Binary variables and fixed effects: generalizing conditional logit

Thierry Magnac, INRA, Paris-Jourdan and CREST*

First version: October 2000
This version: September 2001

Abstract

We extend the conditional logit approach used in panel data models of binary variables with correlated fixed effects. In a two-period two-state model, we derive the necessary and sufficient conditions on the joint distribution function of the individual and period specific shocks such that the sum of individual binary variables across time is a sufficient statistic for the individual effect. The conditional likelihood function, conditional on the sum of the binary variables, does not therefore depend on individual effects. It leads to $\sqrt{n}$ consistent and normally distributed estimates. Conditions are a lot less stringent than in the conditional logit approach (Rasch, Andersen, Chamberlain) and give justifications to quasi-differencing the binary variables as if they were continuous variables. The "conditional probit" method in particular is shown to be a perfectly legitimate approach. Semiparametric approaches drawn from the cross-section literature can also be readily applied.

* *Address:* 48, Bld Jourdan, 75014 Paris, France, *Email* Thierry.Magnac@ens.fr

1. Introduction[1]

Using conditional likelihood methods in the estimation of panel data models of binary variables with correlated fixed effects, is a well known "semi-parametric" technique since it avoids specifying the distribution of individual effects conditional on covariates (Rasch, 1960, Andersen, 1973, Chamberlain, 1984). It is based on the result that there exists a sufficient statistic for the individual effect which is the number of times the binary variable is equal to 1. By definition of sufficiency, the conditional likelihood function depends on the parameters of interest only while the marginal likelihood function still depends on both parameters of interest and nuisance parameters, the individual effects (Barndorff-Nielsen and Cox, 1993, Lancaster, 2000). The conditional likelihood method is however seen to be restrictive because of the logistic assumption, the only applicable case to my knowledge up to now.

In this paper, we extend conditional likelihood methods by relaxing the assumption of independence between individual and period specific shocks across periods. We use a standard two-period two-state model. The sequence of individual binary variables is described by latent variables which are assumed to be the sum of a linear index of explanatory variables, of the individual effect and of the individual and period specific shocks. We generalize the conditional logit approach by solving the question of the existence of a sufficient statistic. Namely, we present the conditions on the joint distribution function of individual and period specific shocks such that a sufficient statistic for the individual effect is the number of times the binary variable is equal to 1. We give restrictions that should be imposed on the conditional likelihood function. They are a lot less stringent than in the conditional logit approach. These results lead to the construction of an estimating equation relating the expectation of the difference between subsequent binary variables and the difference in the linear index of covariates. The method therefore can be interpreted as a method giving the conditions for quasi-differencing binary variables and allowing standard packages to be used to estimate binary panel data models.

[1] I thank Jean-Pierre Florens and participants at seminars (CREST, Toulouse) and conferences, EC^2 in Dublin, ESEM in Lausanne for helpful comments. The usual disclaimer applies.

Binary models for panel data have been the focus of interest in the literature for a long time (see Arellano and Honoré, 1999, for a recent survey). Most papers use “random” effect models where the distribution function of disturbances is assumed to be independent of covariates. Moreover, most authors parametrically specify this distribution function including the cases where the individual effect is assumed to have a discrete and finite support. Very few papers use non parametric methods (Chen, Heckman and Vytlacil, 1999). Turning to so called fixed effect models where the distribution function of individual effects can depend in a unspecified way on covariates, we can distinguish four different estimation approaches : maximum score (Manski, 1987), maximum rank correlation (Lee, 1999, Abrevaya, 2000), ”pseudo”-regressions (Honoré and Lewbel, 2000) and likelihood methods such as conditional logit. We can first notice that the first three approaches are all based on methods that were first applicable to estimating models of binary variables in a cross section. In the maximum score approach, covariates are strictly exogenous. Under strict exogeneity, Horowitz (1992) extended this approach by proposing smooth maximum score and a panel data application was performed by Charlier, Melenberg and Van Soest (1995). For maximum rank correlation methods, the strict exogeneity assumption is also maintained. In contrast, the adaptation of Lewbel (2000) estimation method to binary variables panel data do not require strict exogeneity but requires that the continuous regressor should be independent of the individual effect. Likelihood approaches require strict exogeneity. Besides the weak or strong exogeneity assumptions, these approaches also differ in terms of their asymptotic properties. Maximum score estimation is not root-n consistent and asymptotic distributions of estimates are non normal but smooth maximum score is “almost” root-n consistent and asymptotically normal (Horowitz, 1992). We do not know much about the efficiency of maximum rank correlation and pseudo regressions though we know they are consistent.

Conditional likelihood approaches are firmly rooted in likelihood theory as are, in a different spirit, approaches based on Cox and Reid (1987) and the Bayesian methods using orthogonal parameters (Lancaster, 2001). Two important results related to conditonal logit were shown by Chamberlain (1992). Under the assumption of independence between individual and period specific shocks across periods, 1) if covariates are bounded, identification is possible only in the logistic case, 2) if covariates are unbounded, consistent estimation at a $\sqrt{n}$ rate is possible only in the logistic case. We are able to generalize the second result of Chamberlain (1992) in two ways. First, the only joint distribution function such that 1) the individual and period random shocks are independent 2) the sum of the binary variables is a sufficient statistic for the individual effect, is the logistic distribution. Inversely, if covariates are unbounded, then consistent estimation at a $\sqrt{n}$ rate is possible only in the case where the sum of the binary variables is a sufficient statistic for the individual effect. The absence of "mixture" between the parameters of interest and the non-parametric part seems to lead to a gain in the convergence rate. These results should be related to results of Chen, Heckman and Vytlacil (1999) in a random effect setting.

Because of the simplicity of the conditional likelihood approach, we believe that it could be the first step to attempt before considering random effect approaches. An empirical illustration is performed on a labour market participation example and show the interest of the method. Section 2 illustrates the principle of sufficiency in the conditional logit case. Section 3 is the main theoretical section. We there derive a characterization of the joint distribution function when there exists a sufficient statistic and derive the conditions on the primitives of the problem. We study estimation and give examples in section 4 and report the results of the empirical illustration in section 5. Section 6 is devoted to extensions and section 7 concludes.

2. The conditional logit model

We shall first concentrate on how the conditional logit approach works. Consider y_{i1}, y_{i2} two binary variables, in an identically and independently distributed sample $i = 1, ., n$ given by the following model:

$$y_{it} = 1 \text{ if and only if } z_{it}\beta + \epsilon_i + u_{it} > 0$$

where (u_{i1}, u_{i2}) is independent of $(z_{i1}, z_{i2}, \varepsilon_i)$:

The method described by Chamberlain consists in specifying that u_{i1} and u_{i2} are independent across periods and have logistic marginal distribution functions. The likelihood of an observation, dropping index i for simplicity, is:

$$\Pr(y_1, y_2 \mid z_1, z_2, \epsilon) = \frac{\exp(\sum_{t=1}^{2} y_t(z_t\beta + \epsilon))}{\prod_{t=1}^{2}(1 + \exp(z_t\beta + \epsilon))}$$

and for $K = 0, 1, 2$, the likelihood of the sufficient statistic, $\sum y_t$, is given by:

$$\Pr(\sum_{t=1}^{2} y_t = K \mid z_1, z_2, \epsilon) = \sum_{\sum_{t=1}^{2} y_t = K} \frac{\exp(\sum_{t=1}^{2} y_t(z_t\beta + \epsilon))}{\prod_{t=1}^{2}(1 + \exp(z_t\beta + \epsilon))}$$

Conditioning on this statistic yields:

$$\begin{aligned}\frac{\Pr(y_1, y_2 \mid z_1, z_2, \epsilon)}{\Pr(\sum_{t=1}^{2} y_t = K \mid z_1, z_2, \epsilon)} &= \frac{\exp(\sum_{t=1}^{2} y_t(z_t\beta + \epsilon))}{\sum_{\sum_{t=1}^{2} y_t = K} \exp(\sum_{t=1}^{2} y_t(z_t\beta + \epsilon))} \\ &= \frac{\exp((\sum_{t=1}^{2} y_t z_t)\beta)}{\sum_{\sum_{t=1}^{2} y_t = K} \exp((\sum_{t=1}^{2} y_t z_t)\beta)}\end{aligned}$$

which is independent of ϵ (and which is non equal to 1 only if $K = 1$). The sufficient statistic $\sum_{t=1}^{2} y_t = K$ provides a global cut for this likelihood

A strong criticism of this approach is that period specific shocks are independent and logistically distributed (Chamberlain, 1984, 1992). If we assume that it is not the case, the principle may however still be used if:

$$\frac{\Pr(y_1, y_2 \mid z_1, z_2, \epsilon)}{\Pr(\sum_{t=1}^{2} y_t = 1 \mid z_1, z_2, \epsilon)} \tag{2.1}$$

is independent of ϵ. Consider $x_i = -z_i\beta - \epsilon$ and assume that period specific shocks are independent of z_1, z_2 and ε. Then :

$$\Pr(y_1 = 1, y_2 = 0 \mid z_1, z_2, \epsilon) = \Pr(u_1 > x_1, u_2 \leq x_2)$$

Then the condition that $\sum_{t=1}^{2} y_t$ is sufficient for the individual effect is that:

$$\frac{\Pr(u_1 > x_1, u_2 \leq x_2)}{\Pr(u_1 \leq x_1, u_2 > x_2)} = c(x_1 - x_2) \tag{2.2}$$

The aim of this paper is to find characterizations of function c and the joint d.f. of (u_1, u_2) such that this property holds.

3. Characterization

The estimating condition (2.2) can easily be used for estimation purposes (see section 4) by using either conditional maximum likelihood or a generalized method of moments. It is however by no means obvious that the function c is unconstrained. We shall first derive what are the general implications of the sufficiency property on the joint d.f. of (u_1, u_2). We will then turn to the conditions that shall be imposed on function c.

3.1. The joint distribution function

We shall limit ourselves to the class of following distribution functions in order to use the usual tools of differential calculus. It should however be noted that the statements below can be translated and adapted to the case where the different functions are defined almost everywhere at the price of additional complexity.

Assumption A1 : *The distribution function of (u_1, u_2) admits a strictly positive and bounded density function with respect to the Lebesgue measure. The marginal distribution function of u_1 is denoted $F(.)$ and its density function, $f(.)$. It is such that $F(0) = \frac{1}{2}$.*

The assumption on the median is the more common normalization assumption. It suffices to include a constant among variables z_i. If we also include a period specific variable in z_1 and not in z_2, then by normalization we shall also assume that :

Assumption A2: $c(0) = 1$

Under these assumptions, the following theorem characterizes some necessary conditions for (2.2) to hold.

