a reduced efficiency relatively to methods that avoid such loss.¹² Table 6 shows the mean sample size used by conditional logit for each DGP for each T (percentages remain stable across the N dimension of the original samples, so that it is enough to look at $N=100$). The comparatively large remaining sample for DGP 3 is explained by the large coefficient of the indicator variable, which implies that a change in that variable, which is independently distributed over time with equal probabilities for the values 1 and 0, is almost sufficient for a

¹² One could argue that no efficiency loss is incurred by dropping uninformative observations. But the non-movers are only uninformative if nothing is known about the individual effects, whereas we assume here that these individual effects are small.

status switch to occur. It is somewhat surprising that the same numbers of observations are retained for DGPs 5 and 6 as for DGPs 1 and 2. The low numbers for DGPs 7 to 9 come from the autocorrelation in x_n , which induces a higher persistence in the dependent variable.

Table 6 around here

Given these considerations we may expect the conditional logit estimator to perform fairly well in all situations in terms of bias, except for DGP 4, but expect it to be dominated by other estimators in terms of precision when it rests on a smallish fraction of the original number of observations, as is the case for DGPs 7 to 9. Generally we also expect it to function better for $T=10$ than for $T=5$, for the same reason.

However, the results of the Monte Carlo study show a somewhat different picture. For the random effects DGP 2, CLE is uniformly the most biased estimator, and it is almost uniformly dominated in terms of precision by both the PPE and by all four ACEs. For DGP 9 (RE with one autocorrelated regressor), CLE performs relatively badly as regards $\mathbf{b}_c / \mathbf{b}_n$, with a bias that increases with N for both values of T , and the worst performance in terms of RMSE and MAE. Things are better for $\mathbf{b}_d / \mathbf{b}_n$ as regards the latter two, but the behaviour in terms of bias is also worrying, as bias does not monotonically decrease with N for fixed T . For DGP 6 the contrast between the two coefficients is starker, as for $\mathbf{b}_c / \mathbf{b}_n$ CLE performs poorly, especially for $T=5$, but is almost uniformly best on all criteria for $\mathbf{b}_d / \mathbf{b}_n$.

For the simple fixed effects DGP 1, CLE performs best on all accounts for $T=10$ and $N=1600$. It is also the least biased estimator for $\mathbf{b}_c / \mathbf{b}_n$, except for $T=5$ and $N=100$, where it is also dominated by all four ACEs in terms of precision. It uniformly dominates PPE except for $T=5$ and $N=100$. For DGP 3, CLE is mostly the least biased estimator and has similar performances in terms of RMSE and MAE as the ACEs, although it is more often dominated when $T=5$. For DGP 4, with $T=5$, the CLE is almost always dominated by the ACEs. Yet, for $T=10$ it is the least biased, but the difference vanishes with increasing N . For DGP 5, CLE is again uniformly best for $\mathbf{b}_d / \mathbf{b}_n$ and it is also almost uniformly best for $\mathbf{b}_c / \mathbf{b}_n$ for $T=10$. For $T=5$ it is dominated by the trimmed sample estimators (TSE). For DGP 7, CLE is the only estimator with a bias below 10% for $\mathbf{b}_c / \mathbf{b}_n$, but the bias appears not to vary monotonically, either with N or with T . In spite of this, CLE only dominates all other estimators in terms of RMSE and MAE for $N=1600$ (both for $T=5$ and $T=10$). Finally, DGP 8 turns out to be the most nasty DGP of all, and again CLE is the only estimator with a bias below 10% for both

coefficients, but the bias does *not* decrease with N . The only serious contender here is the second TSE, which tends to uniformly dominate CLE for N large.

What we have seen so far suggests that we should now separately discuss two groups among the estimators introduced in this paper. We will call the first group IVE, and the second has already been termed TSE.

4.2.3 Instrumental variables estimators

These are the *pooled* instruments and *consistent* instruments GMM estimators discussed at the end of Subsection 3.3, and we shall refer to them as IVEP and IVEC, respectively. They perform well for DGPs 1 to 4 and DGP 9, give mixed results for DGPs 6 and 7, and bad results for DGPs 5 and 8.

For the DGPs where they give good results we discuss in turn the DGPs with random effects (DGPs 2 and 9) and the DGPs with fixed effects (DGPs 1, 3, 4). For DGP 2, the IVEs dominate for almost all T and N . Only the TSEs prove better for $\mathbf{b}_c / \mathbf{b}_n$ for $T=5$ and $N=100$. For DGP 9 the IVEs are almost uniformly best for $T=10$ and are dominated only by PPE for $T=5$. For $N=1600$ they are dominated by the TSEs in terms of bias. For DGP 3 the performance of the IVEs is similar to and sometimes better than the performance of the CLE. For DGP 4, the IVEs are biased, much less than PPE, and only a little more than CLE for $T=10$ and for $T=5$ and $N=1600$. In terms of RMSE and MAE they dominate uniformly for $T=5$. For $T=10$ and $N=1600$, their performances are fairly close to those of CLE.

Turning to DGPs 6 and 7, the IVEs have the largest biases for DGP 6, but uniformly have the lowest RMSE and MAE for $\mathbf{b}_c / \mathbf{b}_n$, while for $\mathbf{b}_d / \mathbf{b}_n$ they are only superseded by CLE. For DGP 7, their bias is also important, about half the size of the bias of PPE, but in terms of RMSE and MAE the results are not bad, especially for IVEC, which dominates CLE for $T=5$, $N=100$, $N=400$, and for $T=10$, $N=100$.

Finally, for DGP 5, the bias for $\mathbf{b}_c / \mathbf{b}_n$ is high, and similar to that obtained with PPE, while for DGP 8 the bias, RMSE and MAE for $\mathbf{b}_c / \mathbf{b}_n$ are much higher than those of PPE, especially for IVEC. The picture is less bleak for $\mathbf{b}_d / \mathbf{b}_n$, but the IVEs are uniformly dominated by CLE and the TSEs.

4.2.4 Trimmed sample estimators

There is no DGP for which the performance of the TSEs is altogether bad. In particular, the TSEs appear to be almost consistent for all DGPs considered here, even though the reduction in bias with increasing N or T sometimes appears very slow (DGPs 1, 4, 7, 8). Their relative performance is often worst for the smallest sample sizes ($T=5$, $N=100$). We begin our detailed discussion with DGPs with random effects.

The TSEs perform well for DGP 2, where they are only dominated by the IVEs (almost uniformly for $T=10$, less for $T=5$) and sometimes by PPE ($\mathbf{b}_c / \mathbf{b}_n$, $T=5$). For DGP 6 they obtain the smallest bias for $\mathbf{b}_c / \mathbf{b}_n$ except for $T=5$ and $N=100$. In terms of precision they are clearly dominated by the IVEs for both coefficients, and by the CLE for $\mathbf{b}_d / \mathbf{b}_n$. For DGP 9 they are increasingly successful in eliminating bias as N increases, both for $T=5$ and $T=10$, but the first TSE encounters numerical problems for $T=5$, $N=100$, due to a few very small samples after trimming. They are uniformly dominated by the IVEs in terms of precision, and by PPE for $T=5$.