Theorem 3.1. *Assume A1, A2 and :*

$$\forall(x_1, x_2); \frac{\Pr(u_1 > x_1, u_2 \leq x_2)}{\Pr(u_1 \leq x_1, u_2 > x_2)} = c(x_1 - x_2) \tag{3.1}$$

Then :

1. *$c(h)$ is a decreasing function from $+\infty$ to 0 and is twice differentiable.*
2. *The marginal d.f. of u_2 is equal to the marginal d.f. of u_1:*

$$\Pr(u_2 \leq x_2) = F(x_2)$$

3. *The joint d.f. of* (u_1, u_2) *is given by :*

$$\Pr(u_1 \leq x_1, u_2 \leq x_2) = \frac{F(x_2) - c(x_1 - x_2)F(x_1)}{1 - c(x_1 - x_2)} \text{ when } x_1 \neq x_2 \tag{3.2}$$

$$\Pr(u_1 \leq x_1, u_2 \leq x_1) = F(x_1) + \frac{f(x_1)}{c'(0)} \tag{3.3}$$

4. $c'(0) < 0$

5. $F(.)$ *is differentiable three times and* f" *is bounded.*

Proof. It is easy to prove point 1, by using monotonicity properties of probability functions, limit conditions and assumption A1. For points 2 and 3, see appendix A. The second expression in point 3 is derived by noting that expression (3.2) is technically undefined when $x_1 - x_2 = 0$. We consider the limit of equation (3.2) when $x_2 \to x_1$. Expanding $F(x_2) \simeq F(x_1) + (x_2 - x_1)f(x_1)$ and $c(x_1 - x_2) \simeq 1 + c'(x_1 - x_2)(x_1 - x_2)$, and taking limits, we get (3.3). It is only defined when $c'(0) < 0$ (point 4). As the joint density function is bounded because of (A1), we can differentiate (3.3) two times and the result is bounded. Therefore, $F(.)$ necessarily is three-times differentiable and f" is bounded (point 5). ■

The fact that u_1 and u_2 have identical distribution functions is general. It should be reminiscent of the property of exchangeability of variables in a sequence which are at the heart of the method of conditional logit and also of the score method developed by Manski (1987). This property is here shown to be the consequence of the sufficiency property and we also allowed for an additional difference in intercepts through A2.

Theorem 3.1 gives also a characterization of the joint probability function in terms of the two functions $F(.)$ and $c(.)$ only. The sufficiency property, as expected, reduces the dimension of the problem from a function with two arguments to two functions of one argument. At this stage of the paper, theorem 3.1 gives, however, a necessary condition only and this characterization might lead to an improper joint distribution function. The aim of this section is therefore to look for necessary and sufficient conditions on the two primitives $F(.)$ and $c(.)$ that lead to a proper d.f.

We nevertheless begin with an interesting application of this theorem when u_1 and u_2 are independent.

Corollary 3.2. *Assume A1, A2, condition (3.1) and that* u_1 *and* u_2 *are indepen-*

dent. Then, there exists $\mu > 0$ such that:

$$F(x) = \frac{1}{1+\exp(-\mu x)} \qquad c(h) = \exp(-\mu h)$$

Proof. Condition (3.3) and the independence assumption imply that :

$$(F(x))^2 = F(x) + \frac{f(x)}{c'(0)} \tag{3.4}$$

For any x, $0 < F(x) < 1$, and denote:

$$\lambda(x) = -\log(\frac{1-F(x)}{F(x)})$$

(3.4) implies that :

$$\lambda'(x) = \frac{f(x)}{F(x)(1-F(x))} = -c'(0) = \mu$$

Integrating this equation and imposing $F(0) = \frac{1}{2}$, we get the expression of $F(x)$. Using (3.2) we get the expression of $c(h)$. ■

In the present paper, we do not impose independence between individual and period specific random variables and a broader class of functions will be shown to lead to a sufficiency property. What we can infer from this corollary is that the intersection of the set of joint distribution functions satisfying this sufficiency property and the set of independent joint distribution functions is reduced to one family, the logistic distribution function. Chamberlain (1992) directly proved that the logistic distribution function is the only possible d.f. consistent with a non-zero semi-parametric efficiency bound when random shocks are independent. This property seems to be similar to the result derived by Chamberlain (1992) who directly proved that the logistic distribution function is the only possible d.f. consistent with a non-zero semi-parametric efficiency bound when random shocks are independent. We shall relate his result to ours in the section about extensions later on.

In our case, it is still not clear that (3.2) and (3.3) define a proper joint distribution function and that the class of joint distribution functions defined by Theorem 3.1 is not reduced to the logistic case. A necessary condition is that the joint density function be strictly positive and bounded and therefore satisfies condition A1. To define the joint density function from equation (3.2), let :

$$\forall x_1 \neq x_2; s(x_1 - x_2) \equiv \frac{c'(x_1 - x_2)}{(1 - c(x_1 - x_2))^2}$$

Proposition 3.3. *Denote :*

$$\forall x_1 \neq x_2; g(x_1, x_2) \equiv s(x_1 - x_2)(f(x_1) + f(x_2)) + s'(x_1 - x_2)(F(x_1) - F(x_2))$$

and by extension:

$$g(x_1, x_1) = \lim_{x_2 \to x_1} g(x_1, x_2)$$

Necessary and sufficient conditions for (3.2) to verify assumption A1 is:

$$\forall (x_1, x_2); \quad 0 < g(x_1, x_2) < +\infty \tag{3.5}$$

Proof. As shown in appendix B, $g(x_1, x_2)$ is the density function that can be derived from (3.2). Furthermore :

$$\lim_{x_1, x_2 \to -\infty} \Pr(u_1 \leq x_1, u_2 \leq x_2) = 0 \qquad \lim_{x_1, x_2 \to +\infty} \Pr(u_1 \leq x_1, u_2 \leq x_2) = 1$$

■

It is difficult to directly tackle this condition. We shall proceed step by step. The first step shows that there is a one-to-one relationship between function $c(.)$ and the distribution function of the difference, $u_1 - u_2$.

3.2. The relationship between $c(.)$ and the distribution function of the difference $u_1 - u_2$

We first show how to derive the distribution function of the difference $u_1 - u_2$ from the function of $c(.)$. It also yields additional necessary conditions that must be imposed on $c(.)$.

Let $\phi(.)$ be the marginal d.f. of $u_1 - u_2$ which admits a bounded and positive density function $\varphi(.)$ by assumption A1. Then :

Proposition 3.4. *Under assumptions A1, A2 and condition (3.1) then :*

$$\forall h; 0 < \frac{d^2}{dh^2} \frac{h}{1 - c(h)} < +\infty \qquad \lim_{h \to +/-\infty} \frac{c'}{(1-c)^2} h = 0 \tag{3.6}$$

Under these conditions :

$$\phi(h) = \frac{d}{dh} \frac{h}{1 - c(h)}$$

Proof. See appendix C. ■

We can also solve the inverse problem deriving $c(.)$ from the d.f. $\phi(h)$. When we solve this inverse problem, we find additional conditions, some of those being related to the necessary and sufficient condition (3.5).

Proposition 3.5. *Let $\phi(h)$ the distribution function of $u_1 - u_2$. Under assumptions A1 and A2, and condition (3.1), we necessarily have:*

$$\int_0^{+\infty} \tau\varphi(\tau)d\tau = \int_0^{-\infty} \tau\varphi(\tau)d\tau < +\infty \tag{3.7}$$

$$\lim_{h\to+\infty} h(1-\phi(h)) = 0 \qquad \lim_{h\to-\infty} h\phi(h) = 0 \tag{3.8}$$

$$\exists\beta_0 > 0; \lim_{h\to+/-\infty} \frac{|h\varphi(h)|}{\int_h^{+\infty} \tau\varphi(\tau)d\tau} > \beta_0 \tag{3.9}$$

Under these conditions:

$$c(h) = 1 - \frac{h}{h\phi(h) + \int_h^{+\infty} \tau\varphi(\tau)d\tau}$$

is a decreasing function from $+\infty$ to 0, is twice differentiable, $c(0) = 1, c'(0) < 0$ and $c(.)$ verifies (3.6).

Proof. See appendix D. ■

The first condition is related to the fact that u_1 and u_2 are identically distributed and implies that :

$$E(u_1 - u_2) = 0$$

even when Eu_1 does not exist. The other conditions are regularity conditions on the d.f. at infinity. These regularity conditions are for instance verified if $\phi(.)$ is the normal d.f. and, as we will see in the next section, by many other standard distributions.

The last proposition gives a way to construct function $c(.)$ for a given d.f. $\phi(.)$. It also gives additional conditions on $c(.)$ that we now translate:

Corollary 3.6. *If $c(.)$ verifies the conditions of proposition 3.5 then :*

$$\lim_{h\to-\infty} h\frac{d}{dh}(\frac{h}{1-c}) = 0 \text{ and } \lim_{h\to+\infty} h(1-\frac{d}{dh}(\frac{h}{1-c})) = 0 \tag{3.10}$$

$$\exists\beta_0 > 0; \min(-\lim_{h\to+\infty} \frac{s'(h)}{s(h)}, \lim_{h\to-\infty} \frac{s'(h)}{s(h)}) > \beta_0$$

Proof. See appendix E. ■

We finish this section by noting the fundamental symmetry of the problem in terms of the disturbances u_1 and u_2 which we will use to simplify the proofs given in the next section. If we change u_1 into u_2 and u_2 into u_1, we simply change $\phi(h)$ into $\phi(-h)$ and $c(h)$ into $1/c(h)$. The marginal distributions $F(.)$ are not changed since they are identical. The symmetry can be seen from the original equation (3.1) or directly from the expression of $c(.)$ because we get after some manipulations and because $E(u_1 - u_2) = 0$:

$$\frac{1}{c(h)} = 1 - \frac{h}{h(1-\phi(h)) + \int_{-\infty}^{h} \tau\varphi(\tau)d\tau}$$

Equation (3.5) is also affected in the same way since it can be equivalently written as (setting $x_1 = x + h$, $x_2 = x$):

$$\forall (h,x); 0 < s(h)(f(x+h) + f(x)) + s'(h)(F(x+h) - F(x)) < +\infty \qquad (3.11)$$

or as:

$$\forall (h,x); 0 < s(-h)(f(x-h) + f(x)) + s'(-h)(F(x-h) - F(x)) < +\infty$$

which is equivalent to:

$$\forall (h,x); 0 < s(-h)(f(x) + f(x+h)) - s'(-h)(F(x+h) - F(x)) < +\infty \qquad (3.12)$$

Equations (3.11) and (3.12) are therefore equivalent to, for any $h \geq 0$ and any x:

$$\begin{aligned} 0 &< s(h)(f(x+h) + f(x)) + s'(h)(F(x+h) - F(x)) < +\infty \\ 0 &< s(-h)(f(x+h) + f(x)) - s'(-h)(F(x+h) - F(x)) < +\infty \end{aligned}$$

We can therefore limit the developments above to the case of $h \geq 0$ **provided that** we verify the conditions bearing on the straight representation $\phi(h)$ (resp. $c(h)$) and on the reverse representation $1 - \phi(h)$ (resp. $1/c(h)$). We shall therefore assume from now on that $h \geq 0$ and that $\phi(h)$ can either be interpreted as the proper $\phi(h)$ or $1 - \phi(h)$.

Given this simplification, we can tackle the implications of the necessary and sufficient condition (3.5).