For DGPs 1, 3 and 4 they have larger biases than CLE and the IVEs and are also uniformly less precise than the latter. For $T=5$, $N=100$ they are more precise than CLE. For DGP 3 their performance comes close to that of the IVEs, while for DGP 4 they are a bit more biased than CLE and the IVEs, but come close in terms of precision for $\mathbf{b}_d / \mathbf{b}_n$. For DGP 5 their only contender is CLE, which dominates them uniformly for $T=10$. For $T=5$ CLE is more precise for $\mathbf{b}_d / \mathbf{b}_n$ (and also less biased for $N=100$), but the TSEs dominate uniformly for $\mathbf{b}_c / \mathbf{b}_n$. For DGP 7, they perform badly in terms of bias for $\mathbf{b}_c / \mathbf{b}_n$, with biases similar to those obtained for PPE, while for $\mathbf{b}_d / \mathbf{b}_n$ their performance comes close to those of the IVEs and CLE. For DGP 8, the second TSE realises the smallest bias for $\mathbf{b}_d / \mathbf{b}_n$ (except for $N=100$ where CLE does), and the second smallest bias for $\mathbf{b}_c / \mathbf{b}_n$, the successful contender being CLE (except for $T=10$ and $N=1600$ where they have the smallest biases). In terms of precision, the second TSE dominates uniformly for $\mathbf{b}_c / \mathbf{b}_n$ and almost uniformly for $\mathbf{b}_d / \mathbf{b}_n$ (except for $N=100$), and again the only contender is CLE.

These results are also documented in the density graphs of Figures 1 to 3, where the lines without symbols correspond to the PPE and the CLE. These are easily distinguished since the PPE is more concentrated and more biased in all cases.

Figures 1 to 3 about here

Thus the performance of the TSEs appears acceptable for all DGPs, as long as the sample is not very small.

5 Conclusion

In this paper we have introduced GMM estimators of the fixed effects probit model based on a first order Taylor approximation of the conditional expectation of the binary dependent variable given the regressors and the individual effects, $E(y_{it} | x_i, c_i)$, around a value \tilde{c} . The corresponding moment restrictions are approximately satisfied for small departures of c_i from \tilde{c} , and therefore the estimators based on this moment restriction will be almost consistent. The paper addresses questions of identification of the coefficients of time invariant regressors, and of the informativeness of the observation of non-movers in this situation. We also showed how heteroscedasticity over time can be accommodated, and how predetermined regressors could be dealt with.

In a Monte Carlo study we compared the performance of our estimators with those of the pooled probit and conditional logit (CLE) estimators, which are often used in practice. While we have not placed our estimators in a favourable situation, because the nine DGPs examined entail substantial variation in the individual effects, we found encouraging results: when the individual effects are correlated with the regressors and the mean of the dependent variable is near 0.5, our estimators outperform the CLE. They clearly dominate the CLE in the presence of random effects. When the sample mean of the dependent variable is far from 0.5, we resort to sample trimming in order to reduce the resulting bias. This turns out to work well in some situations where the bias of the CLE increases with sample size.

In brief, these estimators which can be used with larger numbers of regressors than most of the available semiparametric estimators, and also in situations where the scarcity of movers would defeat the CLE, should prove useful to the practitioner. An obvious extension of this work would be to devise tests based on comparisons between these estimators and possibly also with the CLE.

References

- Abrevaya, J. (2000): "Rank Estimation of a Generalized Fixed Effects Regression Model", *Journal of Econometrics*, 95, 1-23.
- Andersen, P.K. (1970): "Asymptotic Properties of Conditional Maximum Likelihood Estimators", *Journal of the Royal Statistical Society, Series B*, 32, 283-301.

- Arellano, M. and B. Honoré (2002): "Panel Data Models: Some Recent Developments", forthcoming in J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, Vol. V, North-Holland, Amsterdam.
- Avery, R., L.P. Hansen, and V.J. Hotz (1983): "Multiperiod Probit Models and Orthogonality Condition Estimation", *International Economic Review*, 24, 21-35.
- Bertschek, I., and M. Lechner (1998): "Convenient Estimators for the Panel Probit Model", *Journal of Econometrics*, 87, 329-371.
- Butler, J.S. and Moffitt, R. (1982): "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model", *Econometrica*, 50, 761-764.
- Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data", *Review of Economic Studies*, 47, 225-238.
- Chamberlain, G. (1984): "Panel Data", in Z. Griliches und M.D. Intriligator, Ed., *Handbook of Econometrics*, Vol. II, Ch. 22, North-Holland, Amsterdam.
- Chamberlain, G. (1992): "Binary Response Models for Panel Data: Identification and Information", *mimeo*, Harvard University.
- Charlier, E., B. Melenberg, and A. van Soest (1995): "A Smoothed Maximum Score Estimator for the Binary Choice Panel Model and an Application to Labour Force Participation", *Statistica Neerlandica*, 49, 324-342.
- Chen, S. (1998): "Root-N Consistent Estimation of a Panel Data Sample Selection Model", *unpublished discussion paper*, Princeton University.
- Chib, S. (1996): "Inference in Panel Data Models via Gibbs Sampling", Chapter 24 in L. Mátyás and P. Sevestre (eds.) *The Econometrics of Panel Data: Handbook of Theory and Applications*, 2nd edition (augmented) Kluwer, Amsterdam.
- Grant, D. (2000): "Estimating a Discrete Choice Fixed Effects Model Using Weighted Least Squares", *unpublished discussion paper*, Georgia Southern University, Statesboro.
- Heckman, J.J. (1981): "Statistical Models for Discrete Panel Data," in C. Manski and D. McFadden, Hrsg, Ed., *Structural Analysis of Discrete Data*, MIT Press, Cambridge, MA, 114-178.
- Heckman, J.J. (1981): "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time - Discrete Data Stochastic Process and Some Monte-Carlo Evidence," in C. Manski and D. McFadden, Ed., *Structural Analysis of Discrete Data*, MIT Press, Cambridge, MA, 179-195.
- Honore, B. and E. Kyriazidou (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables", *Econometrica*, 68, 839-874.
- Honore, B. and A. Lewbel (2002): "Semiparametric Binary Choice Panel Data Models Without Strictly Exogenous Regressors", *unpublished discussion paper*, Boston College.
- Horowitz, J. (1992): "A Smoothed Maximum Score Estimator for the Binary Response Model", *Econometrica*, 60, 505-531.
- Hsiao, C. (1986): *Analysis of Panel Data*. Cambridge: Cambridge University Press.
- Kyriazidou, E. (1995): *Essays in Estimation and Testing of Econometric Models*, Ph.D. dissertation, Northwestern University.
- Lancaster, T. (1999): "Panel Binary Choice with Fixed Effects", *unpublished discussion paper*, Brown University.
- Lancaster, T. (2000): "The Incidental Parameter Problem Since 1948", *Journal of Econometrics*, 95, 391-413.
- Lancaster, T. (2001): "Orthogonal Parameters and Panel Data", *unpublished discussion paper*, Brown University.
- Lechner, M. and J. Breitung (1996): "Some GMM Estimation Methods and Specification Tests for Nonlinear Models ", Chapter 22 in L. Mátyás and P. Sevestre (eds.) *The Econometrics of Panel Data: Handbook of Theory and Applications*, 2nd edition (augmented) Kluwer, Amsterdam.
- Lee, M.J. (1999): "A root-N Consistent Semiparametric Estimator for Related-Effect Binary Response Panel Data", *Econometrica*, 67, 427-433.
- Maddala, G.S. (1987): "Limited Dependent Variable Models Using Panel Data", *Journal of Human Resources*, 22, 307-338.
- Magnac, T. (2002): "Binary Variables and Fixed Effects: Generalizing the Conditional Logit", *unpublished discussion paper*, INRA and CREST, Paris.

- Manski, C. (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data", *Econometrica*, 55, 357-362.
- Newey, W.K. (1993): "Efficient Estimation of Models with Conditional Moment Restrictions", in G.S. Maddala, C.R. Rao, and H.D. Vinod (eds.), *Handbook of Statistics, Vol. 11: Econometrics*, Ch. 16, Amsterdam: North-Holland.
- Newey, W.K. (1994): "The Asymptotic Variance of Semiparametric Estimators", *Econometrica*, 62, 1349-1382.
- Newey, W.K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing", in R.F. Engle and D.L. McFadden (eds.), *Handbook of Econometrics, Vol. 4*, 2113-2245, Amsterdam: North-Holland.
- Oswald, A. (1998): "The Missing Piece of the Unemployment Puzzle", *unpublished discussion paper*, Warwick University.
- Rasch, G. (1960): *Probabilistic Models for Some Intelligence and Attainment Tests*, Denmark Paedagogiske Institut, Copenhagen.
- Wooldridge, J.M. (2000): "The Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity", *unpublished discussion paper*, Michigan State University.

Appendix A: Heteroscedasticity over time

When allowing for heteroscedasticity over time, we still have to set one of the variances to 1, say the first, and we define \mathbf{s} as the vector $\mathbf{s} = (\mathbf{s}_2, \dots, \mathbf{s}_T)'$. Now we replace (3) with

$$E(Y_t | X = x, C = c) = \Phi\left(\frac{x_t \mathbf{b}_0}{\mathbf{s}_t}\right); \quad t = 1, \dots, T. \quad (3')$$

and (4) with

$$\Phi\left(\frac{x_t \mathbf{b} + c}{\mathbf{s}_t}\right) \Big|_{c=\tilde{c}} \approx \Phi\left(\frac{x_t \mathbf{b} + \tilde{c}}{\mathbf{s}_t}\right) + \frac{c - \tilde{c}}{\mathbf{s}_t} f\left(\frac{x_t \mathbf{b} + c}{\mathbf{s}_t}\right); \quad t = 1, \dots, T. \quad (4')$$

Hence we have:

$$C - \tilde{c} = \frac{E(Y_t | X, C) - \Phi\left(\frac{x_t \mathbf{b} + \tilde{c}}{\mathbf{s}_t}\right)}{f\left(\frac{x_t \mathbf{b} + \tilde{c}}{\mathbf{s}_t}\right)} \mathbf{s}_t = \frac{E(Y_s | X, C) - \Phi\left(\frac{x_s \mathbf{b} + \tilde{c}}{\mathbf{s}_s}\right)}{f\left(\frac{x_s \mathbf{b} + \tilde{c}}{\mathbf{s}_s}\right)} \mathbf{s}_s, \quad s, t = 1, \dots, T.$$

The moment condition (5) becomes:

$$m_{ts}(Z; \mathbf{b}, \mathbf{s}) = \frac{Y_t - \Phi\left(\frac{x_t \mathbf{b} + \tilde{c}}{\mathbf{s}_t}\right)}{f\left(\frac{x_t \mathbf{b} + \tilde{c}}{\mathbf{s}_t}\right)} \mathbf{s}_t - \frac{Y_s - \Phi\left(\frac{x_s \mathbf{b} + \tilde{c}}{\mathbf{s}_s}\right)}{f\left(\frac{x_s \mathbf{b} + \tilde{c}}{\mathbf{s}_s}\right)} \mathbf{s}_s, \quad s, t = 1, \dots, T; \quad s \neq t. \quad (5')$$

As regards the identification of \mathbf{s} it is easy to check that

$$E\left\{\frac{\partial m_{ts}(Z; \mathbf{b}_0)}{\partial \mathbf{s}_t} \Big| X = x\right\} \Big|_{x_t^k = x_s^k} = \frac{E[(c - \tilde{c}) | X = x]}{\mathbf{s}_t} + \frac{x_t \mathbf{b}_0 + \tilde{c}}{\mathbf{s}_t},$$

and thus that \mathbf{s} will in general be identified unless both \mathbf{b}_0 and \tilde{c} are equal to zero.

We skip the details of the adjustment to the determination of \tilde{c} . In Section 3.2, we have to replace \mathbf{b} with the vector $\mathbf{q} = (\mathbf{b}, \mathbf{s})$, and (17) becomes

$$\tilde{d}_{ts}(x; \mathbf{q}) = \left[(x_s - x_t)', 0, \dots, 0, \frac{x_t \mathbf{b} + \tilde{c}}{\mathbf{s}_t}, 0, \dots, 0, -\frac{x_s \mathbf{b} + \tilde{c}}{\mathbf{s}_s}, 0, \dots, 0 \right]. \quad (17')$$

We will refrain from giving the bulky new expression for $\tilde{\mathbf{w}}_{ts}(x; \mathbf{q})$ and conclude by noting that additional moments for \mathbf{s} would increase efficiency.

Appendix B: Choice of the expansion point

Precisely because of the fact that the fixed effect absorbs the constant term and the time constant variables, choosing $\tilde{c} = 0$ is not attractive and some data dependent choice is called for.

The following equation defines the residual v_{it} :

$$y_{it} = \Phi(x_{it}\mathbf{b}_0 + c_i) + v_{it} \Rightarrow x_{it}\mathbf{b}_0 + c_i = \Phi^{-1}(y_{it} - v_{it}). \quad (\text{B.1})$$

As the next step we consider the first order Taylor expansion of the right hand side of (8) around the grand mean of y_{it} (\bar{y}) and go through simple algebraic steps justifying the choice of $\tilde{c} = \Phi^{-1}(\bar{y})$.

$$\begin{aligned} \Phi^{-1}(y_{it} - v_{it}) &\approx \Phi^{-1}(\bar{y}) - (y_{it} - v_{it} - \bar{y}) \frac{1}{\mathbf{f}[\Phi^{-1}(\bar{y})]}; & \bar{y} &= \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N y_{it}; \\ x_{it}\mathbf{b}_0 + c_i &\approx \Phi^{-1}(\bar{y}) - (y_{it} - v_{it} - \bar{y}) \frac{1}{\mathbf{f}[\Phi^{-1}(\bar{y})]}; \\ (x_{it} - \bar{x})\mathbf{b}_0 + (c_i + \bar{x}\mathbf{b}_0) &\approx \Phi^{-1}(\bar{y}) - (y_{it} - v_{it} - \bar{y}) \frac{1}{\mathbf{f}[\Phi^{-1}(\bar{y})]}; & \bar{x} &= \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N x_{it}; \\ (c_i + \bar{x}\mathbf{b}_0) &\approx \Phi^{-1}(\bar{y}) - (y_{it} - v_{it} - \bar{y}) \frac{1}{\mathbf{f}[\Phi^{-1}(\bar{y})]} - (x_{it} - \bar{x})\mathbf{b}_0; \\ \tilde{c} &\approx \Phi^{-1}(\bar{y}) - \frac{1}{N} \sum_{i=1}^N (y_{it} - v_{it} - \bar{y}) \frac{1}{\mathbf{f}[\Phi^{-1}(\bar{y})]} - \frac{1}{N} \sum_{i=1}^N (x_{it} - \bar{x})\mathbf{b}_0; & \tilde{c} &= \frac{1}{N} \sum_{i=1}^N (c_i + \bar{x}\mathbf{b}_0); \\ \tilde{c} &\approx \Phi^{-1}(\bar{y}) - \frac{\bar{v}}{\mathbf{f}[\Phi^{-1}(\bar{y})]}; & \bar{v} &= \frac{1}{NT} \sum_{t=1}^T \sum_{i=1}^N v_{it}. \end{aligned} \quad (\text{B.2})$$

Since in large samples $\bar{v} \approx 0$, choosing the expansion point as $\tilde{c} = \Phi^{-1}(\bar{y})$ appears to be natural. Note that in this case the explanatory variables should enter the estimation centred around their grand mean (over time and units).