3.3. Conditions on the marginal distribution function $F(.)$

We are now in a position to prove that there exist distribution functions $F(.)$ such that condition (3.5) is satisfied. To prove the point more easily, we have to restrict further the conditions given in proposition 3.5. We write an additional condition on the d.f. $\phi(.)$:

$$\exists \alpha_0 > 0 \text{ such that } \forall h \geq 0; \frac{\alpha_0 \mathrm{sh}(\alpha_0 h)}{\mathrm{ch}(\alpha_0 h) - 1} - \frac{2}{h} < \frac{h\varphi(h)}{\int_h^{+\infty} \tau\varphi(\tau)d\tau} \tag{3.13}$$

where sh(.) and ch(.) are the hyperbolic sine and cosine functions. [2]

This condition is not as stringent as it may first appear. First, note that as the function on the left hand side is increasing from 0 when $h = 0$ to α_0 when $h \to +\infty$, condition (3.9) on the limit behaviour of the term on the RHS, is a consequence of condition (3.13). Second, we show after the main result[3] that under very weak conditions, condition (3.9) implies condition (3.13).

The following proposition gives the conditions on the marginal distribution function $F(.)$.

Proposition 3.7. *For any function $\phi(.)$ satisfying (3.7, 3.8, 3.13), there exist functions $F(.)$ satisfying (3.5). They are necessarily such that :*

$$\forall x; \frac{f"(x)}{f(x)} < \alpha_0^2 \leq \frac{6\varphi(0)}{\int_0^{+\infty} \tau\varphi(\tau)d\tau} \tag{3.14}$$

where α_0 is defined by (3.13).

Proof. See appendix F. ■

It means that f should not be "too" convex. It is the consequence of (3.13) when we solve for equation (3.5). The tails of the distribution are thick enough and the normal distribution, say, would not qualify for this condition. There are many distribution functions, however, that satisfy theis condition, in particular, the double-exponential or Laplace distribution as shown in appendix F.

[2] $\mathrm{ch}(h) = \dfrac{\exp(h) + \exp(-h)}{2}$ $\quad \mathrm{sh}(h) = \dfrac{\exp(h) - \exp(-h)}{2}$

[3] The order is commanded by the fact that this statement uses developments that are given in the proof of proposition 3.7.

We now weaken condition (3.13) which is difficult to verify in most cases to condition (3.9) which is easy to verify.

Lemma 3.8. *Let $\varphi(h)$ be a strictly positive density function such that :*

i). It is continuous except possibly at a finite number of points.

ii). It verifies (3.9)

Then, it satisfies (3.13).

Proof. **:** See appendix G. ■

Therefore, for most density distributions that one is willing to consider , the important condition to verify is (3.9). We can also write this result in terms of function $c(.)$. Condition (3.13) can be rewritten using the one to one relationship between $\phi(.)$ and $c(.)$ as :

$$\exists\alpha_0 > 0; \forall h > 0; \frac{\alpha_0 \mathrm{sh}(\alpha_0 h)}{\mathrm{ch}(\alpha_0 h) - 1} < -\frac{s'(h)}{s(h)} \tag{3.15}$$

An equivalent result to lemma 3.8 can then be written for function $c(.)$. The continuity assumption on $\varphi(.)$ translates into a continuous second derivative $c"(.)$.

We can now prove that the set of distributions satisfying theorem 3.1 is not empty because the distribution function that it defines is a proper joint distribution function. We summarize all the results and state that, given a function $c(.)$, theorem 3.1 defines a proper joint distribution function.

Theorem 3.9. *Let $\mathcal{C}$ be the set of functions from $\mathbb{R}$ to $\mathbb{R}$ such that :*

i. $c(h)$ is a decreasing function from $+\infty$ to 0 and it is twice differentiable.

ii. $c(0) = 0; c'(0) < 0$.

iii. $c(.)$ verifies (3.6), (3.10) and (3.15)

and let $F \in \mathcal{F}(c)$ the non empty set of functions defined in proposition 3.7.

Let $c(.) \in \mathcal{C}$. Then (3.2) defines a joint distribution function that verifies A1 and (3.1).

Proof. The necessary and sufficient condition (3.5) is verified by using proposition 3.7. ■

In the next section, we will consider different examples for function $c(.)$ or its equivalent representation the distribution function $\phi(.)$. We can however reconsider now the example of conditional logit.

3.4. Conditional logit

In this case, $c(h) = \exp(-h)$. It is a decreasing function from $+\infty$ to 0, twice differentiable, equal to 1 when $h = 0$ and such that $c'(0) = -1 < 0$. Its associated function $s(h)$ is given by:

$$s(h) = \frac{c'}{(1-c)^2} = -\frac{2\exp(-h)}{(1-\exp(-h))^2}$$

Using the expression of the density:

$$\varphi(h) = \frac{d^2}{dh^2}\frac{h}{1-c} = 2s(h) + hs'(h)$$

one gets:

$$\varphi(h) = \frac{\exp(-h)\left(h(1+\exp(-h)) - 2(1-\exp(-h)\right)}{(1-\exp(-h))^3}$$

which can be shown to be symmetric and positive everywhere and such that $\varphi(0) = 1/6$. It also verifies (3.10). More interestingly, the last condition (3.15) can be written as:

$$\exists\alpha_0; \forall h > 0; \frac{\alpha_0 \mathrm{sh}(\alpha_0 h)}{\mathrm{ch}(\alpha_0 h) - 1} < -\frac{s'(h)}{s(h)} = \frac{1+\exp(-h)}{1-\exp(-h)}$$
$$\Leftrightarrow \exists\alpha_0; \forall h > 0; \alpha_0\frac{1+\exp(-\alpha_0 h)}{1-\exp(-\alpha_0 h)} < \frac{1+\exp(-h)}{1-\exp(-h)}$$

which is verified for any $\alpha_0 < 1$. As the expression on the RHS is increasing with α_0, this expression is verified for any $\alpha_0 < 1$.

It can then be verified that if $F(.)$ is a logistic distribution function:

$$\frac{f"(x)}{f(x)} = \frac{\exp(-2x) - 4\exp(-x) + 1}{\exp(-2x) + 2\exp(-x) + 1} < 1$$

and the pair $(c(.), F(.))$ gives a proper joint distribution function as expected. It is however far to be the only marginal distribution function which can be associated to this function $c(.)$. Any function such that $\frac{f"(x)}{f(x)} < 1$ would qualify.

More generally now, one may wonder whether the conditions on $c(.)$ and $F(.)$ bound the correlation between the disturbances u_1 and u_2. It is what we now study.

3.5. Bounds on the correlation

We shall from now on assume that all moments of the distributions exist. In that case we use :

$$V(u_1 - u_2) = 2(V(u_1) - Cov(u_1, u_2))$$

since u_1 and u_2 have the same marginal d.f. F. The coefficient of correlation is therefore bounded by:

$$\rho \geq 1 - \frac{V(u_1 - u_2)}{2Vu_1}$$

We consider bounds on $V(u_1 - u_2)$ and Vu_1. The latter bound is the most straightforward to derive. Assuming that the following integration by parts is legitimate, we have :

$$\begin{aligned} 1 &= \int_{-\infty}^{+\infty} f(x)dx = [xf(x)]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} xf'(x)dx = -\int_{-\infty}^{+\infty} xf'(x)dx \\ &= -\left[x^2 f'(x)/2\right]_{-\infty}^{+\infty} + \int_{-\infty}^{+\infty} \frac{x^2}{2} f"(x)dx = \int_{-\infty}^{+\infty} \frac{x^2}{2} f"(x)dx \end{aligned}$$

Using the inequality restriction (3.14) yields :

$$1 \leq \int_{-\infty}^{+\infty} \frac{x^2}{2} \alpha_0^2 f(x)dx = \frac{\alpha_0^2}{2} V(u_1)$$

As parameter α_0 is given by condition (3.13), it also affects $V(u_1 - u_2)$. This relationship is bounded by the following expression.

Lemma 3.10.

$$V(u_1 - u_2) \leq \frac{2\pi^2}{3\alpha} \int_0^{+\infty} \tau\varphi(\tau)d\tau$$

Proof. See appendix. ■

The coefficient of correlation is therefore bounded by :

$$\rho \geq 1 - \frac{\pi^2 \alpha}{6} \int_0^{+\infty} \tau\varphi(\tau)d\tau$$

Considering that $\alpha^2 \leq \frac{6\varphi(0)}{\int_0^{+\infty} \tau\varphi(\tau)d\tau}$ as shown in lemma F.1 (see appendix F), we have therefore that :

$$\rho \geq 1 - \pi^2 \left(\varphi(0) \int_0^{+\infty} \tau\varphi(\tau)d\tau\right)^{1/2}$$

In the examples that will be presented in the next section, this bound is always less than -1.

4. Estimation and parametric examples

We now use these results to derive the estimating equations and give some parametric examples where we simply use the property of quasi-differencing the binary variables as if they were continuous variables.

4.1. Estimation principle

Since $x_1 - x_2 = (z_1 - z_2)\beta$, an estimating condition is given by using (2.1) and (2.2):

$$\Pr(y_1 = 1, y_2 = 0 \mid z_1, z_2, \sum_{t=1}^{2} y_t = 1) = \frac{c((z_1 - z_2)\beta)}{1 + c((z_1 - z_2)\beta)} \tag{4.1}$$

On the restricted sample, such that $\sum_{t=1}^{2} y_t = 1$ we then also have that:

$$E(y_1 - y_2 \mid z_1, z_2, \sum_{t=1}^{2} y_t = 1) = G((z_1 - z_2)\beta) \tag{4.2}$$

where $G = \frac{c-1}{1+c}$ is a monotonic function taking values between -1 and $+1$, $G(0) = 0$.

Equation (4.1) gives a way to estimate parameters β by maximum likelihood and equation (4.2) by the method of moments. It should be noted that we are back to the static case because the dependent variable can only take two values 1 and -1. Standard binary methods therefore can be used provided that the conditions developed in the previous section are satisfied.

As far as identification in the semiparametric case is concerned, the normalization restriction $c(0) = 1$ implies that the distribution function given by (4.1) is such that its median is equal to 1/2. By analogy, the other identification conditions are given in Manski, 1988 or Horowitz, 1998. There is at least one continuously distributed regressor and by normalization its coefficient is set equal to 1. It simply implies that function $c(h)$ is rescaled into $c(\lambda h)$.

For estimation purposes, it is also easy to translate the different methods that have been proposed in the literature and that we already quoted, and in particular the efficient estimator developed by Klein and Spady (1993). It should be remarked that the conditions imposed on $c(.)$ do not affect the results developed in that paper.

Finally, for extensions to many periods, we can use the principle of differencing across any two periods following the idea of Manski (1987).

4.2. Parametric examples

The simplest parametric example is the standard case of logistic distributions. In that case function $c(h) = \exp(-h)$ and the distribution function given by (4.1) is the logistic distribution. There are two routes to depart from this assumption. The first route is to use standard distributions in (4.1), for instance the normal d.f. that we call Conditional Probit (Example 1). The second route is to specify the distribution function for the difference between u_1 and u_2. We shall study logistic differences (Example 2) and normal differences (Example 3). In each case, we shall show that the conditions presented in the previous section are satisfied.

4.2.1. Conditional Probit

Assume that in equation (4.1):

$$\frac{c(h)}{1+c(h)} = \Phi_0(-h) \text{ or } c(h) = \frac{\Phi_0(-h)}{\Phi_0(h)}$$

where Φ_0 is the zero-mean unit-variance normal d.f. Function $c(.)$ is a strictly decreasing function which is twice differentiable and such that $c(0) = 1$. Its derivative is given by:

$$c'(h) = -\frac{\varphi_0(h)}{(\Phi_0(h))^2} \qquad c'(0) = -\frac{4}{\sqrt{2\pi}} < 0$$

where $\varphi_0(h)$ is the normal density. Conditions (3.6) translate into first :

$$\lim_{h \to +/-\infty} \frac{\varphi_0(h)}{(2\Phi_0(h)-1)^2} h = 0$$

which is satisfied and into :

$$0 < \varphi_c(h) = \frac{\partial^2}{\partial h^2} \frac{h\Phi_0(h)}{2\Phi_0(h)-1} < +\infty$$

where $\varphi_c(h)$ is the density function of the difference between u_1 and u_2. It is straightforward to show that :

$$\varphi_c(h) = \frac{4\varphi_0(h)}{(2\Phi_0(h)-1)^3} \left((\frac{h^2}{2} - 1)(\Phi_0(h) - 1/2) + h\varphi_0(h) \right) \qquad (4.3)$$

We just consider the case of $h > 0$ as the d.f. is symmetric. When h tends to zero,

$$\Phi_0(h) - 1/2 \sim h\varphi_0(0) - (h^3/6)\varphi_0(0)$$

$$\varphi_0(h) \sim \varphi_0(0) - (h^2/2)\varphi_0(0)$$

the argument between brackets is equivalent to:

$$(\frac{h^2}{2} - 1)(\Phi_0(h) - 1/2) + h\varphi_0(h) \sim (h^3/6)\varphi_0(0)$$

and therefore is positive. Then, if h tends to zero :

$$\varphi_c(h) \sim \frac{4h^3\varphi_0^2(0)/6}{8h^3\varphi_0^3(0)} = \frac{1}{12\varphi_0(0)}$$

which is bounded and positive.

The derivative of the argument between brackets in (4.3) is :

$$h(\Phi_0(h) - 1/2) - \frac{h^2\varphi_0(h)}{2}$$

As :

$$\Phi_0(h) - 1/2 = \int_0^h \varphi_0(\tau)d\tau \geq h\varphi_0(h)$$

the derivative is always positive. $\varphi_c(h)$ is therefore positive and bounded because of the exponential term in $\varphi_0(h)$ appearing in (4.3).

Finally, condition (3.9) can be written as:

$$\exists\beta_0; \lim_{h\to+\infty} \frac{h\varphi_c(h)}{\int_h^{+\infty} \tau\varphi_c(\tau)d\tau} > \beta_0$$

As shown in the appendix :

$$\int_h^{+\infty} \tau\varphi_c(\tau)d\tau = -h^2\frac{c'}{(1-c)^2}$$

and therefore:

$$\frac{h\varphi_c(h)}{\int_h^{+\infty} \tau\varphi_c(\tau)d\tau} = \frac{4h}{\Phi_0(h) - 1/2}\left((\frac{h^2}{2} - 1)(\Phi_0(h) - 1/2) + h\varphi_0(h)\right)$$

which tends to $+\infty$ when h tends to $+\infty$. As $\varphi_c(.)$ satisfies (3.9) and is continuous, we can apply lemma 3.8 to prove (3.13).

All the conditions in theorem 3.9 are therefore proven and conditional Probit is a valid method. In conclusion, one can wonder if the correlation is restricted in that case. As $\varphi_c(0) = \frac{1}{12\varphi_0(0)}$ and :

$$\int_0^{+\infty} \tau\varphi_c(\tau)d\tau = \lim_{h\mapsto 0} \frac{\varphi_0(h)h^2}{(2\Phi_0(h)-1)^2} = \frac{1}{4\varphi_0(0)}$$

the correlation is bounded by :

$$\rho \geq 1 - \pi^2 \left(\frac{1}{48\varphi_0^2(0)}\right)^{1/2} = 1 - \frac{\pi^2}{4\sqrt{3}}(2\pi)^{1/2} = -2.57$$

and the bound is not limiting.

4.2.2. Normal differences

Assume now that $u_1 - u_2$ is normally distributed of variance equal to 1. Then $\int_h^{+\infty} \tau\varphi_0(\tau)d\tau = \frac{1}{(2\pi)^{1/2}}\exp(-\frac{h^2}{2})$ and the corresponding function $c(.)$ is :

$$c(h) = 1 - \frac{h}{h\Phi_0(h) + \exp(-\frac{h^2}{2})/\sqrt{2\pi}}$$

Conditions are easier to verify. Condition (3.9) is satisfied since:

$$\frac{h\varphi_0(h)}{\int_h^{+\infty} \tau\varphi_0(\tau)d\tau} = h$$

and as $\varphi_0(.)$ is continuous, we can apply lemma 3.8 to get condition (3.13).

Concerning the correlation, we get:

$$\rho \geq 1 - \pi^2 \left(\varphi_0^2(0)\right)^{1/2} = 1 - \frac{\pi^{3/2}}{\sqrt{2}} = -2.93$$

and the bound is not limiting.

4.2.3. Logistic differences

Assume finally that $u_1 - u_2$ is logistically distributed of variance equal to 1. Then:

$$\int_h^{+\infty} \tau\varphi_0(\tau)d\tau = \int_h^{+\infty} \frac{\tau\exp(-\tau)}{(1+\exp(-\tau))^2}d\tau$$

$$= \left[\tau(\frac{1}{1+\exp(-\tau)} - 1)\right]_h^{\infty} + \int_h^{+\infty} \frac{\exp(-\tau)}{1+\exp(-\tau)}d\tau$$

$$= h\frac{\exp(-h)}{1+\exp(-h)} + \log(1+\exp(-h))$$

and the corresponding function $c(.)$ is :

$$c(h) = 1 - \frac{h}{h + \log(1 + \exp(-h))} = 1 - \frac{h}{\log(1 + \exp(h))}$$

Concerning the correlation, we get:

$$\rho \geq 1 - \pi^2 \left(\frac{\log(2)}{4}\right)^{1/2} = -3.10$$

and the bound is not limiting.

5. An application

We now present an illustration of this estimation method on the prototypical binary model since we consider female participation in the labor market. We use data from the French Labour Force Survey in 1998 and 1999. The LFS is a rolling panel. Our working sample gathers married women present in 1998 and 1999 and aged from 25 to 55. Partners are employed and partners' income is reported in both years. The sample size is equal to 11296. Participation patterns are presented in table 1 where participation means employment or searching for a job. Only 5.8% of the sample change participation status between 1998 and 1999 of whom 54% move from non participation into participation. Therefore the sample size of movers, that we use for conditional likelihood estimation, is reduced to 651. Explanatory variables are the number of children in different age ranges (0-3, 4-6, 7-18), spouse income and a quadratic polynomial in age.

Results of conditional likelihood estimation for different specifications are reported in table 2. In order to facilitate the comparison across specifications, we normalized to 1 the coefficient of the number of children in the 0-3 age range. It was the most precisely estimated coefficient in all specifications. Variables are centered, by subtracting to them their average values, and the coefficient of the intercept can be interpreted as the macro effect of year 99 with respect to year 98 – a period of decreasing unemployment.

Different specifications lead to very similar estimates, in particular, when symmetric distributions are used. Whether it be conditional Logit or Probit, or Normal Differences – the distribution function of $u_1 - u_2$ is normal – estimates of all coefficients are almost equal and likelihood values are not significantly different. In this sense, Conditional Logit estimates are robust in this sample.

We also tried the method of estimation of Klein and Spady (1993). Because of results appearing in table 2, we considered that the necessary continuously distributed variable was Age (in months) – because the effect of Spouse income was never significant. The method did not converge and a detailed investigation of this problem showed that the model was not semi-parametrically identified. The effect of the continuous variable, Age, is much smaller, in terms of range and strength, than the effect of the children variables. The latter are therefore not identified (see Horowitz, 1998 for an example). Some parametric assumption is needed in the present sample and we therefore stick to the results presented in table 2.

It is also of some interest to compare the conditional likelihood results of this type of estimation in cases of Logit and Probit using normally distributed individual effects. We also present results when time-averages of the explanatory variables are included as explanatory variables in order to control for the possible dependence between individual effects and explanatory variables (Arellano and Honoré, 1999). Overall, Logit and Probit estimates are very close as they were in the conditional estimation. The correlation across time between random terms is very large (.98 using Logit, .94 using Probit) because persistence is high and changes between states are few. Differences between results when controling or not for time averages of variables are noticeable and the estimated coefficients of time averages are very significant. The control procedure has no impact on the estimates of the correlation coefficients, a moderate impact for children variables and spouse income and a large impact for the age variable. In particular, it does not seem to be possible to reconcile the estimated coefficient of age, using fixed effects, table 2, with the estimated coefficient of age^2, using random effects, table 3. Age and individual effects are unsurprisingly very much related on such a short panel.

The random effect estimation performs quite well with respect to the children variables when we compare estimates with the fixed effect results. We can test, by an Hausman procedure, that the estimated relative effects of children (4-6, 7-

18) are equal in the random effects Logit case – the efficient one under the null hypothesis – and the conditional Logit case – the robust one against dependence between individual effects and covariates and normality of individual effects. The test statistic is equal to 5.22 with two degrees of freedom. At a level of 5%, we cannot reject that coefficients are equal.

6. Extensions

Other questions of interest are whether the implied specification for the joint distribution function can be tested and whether the other primitive, the marginal distribution function of both disturbances, can be recovered from the data. Another question is how to extend the results of Chamberlain (1992) to this setting.

6.1. Specification tests

As before, we write the indices $x_t = z_t\beta$ where z_t are the exogenous covariates at period $t = 1, 2$. We also assume that parameters β and function $c(.)$ are identified up to a scale parameter under the identification conditions given in Manski (1988). In particular, x_t is a continuous variable because one variable among z_t is continuous (say $z_t^{(1)}$). What is identified is $\dfrac{x_t}{\beta^{(1)}} = z_t^{(1)} + z_t^{(-1)}\dfrac{\beta}{\beta^{(1)}}$ where $z_t^{(-1)}$ stands for other z_t variables.