Appendix C: Trimmed estimators

This appendix gives the details of the computation of the trimmed estimators used in the Monte Carlo study. The first trimmed estimator is obtained through the following steps:

- Step 1 Estimate 'standard' fixed effect probit model (GMM with pooled instruments : IVEP). This yields, say, $\tilde{\mathbf{b}}_N$.
- Step 2 Order units according to the mean of predicted index $(\bar{\hat{y}}_i = \frac{1}{T} \sum_{t=1}^T \hat{y}_{it})$, with $\hat{y}_{it} = x_{it} \tilde{\mathbf{b}}_N$.
- Step 3 Reduce the number of units according to the value of $\bar{\hat{y}}_i$, until the mean of predicted index is close to 0, or until $|\bar{y} - 0.5|N$ observations have been deleted.
- Step 4 Compute new value for \tilde{c} based on reduced sample, using equation (9). Redo centering of explanatory variables in the reduced sample.
- Step 5 Estimate 'standard' fixed effect probit model (GMM with consistent instruments: IVEC).

The second trimmed estimator is obtained by going through steps 1 to 5 a second time, using the estimate of the first TSE to build the instruments.

Table 1: Summary description of individual effects in the different DGPs.

DGP	1	2	3	4	5	6	7	8	9
<i>T=5</i>									
Min	-2.79	-2.24	-3.29	-2.42	-4.12	-3.76	-2.34	-3.44	-3.45
Max	2.66	2.81	1.30	2.26	1.25	1.08	2.05	0.97	1.54
Mean	-0.01	-0.00	-0.98	0.00	-1.22	-1.20	0.00	-1.18	-1.19
Std	0.70	0.71	0.73	0.70	0.72	0.71	0.71	0.71	0.70
$r(c, x_n)$	0.40	0	0.42	0.40	0.41	0	0.90	0.90	0
<i>T=10</i>									
Min	-2.66	-2.20	-3.34	-2.22	-3.61	-3.54	-2.16	-3.38	-3.84
Max	2.69	2.71	1.55	2.69	1.27	1.27	2.15	1.16	1.06
Mean	0.04	-0.03	-0.98	-0.05	-1.17	-1.21	0.04	-1.18	-1.21
Std	0.72	0.73	0.70	0.71	0.70	0.70	0.71	0.71	0.72
$r(c, x_n)$	0.31	0	0.30	0.30	0.29	0	0.85	0.84	0

Table 2: Simulation results: Estimated coefficients for DGP 1 (correlated individual effects)

	Bias in %		Root MSE x 10		Median abs. err. x 10	
	b_c / b_n	b_d / b_n	b_c / b_n	b_d / b_n	b_c / b_n	b_d / b_n
<i>N</i> = 100						
<i>T</i> = 5, 1000 replications						
<i>pooled probit</i>	-27.6	-21.8	3.97	1.54	2.51	1.13
conditional logit	-9.8	-1.9	4.73	1.45	2.98	0.95
pooled instruments	-9.2	-0.9	3.46	1.38	2.13	0.90
consistent instr.	-10.2	0.5	3.43	1.36	2.08	0.88
trimmed sample 1	-14.4	1.0	3.83	1.39	2.26	0.92
trimmed sample 2	-15.0	1.1	3.93	1.40	2.28	0.91
<i>N</i> = 400						
<i>T</i> = 5, 1000 replications						
<i>pooled probit</i>	-24.1	-22.3	2.77	1.24	2.29	1.13
conditional logit	-4.5	0.4	1.29	0.72	1.29	0.49
pooled instruments	-5.4	-0.1	1.67	0.69	1.03	0.47
consistent instruments	-6.1	-0.2	1.67	0.68	1.04	0.45
trimmed sample 1	-10.0	0.9	1.98	0.71	1.21	0.46
trimmed sample 2	-10.7	1.0	2.06	0.69	1.28	0.46
<i>N</i> = 1600						
<i>T</i> = 5, 1000 replications						
<i>pooled probit</i>	-22.7	-22.2	2.36	1.14	2.28	1.11
conditional logit	-2.0	0.6	1.07	0.37	0.68	0.25
pooled instruments	-3.9	0.2	0.88	0.34	0.59	0.24
consistent instruments	-4.6	0.1	0.89	0.34	0.60	0.24
trimmed sample 1	-8.3	1.3	1.19	0.35	0.85	0.25
trimmed sample 2	-9.1	1.4	1.27	0.35	0.92	0.25
<i>N</i> = 100						
<i>T</i> = 10, 500 replications						
<i>pooled probit</i>	-19.1	-17.3	2.70	1.19	1.82	0.90
conditional logit	-2.9	-1.2	2.34	0.88	1.51	0.59
pooled instruments	-5.2	-0.1	2.02	0.89	1.24	0.62
consistent instruments	-6.0	-0.1	2.03	0.90	1.25	0.60
trimmed sample 1	-8.2	0.4	2.25	0.91	1.33	0.61
trimmed sample 2	-8.3	0.4	2.25	0.91	1.37	0.64
<i>N</i> = 400						
<i>T</i> = 10, 500 replications						
<i>pooled probit</i>	-18.9	-17.8	2.10	0.97	1.90	0.91
conditional logit	-2.0	-0.1	1.14	0.44	0.79	0.29
pooled instruments	-4.6	-0.1	1.09	0.44	0.70	0.33
consistent instruments	-5.3	-0.1	1.11	0.44	0.73	0.33
trimmed sample 1	-7.4	0.5	1.26	0.46	0.82	0.33
trimmed sample 2	-7.6	0.6	1.28	0.46	0.83	0.33
<i>N</i> = 1600						
<i>T</i> = 10, 500 replications						
<i>pooled probit</i>	-17.9	-17.8	1.85	0.91	1.79	0.89
conditional logit	-0.4	0.0	0.61	0.21	0.41	0.15
pooled instruments	-3.4	-0.0	0.62	0.23	0.45	0.16
consistent instruments	-4.0	0.1	0.66	0.23	0.47	0.16
trimmed sample 1	-6.0	0.7	0.82	0.23	0.61	0.15
trimmed sample 2	-6.2	0.8	0.84	0.23	0.62	0.16

Note: bold: best performance; italic: worst (or worst after pooled probit); grey: better than conditional logit, or same.

Table 3: Simulation results: Estimated coefficients for DGP 2 (uncorrelated indiv. effects)

	Bias in %		Root MSE x 10		Median abs. err. x 10	
	b_c / b_n	b_d / b_n	b_c / b_n	b_d / b_n	b_c / b_n	b_d / b_n
N = 100 <i>T</i> = 5, 1000 replications						
pooled probit	-2.3	0.2	2.73	1.31	1.78	0.85
<i>conditional logit</i>	-12.9	1.1	4.83	1.50	2.87	1.00
pooled instruments	-1.0	1.0	2.87	1.29	1.88	0.87
consistent instruments	-0.9	1.0	2.79	1.28	1.88	0.88
trimmed sample 1	-1.5	-0.2	2.78	1.37	1.77	0.91
trimmed sample 2	-1.5	-0.1	2.86	1.39	1.74	0.93
N = 400 <i>T</i> = 5, 1000 replications						
pooled probit	-0.5	-0.4	1.25	0.66	0.80	0.44
<i>conditional logit</i>	-4.9	0.6	2.18	0.75	1.36	0.49
pooled instruments	-0.1	-0.6	1.25	0.64	0.82	0.42
consistent instr.	-0.2	-0.5	1.24	0.64	0.81	0.43
trimmed sample 1	-0.2	-0.7	1.30	0.66	0.85	0.43
trimmed sample 2	-0.0	-0.7	1.34	0.68	0.90	0.45
N = 1600 <i>T</i> = 5, 1000 replications						
pooled probit	-0.4	0.1	0.60	0.33	0.42	0.23
<i>conditional logit</i>	-2.6	0.1	1.07	0.37	0.76	0.25
pooled instruments	-0.5	-0.0	0.60	0.32	0.38	0.22
consistent instr.	-0.5	-0.0	0.60	0.32	0.38	0.22
trimmed sample 1	-0.5	-0.1	0.63	0.33	0.41	0.23
trimmed sample 2	-0.5	-0.1	0.65	0.34	0.41	0.25
N = 100 <i>T</i> = 10, 500 replications						
pooled probit	-2.2	-0.1	1.83	0.92	1.20	0.60
<i>conditional logit</i>	-3.6	0.6	2.51	0.88	1.57	0.59
pooled instruments	-1.4	-0.5	1.78	0.85	1.12	0.57
consistent instr.	-1.5	-0.5	1.77	0.85	1.17	0.57
trimmed sample 1	-1.6	-0.7	1.92	0.87	1.23	0.59
trimmed sample 2	-1.5	-0.9	1.98	0.90	1.18	0.60
N = 400 <i>T</i> = 10, 500 replications						
pooled probit	-0.1	-0.2	0.90	0.46	0.60	0.32
<i>conditional logit</i>	-1.9	-0.5	1.13	0.46	0.77	0.32
pooled instruments	-0.2	-0.1	0.84	0.43	0.57	0.29
consistent instr.	-0.3	-0.1	0.84	0.43	0.58	0.29
trimmed sample 1	-0.3	-0.1	0.88	0.44	0.61	0.28
trimmed sample 2	-0.5	0.0	0.91	0.44	0.61	0.27
N = 1600 <i>T</i> = 10, 500 replications						
pooled probit	0.1	0.1	0.46	0.24	0.34	0.16
<i>conditional logit</i>	-0.7	0.4	0.58	0.22	0.40	0.14
pooled instruments	0.1	0.2	0.42	0.22	0.28	0.15
consistent instr.	0.1	0.2	0.42	0.22	0.28	0.15
trimmed sample 1	0.1	0.2	0.44	0.23	0.31	0.15
trimmed sample 2	0.1	0.2	0.45	0.23	0.32	0.16

Note: bold: best performance; italic: worst performance (or worst after pooled probit); grey: better than conditional logit.

Table 4: Simulation results: Estimated coefficients for DGP 7 (as DGP 1, but x_n AR(1))

	Bias in %		Root MSE x 10		Med. abs. err. x 10	
	b_c / b_n	b_d / b_n	b_c / b_n	b_d / b_n	b_c / b_n	b_d / b_n
N = 100 T = 5, 1000 replications						
<i>pooled probit</i>	-48.4	-39.7	6.31	2.13	4.06	1.97
conditional logit	-7.9	1.6	10.7	1.76	7.17	1.10
pooled instruments	-19.5	2.3	7.29	1.86	5.17	1.18
consistent instruments	-31.8	-0.3	6.45	1.71	3.96	1.09
trimmed sample 1	-40.1	-0.0	7.27	1.72	4.53	1.10
<i>trimmed sample 2</i>	-47.0	0.7	7.96	1.74	4.83	1.10
N = 400 T = 5, 1000 replications						
<i>pooled probit</i>	-42.3	-38.7	4.72	1.97	4.10	1.95
conditional logit	-3.5	0.6	4.74	0.89	2.97	0.63
pooled instruments	-23.3	1.2	4.21	0.83	2.82	0.56
consistent instruments	-28.0	0.6	4.09	0.78	2.52	0.52
trimmed sample 1	-37.8	1.2	5.00	0.79	3.49	0.54
<i>trimmed sample 2</i>	-46.5	1.4	5.89	0.79	4.31	0.54
N = 1600 T = 5, 1000 replications						
<i>pooled probit</i>	-40.7	-39.0	4.11	1.96	3.86	1.95
conditional logit	-5.3	0.2	2.40	0.40	1.58	0.28
pooled instruments	-22.7	-0.4	2.87	0.42	2.31	0.29
consistent instruments	-24.6	-0.3	2.89	0.40	2.44	0.28
trimmed sample 1	-34.2	0.3	3.82	0.41	3.37	0.28
<i>trimmed sample 2</i>	-43.2	0.4	4.71	0.41	4.28	0.28
N = 100 T = 10, 500 replications						
<i>pooled probit</i>	-43.7	-37.8	5.22	1.97	4.00	1.92
conditional logit	-8.9	0.3	5.24	1.07	3.46	0.74
pooled instruments	-22.1	-0.4	4.45	1.06	2.86	0.75
consistent instruments	-26.0	-1.2	4.36	1.00	2.49	0.70
trimmed sample 1	-36.2	-0.5	5.30	1.02	3.37	0.71
<i>trimmed sample 2</i>	-43.5	-0.3	6.11	1.02	3.94	0.70
N = 400 T = 10, 500 replications						
<i>pooled probit</i>	-38.4	-37.8	4.09	1.91	3.75	1.91
conditional logit	-5.3	0.8	2.60	0.50	1.69	0.36
pooled instruments	-17.9	-0.9	2.73	0.54	1.74	0.36
consistent instruments	-18.4	-0.8	2.64	0.52	1.66	0.35
trimmed sample 1	-27.4	0.1	3.43	0.52	2.58	0.35
<i>trimmed sample 2</i>	-34.7	0.3	4.14	0.52	3.41	0.35
N = 1600 T = 10, 500 replications						
<i>Pooled probit</i>	-36.3	-37.4	3.69	1.87	3.59	1.87
conditional logit	-6.6	0.2	1.51	0.25	1.13	0.16
pooled instruments	-17.5	-0.0	1.99	0.27	1.76	0.18
consistent instruments	-16.8	0.1	1.88	0.26	1.67	0.17
trimmed sample 1	-26.0	1.0	2.77	0.27	2.58	0.17
<i>trimmed sample 2</i>	-33.5	1.2	3.53	0.27	3.36	0.17

Note: bold: best performance; italic: worst performance (or worst after pooled probit); grey: better than conditional logit.

Table 5: Simulation results: Estimated coefficients for DGP 8 (as DGP 7, but $E(Y) = 0.22$)