As y_1 and y_2 are binary variables, its joint distribution function conditional to z_1 and z_2 is completely defined by the three following functions:

$$\left\{\begin{array}{rc} \Pr\{y_1 = 0, y_2 = 0 \mid z_1, z_2\} = & E_\varepsilon \Pr\{y_1 = 0, y_2 = 0 \mid z_1, z_2, \varepsilon\} \\ \Pr\{y_1 = 0 \mid z_1, z_2\} = & E_\varepsilon \Pr\{y_1 = 0 \mid z_1, z_2, \varepsilon\} \\ \Pr\{y_2 = 0 \mid z_1, z_2\} = & E_\varepsilon \Pr\{y_2 = 0 \mid z_1, z_2, \varepsilon\} \end{array}\right.$$

where, in the RHS, ε is integrated out with respect to the conditional distribution function of individual effects, $dG(\varepsilon \mid z_1, z_2)$ say. Using (3.2) and that the marginal distribution functions of u_1 and u_2 are equal to $F(.)$, we can rewrite this system as:

$$\left\{\begin{array}{rc} \Pr\{y_1 = 0, y_2 = 0 \mid z_1, z_2\} = & \dfrac{E_\varepsilon(F(x_2+\varepsilon) \mid z_1, z_2) - c(x_1 - x_2).E_\varepsilon(F(x_1+\varepsilon) \mid z_1, z_2)}{1 - c(x_1 - x_2)} \\ \Pr\{y_1 = 0 \mid z_1, z_2\} = & E_\varepsilon(F(x_1+\varepsilon) \mid z_1, z_2) \\ \Pr\{y_2 = 0 \mid z_1, z_2\} = & E_\varepsilon(F(x_2+\varepsilon) \mid z_1, z_2) \end{array}\right. \tag{6.1}$$

The last two equations express the marginal distribution functions which are directly identifiable from the data, as convolutions of function $F(.)$ and function $G(\varepsilon \mid z_1, z_2)$:

$$\int F(x_t + \varepsilon) dG(\varepsilon \mid z_1, z_2)$$

the integral being taken over the support of ε. These equations are convolution equations which properties relative to identification, remain to be sought. We can however rewrite these integrals as:

$$\int \tilde{F}(\frac{x_t + \varepsilon}{\beta^{(1)}}) d\tilde{G}(\frac{\varepsilon}{\beta^{(1)}} \mid z_1, z_2) = \int \tilde{F}(z_t^{(1)} + z_t^{(-1)} \frac{\beta}{\beta^{(1)}} + \eta) d\tilde{G}(\eta \mid z_1, z_2)$$

and the last two equations of (6.1) can be rewritten accordingly. It implies constraints, for instance, that for any z_1, z_2 such that:

$$z_1^{(1)} + z_1^{(-1)} \frac{\beta}{\beta^{(1)}} = z_2^{(1)} + z_2^{(-1)} \frac{\beta}{\beta^{(1)}}$$

we have:

$$\Pr\{y_1 = 0 \mid z_1, z_2\} = \Pr\{y_2 = 0 \mid z_1, z_2\}$$

Another possible way of testing the global cut property is to reconsider the first equation of system (6.1) and replace the expressions in the first equation by their value given by the two last equations to obtain:

$$\Pr\{y_1 = 0, y_2 = 0 \mid z_1, z_2\} = \frac{\Pr\{y_2 = 0 \mid z_1, z_2\} - c(x_1 - x_2). \Pr\{y_1 = 0 \mid z_1, z_2\}}{1 - c(x_1 - x_2)}$$

As all these expressions can be non parametrically identified, this condition yields a moment condition that can be tested.

6.2. Convergence rate

We can easily generalize one of the results of Chamberlain (1992) to this setting. The semiparametric efficiency bound of the parametric part of the model is zero unless the sum of the binary variables is a sufficient statistic.

First, we use the following theorem adapted from the one stated at page 7, Chamberlain (1992). Define first the vector of probabilities:

$$a(z, c, \theta_0) = \begin{pmatrix} \Pr(u_1 \leq -(z_1\beta + c), u_2 \leq -(z_2\beta + c)) \\ \Pr(u_1 > -(z_1\beta + c), u_2 \leq -(z_2\beta + c)) \\ \Pr(u_1 \leq -(z_1\beta + c), u_2 > -(z_2\beta + c)) \\ \Pr(u_1 > -(z_1\beta + c), u_2 > -(z_2\beta + c)) \end{pmatrix}$$

to be able to write:

Theorem 6.1. *The semi-parametric efficiency bound $I_\Lambda = 0$ for all θ_0 in Θ unless the distribution function of random shocks is such that:*

$$\begin{aligned} \forall z_1, z_2; \exists \psi &= (\psi_1, \psi_2, \psi_3, \psi_4) \in \mathbb{R}^4 \text{ such that:} \\ \forall c &\in \mathbb{R}, \ \psi' a(z, c, \theta_0) = 0 \end{aligned} \tag{6.2}$$

Proof: See Chamberlain (1992) adapting the proof to the case where u_1 and u_2 are not independent.

Second, it remains to be proven that this condition is equivalent to the existence of the sufficient statistic. Note first that if $c \to +\infty$ then necessarily $\psi_4 = 0$ and that if $c \to -\infty$ then necessarily $\psi_1 = 0$ (Chamberlain, 1992). Therefore condition (6.2) is equivalent to:

$$\psi_2 \Pr(u_1 > -(z_1\beta+c), u_2 \leq -(z_2\beta+c)) + \psi_3 \Pr(u_1 \leq -(z_1\beta+c), u_2 > -(z_2\beta+c)) = 0$$

which is equivalent to:

$$\frac{\Pr(u_1 > -(z_1\beta + c), u_2 \leq -(z_2\beta + c))}{\Pr(u_1 \leq -(z_1\beta + c), u_2 > -(z_2\beta + c))} = -\frac{\psi_3}{\psi_2}$$

where this ratio is independent of c. This condition is exactly condition (3.1). Therefore:

Theorem 6.2. *The semi-parametric efficiency bound is equal to zero unless the sum of binary variables is a sufficient statistic for the individual effect.*

As the principle of the proof of this result is very similar to the other result about identification when regressors are bounded, the same extended type of result can be derived.

7. Conclusion

In this paper, we used the property of sufficiency to derive a generalization to conditional logit methods. We presented the conditions under which we can quasi-difference binary data. There are various possible extensions. In particular, it remains to be seen how such an approach would be applied other non-linear models.

References

Abrevaya, J., 2000, "Rank Estimation of a Generalized Fixed Effects Regression Model", *Journal of Econometrics*, 95:1-23.

Andersen, P.K., 1973, *Conditional Inference and Models for Measuring,* Mentalhygienjnisk Forlag:Copenhague.

Arellano, M. and B. Honoré, 1999, "Panel Data Models: Some Recent Developments", in eds E. LEamer and J.J. Heckman, *Handbook of Econometrics*, 5, North Holland: Amsterdam.

Barndorff-Nielsen, O.E., 1978, *Information and Exponential Families in Statistical Theory,* Chichester:Wiley.

Barndorff-Nielsen, O.E., and D., R., Cox, 1994, *Inference and Asymptotics,* Chapman and Hall: London.

Chamberlain, **G.**, 1984, "Panel data", in eds Griliches and Intriligator, *Handbook of Econometrics*, volume 2,

Chamberlain, G., 1992, "Binary Response Models for Panel Data : Identification and Information", Harvard Univ., unpublished manuscript.

Charlier, E.B., B. Melenberg, and A. Van Soest, 1995, "A Smoothed Maximum Score Estimator for the Binary Choice Panel Data Model and an Application to Labour Force Participation", *Statistica Neerlandica*, 49:324-42.

Chen, X., J.J., Heckman, E., Vytlacil, 1999, "Identification and $\sqrt{N}$ Efficient Estimation of Semiparametric Panel Data Models with Binary Dependent Variables and a Latent Factor", unpublished manuscript.

Cox, D.R. and N., Reid, 1987, "Parameter Orthogonality and Approximate Conditional Inference", *Journal of Royal Statistical Society Series B*, 49:1:39.

Gradshteyn, I.S., and I.M., Ryzhik, 1995, *Tables of Integrals, Series and Products*, Academic Press: London.

Honoré, B.E. and E., Kyriazidou, 2000, "Panel Data Discrete Choice Models with Lagged Dependent Variables", *Econometrica*,

Honoré, B.E. and A. Lewbel, 2000, "Semiparametric Binary Choice Panel Data Models without Strictly Exogeneous Regressors", WP, Boston College.

Horowitz, J., 1992, "A Smoothed Maximum Score Estimator for the Binary Response Model", *Econometrica*, 60:505-31

Horowitz, J., 1998, *Semiparametric methods in Econometrics*, Springer: Berlin.

Klein, R.W. and R.H. Spady, 1993, "An Efficient Semiparametric Estimator for Binary Response Models", *Econometrica*, 61:387-421.

Lancaster, T., 1999, "Panel Binary Choice with Fixed Effects", 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", Brown University.

Lee, M.J., 1999, "A Root-n Consistent Semiparametric Estimator for Related-Effect Binary Response Panel Data", *Econometrica*, 67:427-33.

Lewbel, A., 2000, "Semiparametric Qualitative Response Model Estimation with Unknown Heteroskedasticity or Instrumental Variables", *Journal of Econometrics*, 97:145-77.

Manski, C.F., 1987, "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data", *Econometrica*, 55:357-62.

Manski, C.F., 1988, "Identification of Binary Response Models", *Journal of the American Statistical Association,* 83:729-738.

Rasch, G., 1960, Probabilistic Models for Some Intelligence and Attainment Tests", Denmark Paedadagogiske Institut, Copenhaguen.

A. Proof of theorem 3.1

The proof proceeds by reparametrizing the problem as:

$$x_1 = \frac{\Delta + h}{2}, x_2 = \frac{\Delta - h}{2}$$

Let:

$$\begin{aligned} K(h,\Delta) &= P(u_1 > \frac{\Delta+h}{2}, u_2 \le \frac{\Delta-h}{2}) \\ G(h,\Delta) &= P(u_1 \le \frac{\Delta+h}{2}, u_2 > \frac{\Delta-h}{2}) \\ M(h,\Delta) &= P(u_1 \le \frac{\Delta+h}{2}, u_2 \le \frac{\Delta-h}{2}) \end{aligned}$$

By the main property (3.1):

$$K(h,\Delta) = c(h)G(h,\Delta)$$

and by construction :

$$K(h,\Delta) = P(u_2 \le \frac{\Delta-h}{2}) - M(h,\Delta)$$

$$G(h,\Delta) = P(u_1 \le \frac{\Delta+h}{2}) - M(h,\Delta) \tag{A.1}$$

Therefore :

$$\begin{aligned} L(h,\Delta) &= G(h,\Delta) - K(h,\Delta) \\ &= P(u_1 \le \frac{\Delta+h}{2}) - P(u_2 \le \frac{\Delta-h}{2}) \\ &= (1-c(h))G(h,\Delta) \end{aligned}$$

By normalization, $c(0) = 1$, and thus :

$$\forall\Delta; P(u_2 \le \frac{\Delta}{2}) = P(u_1 \le \frac{\Delta}{2}) = F(\frac{\Delta}{2})$$

and :

$$G(h,\Delta) = \frac{F(\frac{\Delta+h}{2}) - F(\frac{\Delta-h}{2})}{1-c(h)}$$

Using (A.1) :

$$M(h,\Delta) = \frac{F(\frac{\Delta-h}{2}) - c(h)F(\frac{\Delta+h}{2})}{1-c(h)}$$

which gives (3.2). As $c(h)$ is a decreasing function from $+\infty$ to 0, it is straightforward to verify the limit conditions when either x_1 or x_2 tend to $+/-\infty$.■

B. Proof of proposition 3.3

The joint density function must be defined everywhere, be positive and bounded. henceforth :

$$\forall(x_1, x_2); 0 < \frac{\partial^2}{\partial x_1 \partial x_2} \Pr(u_1 \leq x_1, u_2 \leq x_2) < \infty$$

Using (3.2):

$$\frac{\partial}{\partial x_2} \Pr(u_1 \leq x_1, u_2 \leq x_2) = \frac{f(x_2)}{(1 - c(h))} + \frac{c'(h)}{(1 - c(h))^2}(F(x_1) - F(x_2))$$

where $h = x_1 - x_2$ and $c'(h) = \frac{dc(h)}{dh}$. Therefore :

$$\frac{\partial^2}{\partial x_1 \partial x_2} \Pr(u_1 \leq x_1, u_2 \leq x_2) =$$

$$\frac{c'(h)}{(1 - c(h))^2}(f(x_2) + f(x_1)) + \frac{\partial}{\partial h}\left[\frac{c'(h)}{(1 - c(h))^2}\right](F(x_1) - F(x_2)) \tag{B.1}$$

Letting $s(h) = \frac{c'(h)}{(1-c(h))^2}$ yields the expression$\square$

C. Proof of proposition 3.4

As the function $s(h)$ is not defined at $h = 0$, we have to distinguish the cases of $h < 0$ and $h > 0$.