	Bias in %		Root MSE x 10		Med. abs. err. x 10	
	b_c / b_n	b_d / b_n	b_c / b_n	b_d / b_n	b_c / b_n	b_d / b_n
N = 100 $T = 5$, 1000 replications						
pooled probit	-45.4	-38.5	5.58	2.21	4.14	1.19
conditional logit	-5.2	4.2	7.27	2.47	3.63	1.39
pooled instruments	-72.6	26.3	<i>11.0</i>	3.37	6.28	1.54
<i>consistent instruments</i>	<i>-83.1</i>	<i>36.7</i>	10.8	3.25	7.66	1.74
trimmed sample 1	-34.2	18.8	7.99	3.55	3.17	1.41
trimmed sample 2	-30.2	17.9	8.56	3.78	3.00	1.41
N = 400 $T = 5$, 1000 replications						
pooled probit	-41.3	-39.4	4.38	2.02	3.99	1.99
conditional logit	5.8	-2.2	2.97	1.01	2.04	0.67
pooled instruments	-69.7	19.0	7.80	1.49	6.85	0.94
<i>consistent instruments</i>	<i>-77.0</i>	<i>31.0</i>	<i>8.50</i>	<i>1.94</i>	7.39	1.47
trimmed sample 1	-18.7	6.0	3.24	1.12	1.78	0.67
trimmed sample 2	-12.4	2.2	2.73	1.08	1.46	0.68
N = 1600 $T = 5$, 1000 replications						
pooled probit	-39.9	-38.7	4.05	1.95	3.95	1.93
conditional logit	6.8	-4.2	1.55	0.54	1.12	0.38
pooled instruments	-66.9	19.3	6.91	1.13	6.61	0.93
<i>consistent instruments</i>	<i>-71.2</i>	<i>31.0</i>	<i>7.34</i>	1.66	7.02	1.52
trimmed sample 1	-12.7	4.9	1.69	0.58	1.20	0.35
trimmed sample 2	-8.1	1.7	1.29	0.54	0.83	0.35
N = 100 $T = 10$, 500 replications						
pooled probit	-40.6	-36.7	4.55	1.94	3.82	1.86
conditional logit	2.5	0.1	3.17	1.20	1.99	0.84
pooled instruments	-57.9	18.8	6.99	1.66	5.27	1.11
<i>consistent instruments</i>	<i>-68.4</i>	<i>30.0</i>	7.93	2.09	6.39	1.51
trimmed sample 1	-12.6	5.0	3.08	1.33	1.62	0.88
trimmed sample 2	-7.9	2.5	2.82	1.36	1.51	0.90
N = 400 $T = 10$, 500 replications						
pooled probit	-38.5	-37.2	3.98	1.88	3.73	1.86
conditional logit	3.9	-1.8	1.61	0.63	1.14	0.47
pooled instruments	-56.9	18.1	6.05	1.14	5.50	0.88
<i>consistent instruments</i>	<i>-64.6</i>	<i>29.0</i>	<i>6.82</i>	<i>1.62</i>	6.17	1.40
trimmed sample 1	-10.0	3.6	1.71	0.69	0.92	0.42
trimmed sample 2	-6.0	1.0	1.40	0.65	0.71	0.41
N = 1600 $T = 10$, 500 replications						
pooled probit	-37.3	-37.6	3.76	1.89	3.72	1.89
conditional logit	6.1	-3.2	0.93	0.34	0.67	0.25
pooled instruments	-54.8	17.1	5.58	0.92	5.50	0.85
<i>consistent instruments</i>	<i>-61.7</i>	<i>27.7</i>	<i>6.26</i>	<i>1.44</i>	6.09	1.37
trimmed sample 1	-7.0	2.0	0.94	0.35	0.64	0.23
trimmed sample 2	-3.9	-0.5	0.70	0.33	0.44	0.24

Note: bold: best performance; italic: worst performance (or worst after pooled probit); grey: better than conditional logit.

Table 4: Mean number of individuals retained for conditional logit and trimmed estimators

DGP	1	2	3	4	5	6	7	8	9
<i>N = 100, T = 5</i>									
conditional logit	72	72	83	64	72	72	41	41	41
trimmed sample 1	93	94	96	92	50	43	93	60	46
trimmed sample 2	92	91	95	91	42	25	91	55	34
<i>N = 100, T = 10</i>									
conditional logit	89	89	96	64	83	89	61	55	55
trimmed sample 1	93	93	97	92	93	43	92	57	43
trimmed sample 2	92	90	96	91	92	23	91	53	31

Bold: largest; italics: smallest.

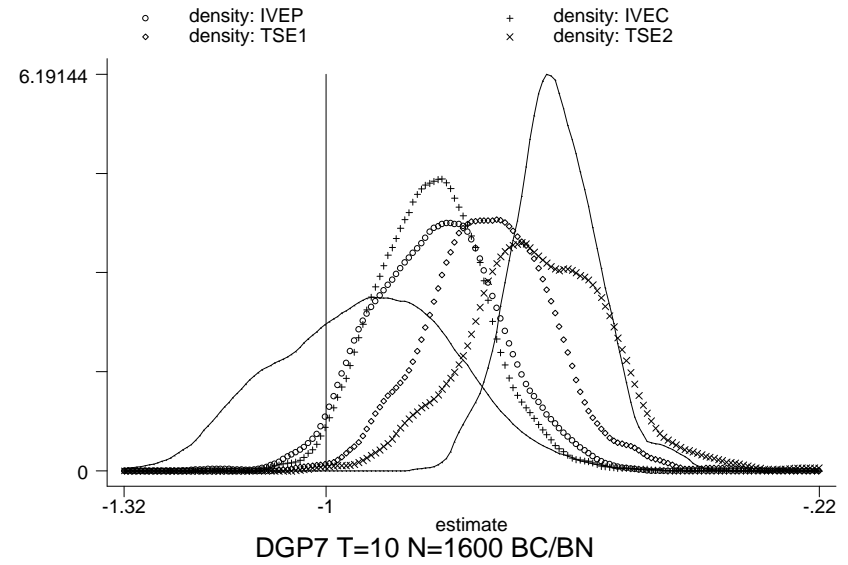
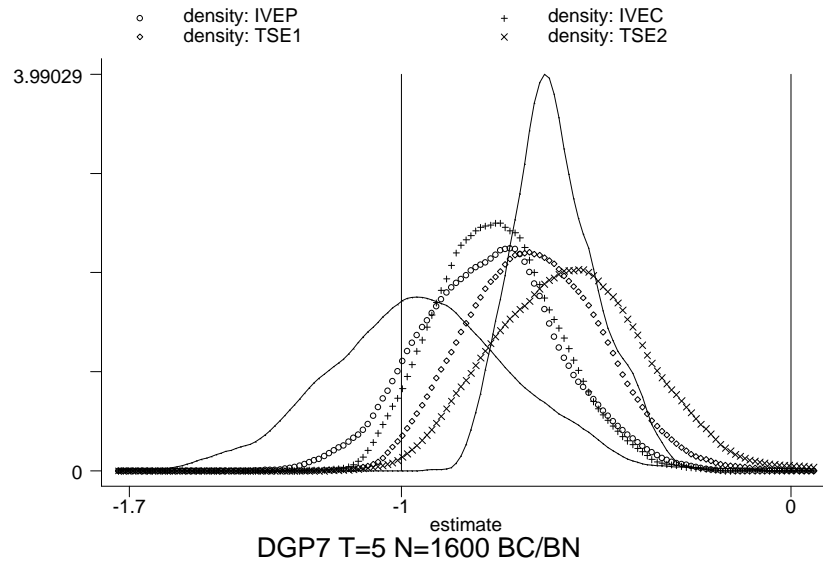
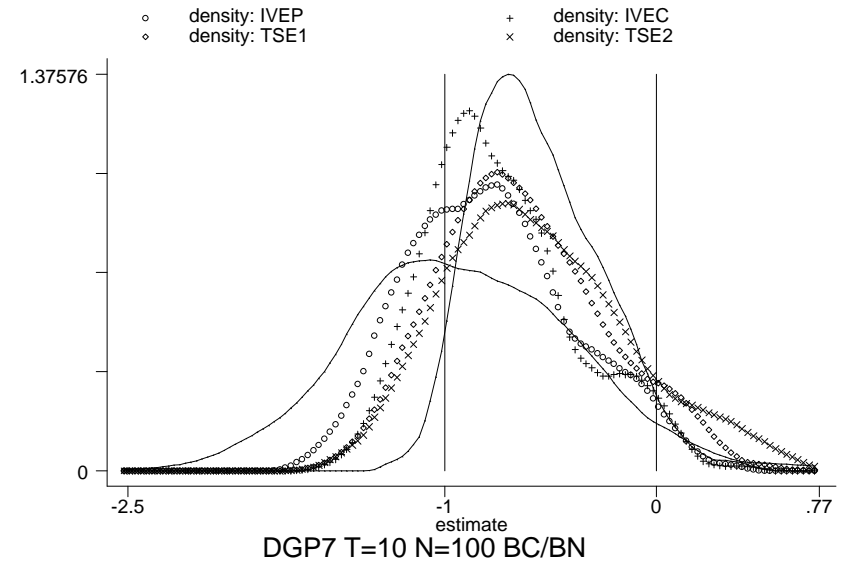
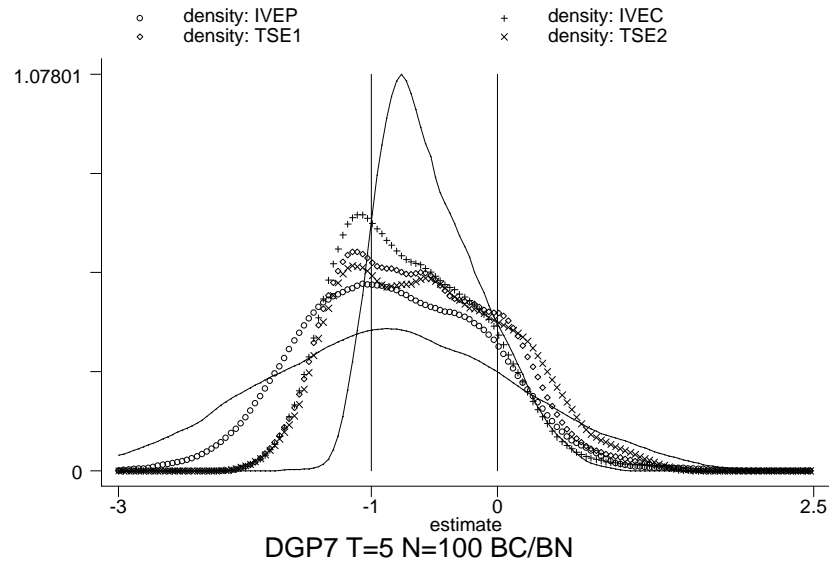