Write for any $h < 0$:

$$\Pr(u_1 - u_2 \leq h) = \int_{x_1 - x_2 \leq h} \frac{\partial^2}{\partial x_1 \partial x_2} \Pr(u_1 \leq x_1, u_2 \leq x_2) dx_1 dx_2$$

Using (B.1), the argument under the integral sign is :

$$s(h)(f(x_2) + f(x_1)) + s'(h)(F(x_1) - F(x_2))$$

Letting $\tau = x_1 - x_2$, we get :

$$\Pr(u_1 - u_2 \leq h) = \int_{-\infty}^{+\infty} \int_{\tau \leq h} \left[s(\tau)(f(x_2) + f(x_2 + \tau)) + s'(\tau)(F(x_2 + \tau) - F(x_2))\right] dx_2 d\tau$$

Consider :

$$\int_{\tau \leq h} s(\tau) d\tau = \frac{1}{1 - c(h)} \qquad \int_{\tau \leq h} s'(\tau) d\tau = s(h)$$

because $\lim_{h \to -\infty} \frac{1}{1-c(h)} = 0$ and **if** $\lim_{h \to -\infty} s(h) = 0$ (Condition T). If not the integral diverges. Also:

$$\int_{\tau \leq h} \left[s(\tau) f(x_2 + \tau) + s'(\tau) F(x_2 + \tau)\right] d\tau = s(h) F(x_2 + h)$$

we get:

$$\Pr(u_1 - u_2 \leq h) = \int_{-\infty}^{+\infty} \left[\frac{1}{1-c(h)} f(x) + s(h)(F(x+h) - F(x)) \right] dx$$

Then:

$$\Pr(u_1 - u_2 \leq h) = \frac{1}{1-c(h)} + s(h) \int_{-\infty}^{+\infty} [F(x+h) - F(x)]\, dx$$

As f" exists and is bounded (see theorem 3.1) then :

$$\exists \lambda > 0 \text{ such that } F(x+h) - F(x) \leq \lambda h (f(x+h) + f(x))$$

And therefore :

$$\forall h; \left| \int_{-\infty}^{+\infty} [F(x+h) - F(x)]\, dx \right| < +\infty$$

We can therefore differentiate the integral expression with respect to h and derivation and integral can be permuted to get :

$$\frac{d}{dh} \int_{-\infty}^{+\infty} [F(x+h) - F(x)]\, dx = 1$$

As the integral takes value 0 when $h = 0$, we obtain :

$$\int_{-\infty}^{+\infty} [F(x+h) - F(x)]\, dx = h$$

Hence replacing $s(h) = \frac{c'(h)}{(1-c(h))^2}$:

$$\Pr(u_1 - u_2 \leq h) = \frac{1}{1-c} + \frac{c'}{(1-c)^2} h$$

$$= \frac{d}{dh} \frac{h}{1-c}$$

As this probability is bounded by $1 = \lim_{h \to \infty} \frac{1}{1-c}$, we necessarily have that:

$$\lim_{h \to -\infty} \frac{c'}{(1-c)^2} h = 0$$

and condition T above is therefore also satisfied.

For any $h > 0$, we can use the same argument using:

$$\Pr(u_1 - u_2 > h) = \int_{-\infty}^{+\infty} \int_{\tau > h} [s(\tau)(f(x_2) + f(x_2 + \tau)) + s'(\tau)(F(x_2 + \tau) - F(x_2))]\, dx_2 d\tau$$

and:

$$\int_{\tau > h} s(\tau) d\tau = 1 - \frac{1}{1-c(h)} \qquad \int_{\tau > h} s'(\tau) d\tau = -s(h)$$

under the condition that $\lim_{h \to +\infty} s(h) = 0$. We end up with:

$$\Pr(u_1 - u_2 > h) = 1 - \frac{d}{dh}\frac{h}{1-c}$$

and:

$$\lim_{h \to +\infty} \frac{c'}{(1-c)^2} h = 0$$

Summarizing, for any $h \neq 0$:

$$\Pr(u_1 - u_2 \leq h) = \frac{1}{1-c} + \frac{c'}{(1-c)^2} h = \frac{d}{dh}\frac{h}{1-c}$$

$$\lim_{h \to +/-\infty} \frac{c'}{(1-c)^2} h = 0$$

By extension, using $1 - c(h) \sim -c'(0)h - c"(0)h^2/2$ and $c'(h) \sim c'(0) + c"(0)h$ when $h \to 0$:

$$\Pr(u_1 - u_2 \leq 0) = \frac{c"(0)}{2c'(0)^2}$$

which implies that $0 < c"(0) < 2c'(0)^2$ and that $c"(.)$ is continuous at zero (condition T').

As $\lim_{h \to -\infty} \frac{1}{1-c} = 0$ and $\lim_{h \to \infty} \frac{1}{1-c} = 1$, by point 1 of Theorem 3.1, the expression above defines a proper cumulative d.f. which admits a positive and bounded density function iff :

$$0 < \frac{d^2}{dh^2}\frac{h}{1-c} < \infty \qquad \lim_{h \to +/-\infty} \frac{c'}{(1-c)^2} h = 0$$

because condition T' is implied by these conditions.■

D. Proof of proposition 3.5

We shall verify the conditions that $c(.)$ is a positive function that is decreasing and such that $c'(0) < 0$. Let $\phi(h)$ be the cumulative distribution function of $u_1 - u_2$, then:

$$\frac{h}{1-c} = \int_0^h \phi(\tau) d\tau + A$$

As $\frac{h}{1-c}$ is a function taking value equal to $-\frac{1}{c'(0)} > 0$ when $h = 0$, $A = -\frac{1}{c'(0)} > 0$. Given that functions $\frac{h}{1-c}$ and $\frac{hc}{1-c}$ are positive functions then:

$$\forall h; \int_0^h \phi(\tau) d\tau + A > 0 \qquad \int_0^h \phi(\tau) d\tau + A - h > 0$$

The second condition can be written as :

$$\int_0^h (\phi(\tau) - 1) d\tau + A > 0$$

Taking the maximum over h yields:

$$\int_{-\infty}^{0} \phi(\tau)d\tau < A \qquad \int_{0}^{\infty} (1-\phi(\tau))d\tau < A \tag{D.1}$$

which are also necessary and sufficient conditions for $c(.)$ to be positive as shown by:

$$c(h) = 1 - \frac{h}{\int_0^h \phi(\tau)d\tau + A}$$

To determine A, we have to use condition (3.5). To determine function $s(h)$, take the derivative of $c(h)$:

$$c'(h) = -\frac{\int_0^h \phi(\tau)d\tau + A - h\phi(h)}{(\int_0^h \phi(\tau)d\tau + A)^2}$$

which is negative under condition (D.1).

Using:

$$\int_0^h \phi(\tau)d\tau = h\phi(h) - \int_0^h \tau\varphi(\tau)d\tau$$

where $\varphi(\tau)$ is the density of $u_1 - u_2$, yields:

$$c'(h) = -\frac{A - \int_0^h \tau\varphi(\tau)d\tau}{(\int_0^h \phi(\tau)d\tau + A)^2}$$

Note that (D.1) implies that $\int_0^h \tau\varphi(\tau)d\tau$ is bounded by A and therefore that:

$$\lim_{h\to+/-\infty} h\varphi(h) = 0 \tag{D.2}$$

We have also :

$$s(h) = \frac{c'(h)}{(1-c(h))^2} = \frac{1}{h^2}(\int_0^h \tau\varphi(\tau)d\tau - A)$$

and equation (3.5) can be written as:

$$\forall(h,x); \quad s(h)(f(x+h)+f(x)) + s'(h)(F(x+h)-F(x)) > 0$$

By the primitive assumptions, $s(h)$ is a negative function and $s'(h)$ is a negative function for $h < 0$ and a positive function for $h > 0$. Equation (3.5) implies that:

$$\forall h > 0, \forall x; \quad \frac{f(x+h)+f(x)}{F(x+h)-F(x)} < -\frac{s'(h)}{s(h)}$$

and:

$$\forall h < 0, \forall x; \quad \frac{f(x+h)+f(x)}{F(x+h)-F(x)} > -\frac{s'(h)}{s(h)}$$

Thus, it implies that:

$$\forall x; \quad 0 < \frac{f(x)}{1-F(x)} < -\lim_{h\to+\infty} \frac{s'(h)}{s(h)}$$

and:

$$\forall x; \quad 0 > -\frac{f(x)}{F(x)} > -\lim_{h\to-\infty}\frac{s'(h)}{s(h)}$$

In order for $F()$ to be positive, it is therefore necessary that:

$$\exists\beta_0 > 0; \min(-\lim_{h\to+\infty}\frac{s'(h)}{s(h)}, \lim_{h\to-\infty}\frac{s'(h)}{s(h)}) > \beta_0 \tag{D.3}$$

Using :

$$\lim_{h\to+/-\infty} -\frac{s'(h)}{s(h)} = \lim_{h\to+/-\infty}\frac{2}{h} - \frac{h\varphi(h)}{\int_0^h \tau\varphi(\tau)d\tau - A} = \lim_{h\to+/-\infty}\frac{h\varphi(h)}{A - \int_0^h \tau\varphi(\tau)d\tau}$$

it is therefore necessary because of (D.2) that:

$$A = \int_0^{+\infty}\tau\varphi(\tau)d\tau = \int_0^{-\infty}\tau\varphi(\tau)d\tau \tag{D.4}$$

and that :

$$\exists\beta_0 > 0; \lim_{h\to+/-\infty}\frac{h\varphi(h)}{\int_h^{+\infty}\tau\varphi(\tau)d\tau} > \beta_0 \tag{D.5}$$