Figure 1: Density of $\hat{\mathbf{b}}_c / \hat{\mathbf{b}}_n$ for correlated individual effects and x_n AR(1), autocorr. 0.9

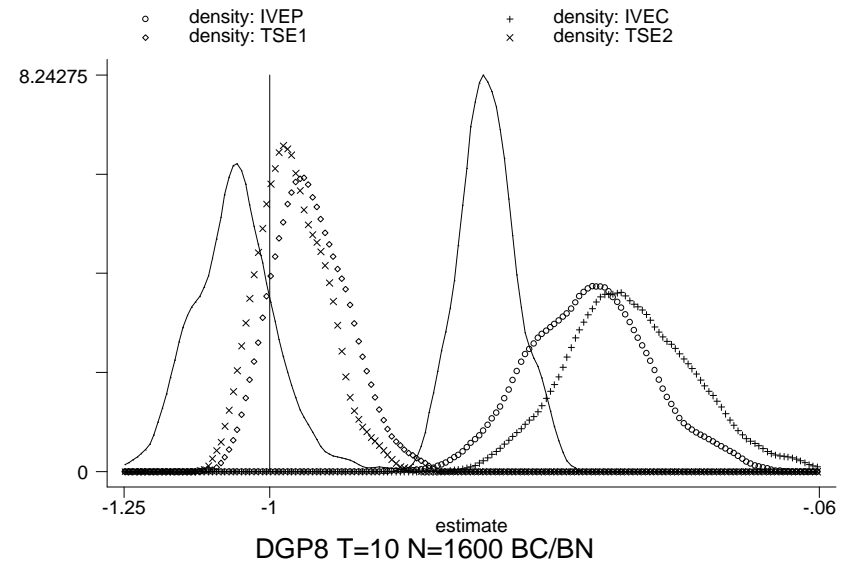
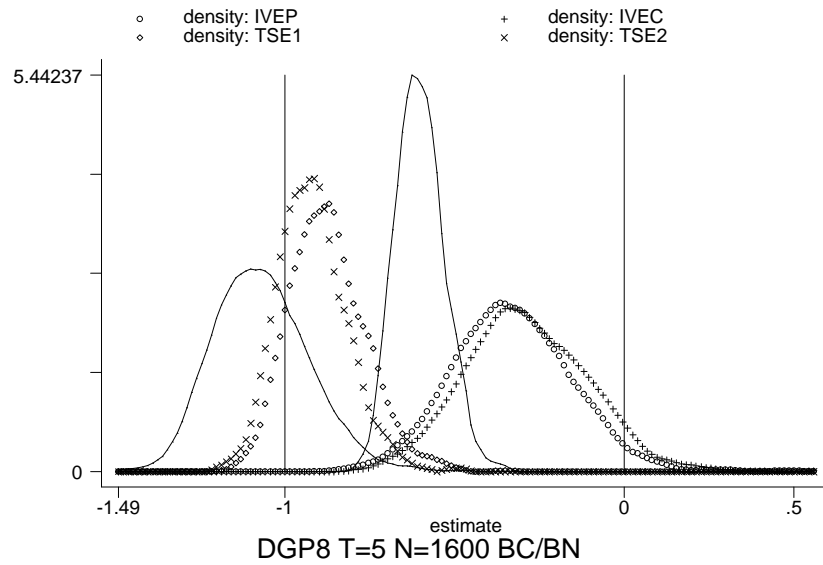
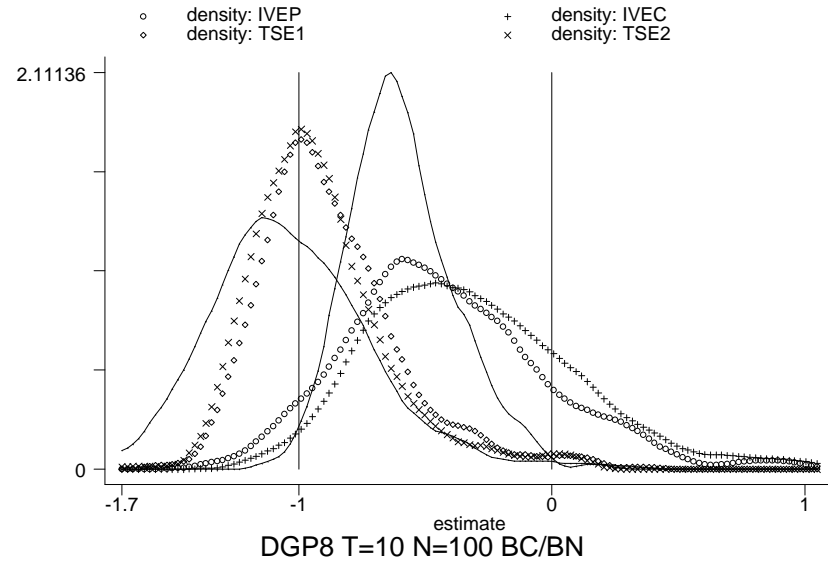
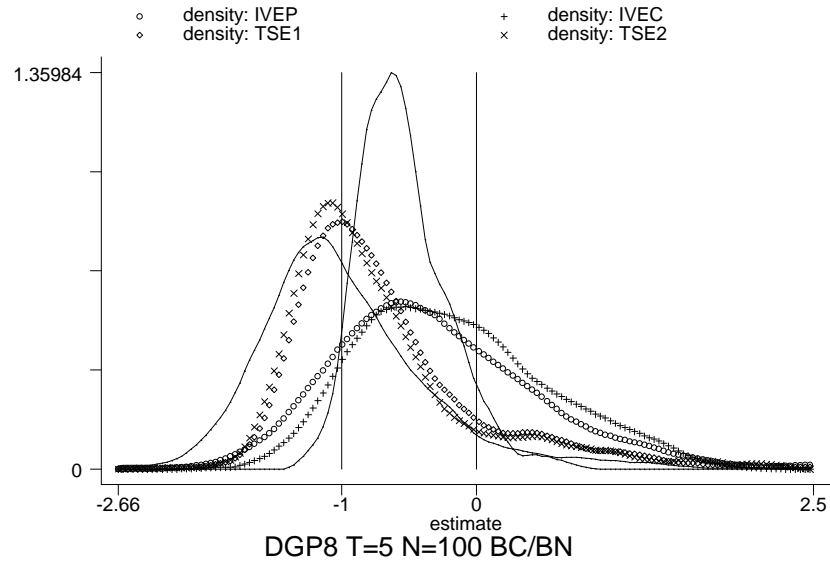