Reconsidering (D.1), they are now equivalent to :

$$\lim_{h\to-\infty} h\phi(h) = 0 \text{ and } \lim_{h\to+\infty} h(1-\phi(h)) = 0 \tag{D.6}$$

which gives the conditions of the proposition

Replacing A by its expression, we get:

$$c(h) = 1 - \frac{h}{h\phi(h) + \int_h^{+\infty}\tau\varphi(\tau)d\tau}$$

E. Proof of corollary 3.6

Continuing the previous proof, we get:

$$s(h) = -\frac{1}{h^2}(\int_h^{+\infty}\tau\varphi(\tau)d\tau)$$

Condition (F.1) can be translated into:

$$\lim_{h\to0^+} h^2 s(h) = \lim_{h\to0^-} h^2 s(h)$$

and is always satisfied because these limits are equal to $1/c'(0)$. Condition (F.2) is translated into condition (D.3) and condition (F.3) into:

$$\lim_{h\to-\infty} h\frac{d}{dh}(\frac{h}{1-c}) = 0 \text{ and } \lim_{h\to+\infty} h(1-\frac{d}{dh}(\frac{h}{1-c})) = 0$$

■

F. Proof of proposition 3.7

Given that $s(h)$ is negative, equation (3.5) is equivalent to :

$$\forall h \geq 0, \forall x; f(x+h) + f(x) < -\frac{s'(h)}{s(h)}(F(x+h) - F(x)) \tag{F.1}$$

We must show that there exist d.f. $F(.)$ satisfying (F.1). As shown in appendix D, we use that :

$$-\frac{s'(h)}{s(h)} = \frac{2}{h} + \frac{h\varphi(h)}{\int_h^{+\infty} \tau\varphi(\tau)d\tau} \tag{F.2}$$

We proceed in four steps using the following lemmas that are proven in the following :

Lemma F.1. *When h tends to zero, (F.1) implies that :*

$$\forall x; \frac{f"(x)}{f(x)} < 6\frac{\varphi(0)}{\int_0^{+\infty} \tau\varphi(\tau)d\tau} \tag{F.3}$$

This condition suggests that we should be interested by the properties of d.f. which satisfy (F.3).

Lemma F.2. *When $\forall x; \frac{f"(x)}{f(x)} \leq \alpha^2$, then :*

$$F(x+h) - F(x) \geq \frac{ch(\alpha h) - 1}{\alpha . sh(\alpha h)}(f(x+h) + f(x))$$

where sh(.) and ch(.) are the hyperbolic sine and cosine functions

The proof is based on convexity inequalities. It is now obvious that this lemma permits to bound the functions in equation (F.1).

Lemma F.3. *Suppose condition (3.13) which we here rewrite:*

$$\exists \alpha_0 > 0; \forall h > 0; \frac{\alpha_0 sh(\alpha_0 h)}{ch(\alpha_0 h) - 1} - \frac{2}{h} < \frac{h\varphi(h)}{\int_h^{+\infty} \tau\varphi(\tau)d\tau}$$

Then if $\frac{f"(x)}{f(x)} \leq \alpha_0^2$, it satisfies equation (F.1).

By making h tend to zero, we can also prove that lemma F.3 implies lemma F.1. It is because the function on the RHS is equivalent to $\frac{\alpha^2 h}{6}$ when h tends to zero, is increasing and tends to α_0 when h tends to $+\infty$. Remark therefore that α_0 is bounded by the bound given in lemma F.1.

It remains to be proven that there exist d.f. $F(.)$.

Lemma F.4. *There exist distribution functions such that :*

$$\frac{f"(x)}{f(x)} \leq \alpha_0^2$$

The proof of proposition 3.7 terminates by considering a d.f. $F(.)$ satisfying lemma F.4. It satisfies the condition of lemma F.2. Imposing condition (3.13), this d.f. satisfies equation (F.1) by lemma F.3. Equation (F.1) is properly defined when $h = 0$ by lemma F.1 .■

F.1. Proof of lemma F.1

We can rewrite (F.1) as :

$$(f(x+h)+f(x)) - 2\frac{(F(x+h)-F(x))}{h} < \frac{h^2\varphi(h)}{\int_h^{+\infty} \tau\varphi(\tau)d\tau}\frac{(F(x+h)-F(x))}{h}$$

When h tends to zero, we can expand the LHS to the third order as $F(.)$ is supposed to be differentiable three times :

$$\frac{F(x+h)-F(x)}{h} \backsimeq f(x) + f'(x)\frac{h}{2} + f"(x)\frac{h^2}{6}$$

$$f(x+h)+f(x) \backsimeq 2f(x) + f'(x)h + f"(x)\frac{h^2}{2}$$

Therefore the LHS is equivalent when $h \to 0$ to :

$$f"(x)\frac{h^2}{6}$$

Using first order expansions, the RHS is equivalent when $h \to 0$ to :

$$h^2\frac{\varphi(0)}{\xi(0)}f(x)$$

As $f(x) > 0$, (F.1) therefore implies that :

$$\forall x; \frac{f"(x)}{f(x)} < \frac{6\varphi(0)}{\xi(0)}$$

■

F.2. Proof of lemma F.2

For any $\lambda \in [0,1]$ let :

$$\tilde{f}(\lambda) = f(\lambda(x+h)+(1-\lambda)x) > 0$$

and therefore $\tilde{f}(0) = f(x)$ and $\tilde{f}(1) = f(x+h)$. The condition $\frac{f"(x)}{f(x)} \leq \alpha^2$ implies that:

$$\frac{\tilde{f}"(\lambda)}{\tilde{f}(\lambda)} \leq \alpha^2 h^2$$

Define function $\gamma(\lambda)$ such that:

$$\gamma(0) = 0, \gamma(1) = 1$$

$$\gamma"(\lambda) - \alpha^2 h^2 \gamma(\lambda) = \alpha^2 h^2 (1 - 2\lambda) \frac{f(x)}{f(x) + f(x+h)} \tag{F.4}$$

and define function $g(\lambda)$ such that:

$$g(\lambda) = \gamma(\lambda) f(x+h) + (1 - 2\lambda + \gamma(\lambda)) f(x)$$

Then:

$$g(0) = f(x), g(1) = f(x+h)$$

$$\frac{g"(\lambda)}{g(\lambda)} = \alpha^2 h^2 \geq \frac{\tilde{f}"(\lambda)}{\tilde{f}(\lambda)}$$

As the degree convexity of $g(.)$ is "larger" than the degree of convexity of $\tilde{f}(.)$, it can be shown that:

Lemma F.5. *$\tilde{f}(\lambda) \geq g(\lambda)$ for any $\lambda \in [0, 1]$*

Proof : Let $\Psi(\lambda) = \tilde{f}(\lambda) - g(\lambda)$. We also have:

$$\Psi^{"}(\lambda) = \tilde{f}(\lambda)\left(\frac{\tilde{f}"(\lambda)}{\tilde{f}(\lambda)} - \frac{g"(\lambda)}{g(\lambda)}\right) + \frac{g"(\lambda)}{g(\lambda)}(\tilde{f}(\lambda) - g(\lambda))$$

Because $\tilde{f}(\lambda) > 0$ and the inequalities above:

$$\Psi(\lambda) \leq 0 \Rightarrow \Psi^{"}(\lambda) \leq 0$$

Assume, by contradiction, $\exists \lambda_0; \Psi(\lambda_0) < 0$. We know that $\Psi(0) = 0$, $\Psi(1) = 0$ and that $\Psi(\lambda)$ is twice differentiable. Therefore, as $\Psi(.)$ is continuous, $\exists(\lambda_1, \lambda_2)$ such that $\lambda_1 < \lambda_0 < \lambda_2$ and such that $\Psi(\lambda_1) = \Psi(\lambda_2) = 0$. Then $\forall \lambda \in]\lambda_1, \lambda_2[$, $\Psi(\lambda) < 0$ and $\Psi"(\lambda) < 0$. It is a contradiction since it is not possible to construct a concave function in a interval where it takes value 0 at the end points and is negative in between.$\blacksquare$

We therefore have :

$$F(x+h) - F(x) = \int_x^{x+h} f(u)du = \int_0^1 f(\lambda(x+h) + (1-\lambda)x) h d\lambda$$

$$= h \int_0^1 \tilde{f}(\lambda) d\lambda \geq h \int_0^1 g(\lambda) d\lambda = h\left(\int_0^1 \gamma(\lambda) d\lambda\right)(f(x) + f(x+h))$$

using the definition of $g(.)$ and $\gamma(.)$. To prove lemma F.2, we shall therefore prove that :

Lemma F.6. $(\int_0^1 \gamma(\lambda)d\lambda) = \dfrac{ch(\alpha h) - 1}{\alpha h.sh(\alpha h)}$

Proof : We integrate equation (F.4) letting $A = \dfrac{f(x)}{f(x) + f(x+h)}$:

$$\gamma"(\lambda) - \alpha^2 h^2 \gamma(\lambda) = \alpha^2 h^2 (1 - 2\lambda) A$$

As a particular solution is $\gamma(\lambda) = -(1 - 2\lambda)A$ and a general solution is :

$$\gamma(\lambda) = K_1 \exp(\alpha h \lambda) + K_0 \exp(-\alpha h \lambda)$$

the solution is :

$$\gamma(\lambda) = K_1 \exp(\alpha h \lambda) + K_0 \exp(-\alpha h \lambda) - (1 - 2\lambda)A$$

Imposing conditions $\gamma(0) = 0$ and $\gamma(1) = 1$ implies that :

$$K_1 = \frac{1 - A(1 + \exp(-\alpha h))}{2\mathrm{sh}(\alpha h)}$$

$$K_0 = \frac{A(1 + \exp(\alpha h)) - 1}{2\mathrm{sh}(\alpha h)}$$

Then integrating $\gamma(\lambda)$ between 0 and 1 yields the result.■

F.3. Proof of lemma F.3

If $\frac{f"(x)}{f(x)} \leq \alpha_0^2$, we can use lemma F.2, condition (3.13) and equation (F.2) to get equation (F.1).

F.4. Proof of lemma F.4

Let $\alpha > 0$. Consider a d.f. such that its density function is :

$$f(x) \propto \alpha^2 \exp(-\alpha^2 \mid x \mid) \text{ if } x \geq B \text{ or } x \leq -B$$

and where anywhere else the density function is concave. Overall we shall impose that f is twice differentiable by smooth pasting at B and $-B$ (f" have discontinuities at B and $-B$ however). Then everywhere f satisfies the condition of the lemma. It is a conjecture that if $f(x)$ is a Laplace density (hence $B = 0$), the whole setting applies.■

G. Proof of lemma 3.8

We first prove the following useful lemma.

Lemma G.1. *Let for any $y > 0$*

$$\Psi(y) = \frac{shy}{y(chy-1)} - \frac{2}{y^2}$$

Then for any $y > 0$, $\Psi(y) \leq 1/6$

Proof : We prove first that, by extension, $\Psi(0) = 1/6$, second that $\Psi'(y) \leq 0$.

i). When $y \to 0$, we can replace hyperbolic functions by their expansions:

$$shy \approx y + y^3/6 \qquad chy \approx 1 + y^2/2 + y^4/24$$

Then :

$$\Psi(y) \approx \frac{y(y + y^3/6) - 2(y^2/2 + y^4/24)}{y^2 y^2/2} = \frac{1}{6}$$

and therefore :

$$\Psi(0) = \lim_{y \to 0} \Psi(y) = \frac{1}{6}$$

ii). We have, using $\frac{d}{dy}\frac{\mathrm{sh}y}{(\mathrm{ch}y-1)} = -\frac{1}{\mathrm{ch}y-1}$:

$$\begin{aligned}\Psi'(y) &= -\frac{\mathrm{sh}y}{y^2(\mathrm{ch}y-1)} - \frac{1}{y(\mathrm{ch}y-1)} + \frac{4}{y^3}\\ &= \frac{-(\mathrm{sh}y+y)}{y^2(\mathrm{ch}y-1)} + \frac{4}{y^3}\end{aligned}$$

We use that for any $y > 0$:

$$\left\{\begin{array}{rl}\mathrm{sh}y \geq & y\\ \mathrm{ch}y - 1 \geq & y^2/2\end{array}\right.$$

$$\Rightarrow \left\{\begin{array}{rl}-(\mathrm{sh}y+y) \leq & -2y\\ \dfrac{1}{y^2(\mathrm{ch}y-1)} \leq & 2/y^4\end{array}\right. \Rightarrow \frac{-(\mathrm{sh}y+y)}{y^2(\mathrm{ch}y-1)} \leq -\frac{4}{y^3}$$

and therefore $\Psi'(y) \leq 0$■

Returning to the proof of lemma 3.8, we consider a density function φ which verifies (3.9) that is :

$$\exists\beta_0 > 0, \exists A > 0 \text{ such that } \forall h > A; \frac{h\varphi(h)}{\int_h^{+\infty}\tau\varphi(\tau)d\tau} > \beta_0$$

Condition (3.13) is written as:

$$\exists\alpha_0 > 0; \forall h > 0; \frac{\alpha_0\mathrm{sh}(\alpha_0 h)}{\mathrm{ch}(\alpha_0 h) - 1} - \frac{2}{h} < \frac{h\varphi(h)}{\int_h^{+\infty}\tau\varphi(\tau)d\tau}$$

Consider first the case of $h > A$. We use two results: i). the limit when $h \to 0$ of $\frac{\alpha_0 \mathrm{sh}(\alpha_0 h)}{\mathrm{ch}(\alpha_0 h)-1} - \frac{2}{h}$ is equal to α_0 and ii):

$$\frac{\partial}{\partial h}\left(\frac{\alpha_0 \mathrm{sh}(\alpha_0 h)}{\mathrm{ch}(\alpha_0 h)-1} - \frac{2}{h}\right) = -\frac{\alpha_0^2}{\mathrm{ch}(\alpha_0 h)-1} + \frac{2}{h^2} \geq 0$$

because $\mathrm{ch}(\alpha_0 h) - 1 \geq (\alpha_0 h)^2/2$. Therefore consider $\alpha_0 \leq \beta_0$ and condition (3.13) is verified for any $h > A$.

Consider now $h \leq A$. (3.13) can be rewritten as :

$$\alpha_0^2 h\left(\frac{\mathrm{sh}(\alpha_0 h)}{\alpha_0 h(\mathrm{ch}(\alpha_0 h)-1)} - \frac{2}{(\alpha_0 h)^2}\right) < \frac{h\varphi(h)}{\int_h^{+\infty} \tau\varphi(\tau)d\tau}$$

$$\Leftrightarrow \alpha_0^2\left(\frac{\mathrm{sh}(\alpha_0 h)}{\alpha_0 h(\mathrm{ch}(\alpha_0 h)-1)} - \frac{2}{(\alpha_0 h)^2}\right) < \frac{\varphi(h)}{\int_h^{+\infty} \tau\varphi(\tau)d\tau}$$

The expression between brackets on the RHS is $\Psi(\alpha_0 h)$ defined in lemma G.1 is less than 1/6. Consider α_1 defined by :

$$\alpha_1 = \min_{0 \leq h \leq A} \frac{\varphi(h)}{\int_h^{+\infty} \tau\varphi(\tau)d\tau}$$

As $\varphi(h)$ is positive, continuous except possibly at a finite number of points, the minimum is taken over a compact set and therefore $\alpha_1 > 0$. Therefore, choose $\alpha_0 \leq (6\alpha_1)^{1/2}$ and (3.13) is satisfied for $h \leq A$. In conclusion, provided that $\alpha_0 \leq \min(\beta_0, (6\alpha_1)^{1/2})$, (3.13) is satisfied. It proves also that if condition (3.13) is verified for α_0 than it is verified for any $\alpha < \alpha_0$.■

H. Proof of lemma 3.10

We start from rewriting condition (3.13) as :

$$\frac{\partial}{\partial h}\left(\log(\mathrm{ch}(\alpha_0 h) - 1) - 2\log(h)\right) < -\frac{\partial}{\partial h}\log \xi(h)$$

where

$$\xi(h) = \int_h^{+\infty} \tau\varphi(\tau)d\tau$$

and therefore :

$$\frac{\partial}{\partial h}\left(\log \frac{(\mathrm{ch}(\alpha_0 h)-1)\xi(h)}{h^2}\right) < 0$$

As the argument within the logarithm is positive, this condition implies that this argument is always less than its value at 0:

$$\frac{(\mathrm{ch}(\alpha_0 h)-1)\xi(h)}{h^2} < \frac{\xi(0)}{2}\alpha_0^2$$

and therefore for any $h > 0$:

$$0 < \xi(h) < \frac{\xi(0)\alpha_0^2}{2} \frac{h^2}{\mathrm{ch}(\alpha_0 h) - 1} \tag{H.1}$$

Consider now :

$$V(u_1 - u_2) = \int_{-\infty}^{0} \tau^2 \varphi(\tau) d\tau + \int_{0}^{+\infty} \tau^2 \varphi(\tau) d\tau$$

and :

$$\int_{0}^{+\infty} \tau^2 \varphi(\tau) d\tau = [-\tau\xi(\tau)]_0^{+\infty} + \int_{0}^{+\infty} \xi(\tau) d\tau$$

Because of (H.1), the first term in the RHS is equal to zero and the second term is bounded by :

$$\frac{\xi(0)}{2} \int_{0}^{+\infty} \frac{\alpha_0^2 h^2}{\mathrm{ch}(\alpha_0 h) - 1} dh = \frac{\xi(0)}{2\alpha_0} \int_{0}^{+\infty} \frac{y^2}{\mathrm{ch}(y) - 1} dy$$

Using $\mathrm{ch}(y) - 1 = (e^y + e^{-y} - 2)/2 = e^{-y}(1 - e^y)^2/2$:

$$\int_{0}^{+\infty} \frac{y^2}{\mathrm{ch}(y) - 1} dy = 2 \int_{0}^{+\infty} \frac{y^2 \exp(y)}{(1 - \exp(y))^2} dy$$

$$= 2 \left(\left[\frac{y^2}{1 - \exp(y)} \right]_0^{+\infty} - 2 \int_{0}^{+\infty} \frac{y}{1 - \exp(y)} dy \right)$$

$$= -4 \int_{1}^{0} \frac{-\log(z)}{1 - 1/z} \left(-\frac{dz}{z}\right) = 4 \int_{0}^{1} \frac{\log(z)}{z - 1} dz = \frac{2\pi^2}{3}$$

where we change variables ($z = \exp(-y)$) and where we used Gradshteyn and Ryzhik (1995, p564). Summarizing:

$$\int_{0}^{+\infty} \tau^2 \varphi(\tau) d\tau = \frac{\xi(0)}{\alpha_0} \frac{\pi^2}{3}$$

Repeating the argument for $h < 0$ by reverting the $u_1 - u_2$ axis , we get :

$$V(u_1 - u_2) < \frac{2\pi^2}{3} \frac{\xi(0)}{\alpha_0}$$

■

	Non participant (1999)	Participant
Non participant (1998)	2201	351
Participant	300	8444

Table 1: Participation flows.

Variables [a]	**Conditional Logit**	**Conditional Probit** [b]	**Normal** [c] **Differences**
Child03	-1.00	-1.00	-1.00
	(-)	(-)	(-)
Child46	-0.673	-0.681	-0.675
	(.093)	(.089)	(.092)
Child718	-0.464	-0.470	-0.466
	(.094)	(.091)	(.093)
Spouse Income	-0.0025	-0.0025	-0.0025
(*Monthly; kF*)	(.016)	(.016)	(.016)
Age	-1.523	-1.563	-1.532
(*in decades*)	(.458)	(.458)	(.457)
Intercept	0.032	0.033	0.032
	(0.034)	(0.033)	(0.034)
LogLikelihood	-381.07	-380.85	-381.01

Notes: a. Estimation method: Conditional Likelihood. The number of observations is equal to 651. Variables are in first differences except age.

b. The probability $c/(1+c)$ is the standard normal.

c. The difference $u_1 - u_2$ is normally distributed.

Table 2: Conditional Likelihood estimation

Variables [a]	RE Logit [b]	RE Logit	RE Probit	RE Probit
Child03	-1	-1	-1	-1
	(-)	(-)	(-)	(-)
Child46	-0.752	-0.643	-0.769	-0.642
	(0.056)	(.140)	(.054)	(.138)
Child718	-0.368	-0.383	-0.363	-.369
	(.029)	(.098)	(.027)	(.093)
Spouse Income	-0.002	0.001	-0.0023	0.001
	(.003)	(.014)	(.003)	(.013)
Age	-2.562	-5.00	-2.84	-5.60
	(.330)	(.665)	(.331)	(.683)
Age^2 [(c)]	-0.074	-0.215	-0.066	-0.204
	(.055)	(.086)	(.052)	(.083)
Mean(Child03)		-0.839		-0.907
		(.108)		(.107)
Mean(Child46)		-0.637		-0.702
		(.095)		(.096)
Mean(Child718)		-0.203		-0.217
		(.080)		(.078)
Mean(SpIncome)		-0.0018		-0.0027
		(.014)		(.014)
$\sigma_\varepsilon^2/(\sigma_\varepsilon^2+\sigma_u^2)$ [(d)]	0.983	0.983	0.945	0.944
	(.00083)	(.00083)	(.0023)	(.0022)
LogLikelihood	-7764.255	-7735.902	-7769.715	-7737.959

Notes: a. The dependent variable is participation status in 1998 and 1999. Explanatory variables such as Mean(Child03) are constructed as the half sum of the values of these variables in 98 and 99.

b. Estimation method: Random effects Logit and Probit. Individual effects are normally distributed. The number of observations is equal to 11296.

c. This variable is constructed in such a way that its first difference is equal to the variable Age. Its coefficient is therefore directly comparable to the estimates of this coefficient reported in Table 2 and the specification remains quadratic in age.

d. It is the ratio of the variance of the individual effect to the total variance of the disturbances.

Table 3: Random effects estimation