Figure 2: Density of $\hat{\mathbf{b}}_c / \hat{\mathbf{b}}_n$, correlated individual effects, x_n AR(1), autocorr. 0.9, $E(Y)=0.22$

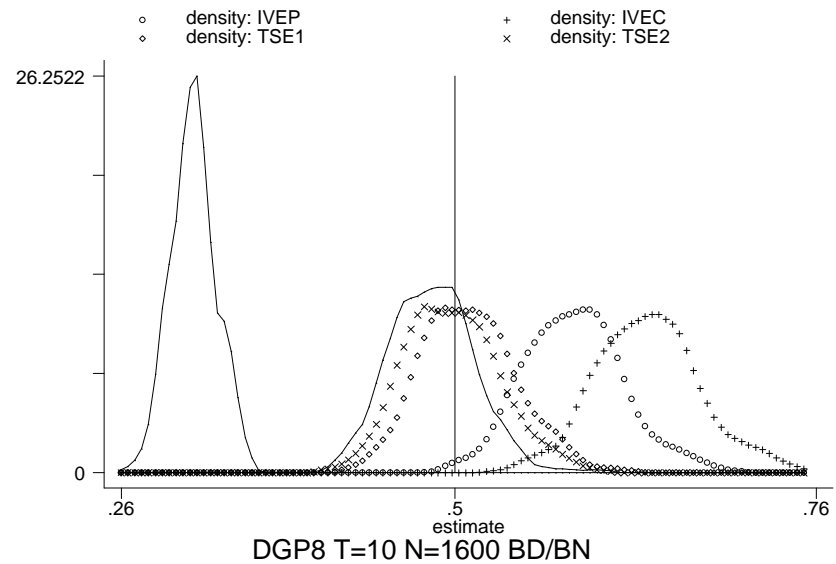
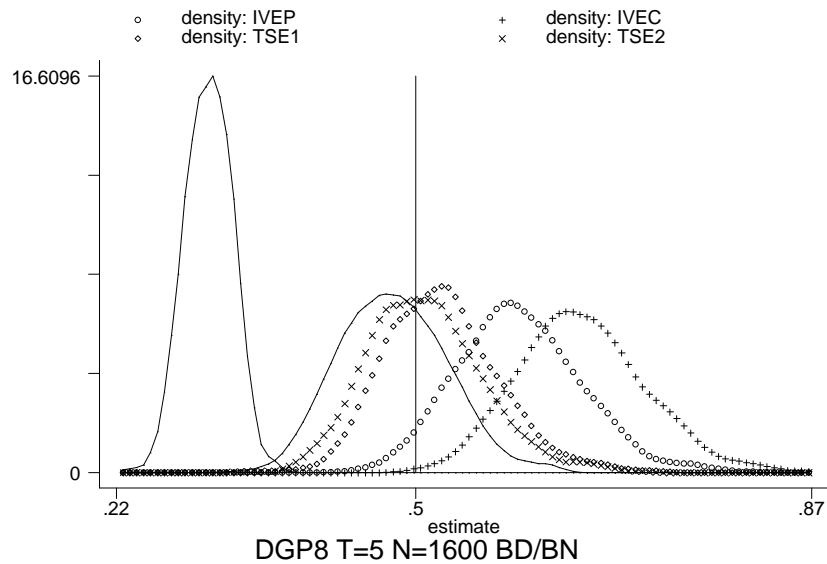
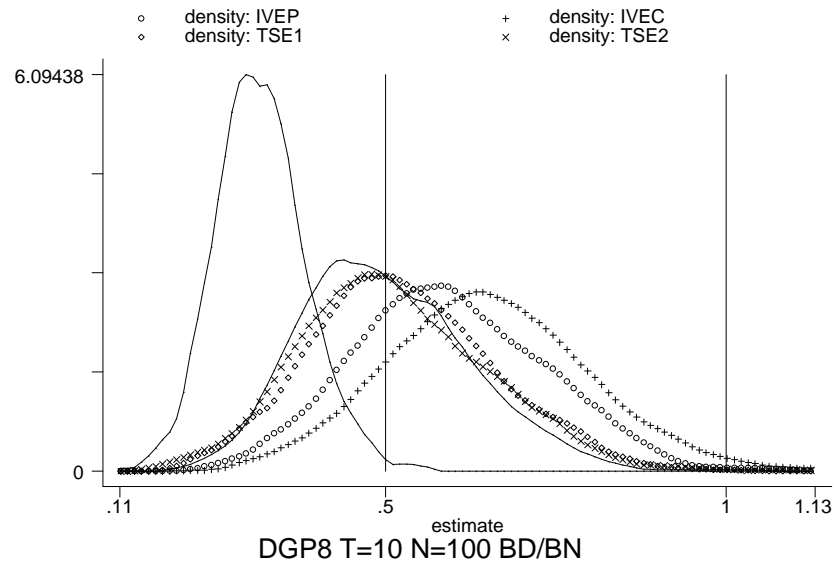
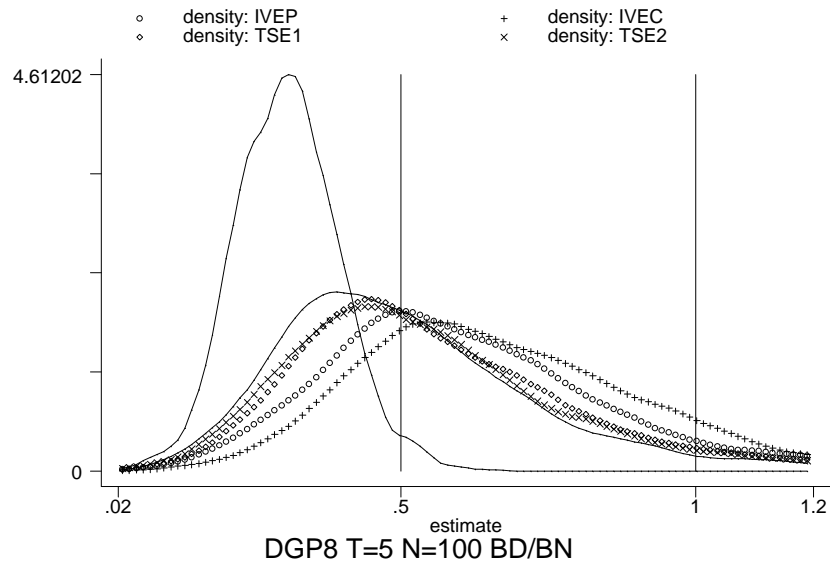


Figure 3: Density of $\hat{\mathbf{b}}_d / \hat{\mathbf{b}}_n$, correlated individual effects, x_n AR(1) with autocorrelation .0.9, and $E(Y)=0.22$