Panel Binary Variables and Individual Effects: Generalizing Conditional Logit

Thierry Magnac, INRA, Paris-Jourdan and CREST*

First version: October 2000
This version: June 2002

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

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 this statistic, does not therefore depend on individual effects. It leads to $\sqrt{n}-$consistent and normally distributed estimators. Conditions under which this property of sufficiency is true are a lot less stringent than in the conditional logit approach (Rasch, Andersen, Chamberlain). In applied work, it gives justifications to quasi-differencing the binary variables as if they were continuous variables. Semiparametric approaches drawn from the cross-section literature can then be readily applied. Finally, the extension of a result of Chamberlain (1992) shows that the semi-parametric efficiency bound is not equal to zero if and only if the property of sufficiency is satisfied. It gives a second motivation to the paper by characterizing all panel binary choice models with individual effects where $\sqrt{n}-$consistent and regular estimators can be constructed.

Résumé

Nous étendons la méthode de Logit conditionnel utilisée pour traiter des données longitudinales dichotomiques avec effect fixes corrélés. Dans un modèle à deux périodes et deux états, nous présentons les conditions nécessaires et suffisantes qu'il faut imposer sur la fonction de répartition jointe des chocs individuels spécifiques aux deux périodes pour que la somme des deux variables binaires soit une statistique exhaustive pour l'effet fixe individuel. La fonction de variasemblance conditionnelle à cette statistique ne dépend donc pas de l'effet individuel. Cela permet de construire des estimateurs convergents en $\sqrt{n}$ et qui sont asymptotiquement normaux. Les conditions sous lesquelles la propriété d'exhaustivité est vraie sont beaucoup moins fortes que pour l'approche par Logit conditionnel (Rasch, Andersen, Chamberlain). Dans le travail appliqué, cela donne une justification eux procédures de quasi-différenciation des variables binaires comme dans le cas de variables continues. Des approches semi-paramétriques utilisées en coupe transversale peuvent être appliquées directement. Finalement, on étend un résultat de Chamberlain (1992) en montrant que la borne d'efficacité semi-paramétrique n'est pas égale à zéro si et seulement si la propriété d'exhaustivité est vérifiée. Cela donne une seconde justification au papier puiqu'on y caractérise tous les modèles de panel à choix binaire avec effets individuels où des estimateurs convergents en $\sqrt{n}$ peuvent être construits.

1. Introduction[1]

Using conditional likelihood methods when estimating panel binary choice models 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). Its properties are derived from the existence of 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 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, I extend conditional likelihood methods by relaxing the assumption of independence between individual-and-period specific shocks. A standard two-period two-state model is used. The pair of individual binary variables is described by a pair of 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. The properties of the conditional logit approach are derived from the proof that a sufficient statistic exists. We generalize conditional logit by characterizing directly models where the sufficient statistic for the individual effect is the number of times the binary variable is equal to 1. Namely, the property of sufficiency provides conditions on the joint distribution function of individual-and-period specific shocks. It also yields 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 the pair of binary variables and the difference in the linear index of covariates. The method of conditional likelihood therefore can be interpreted as a method giving the conditions for quasi-differencing binary variables and allowing

[1] I thank Andrew Chesher, Miguel Delgado, Jean-Pierre Florens and participants at seminars at CREST, Toulouse, Tinbergen I., UCLondon, Bonn, Montréal, Brown U. and Québec and conferences, EC^2 in Dublin, ESEM in Lausanne for helpful comments. The usual disclaimer applies.

standard packages to be used to estimate binary panel data models.

Panel binary choice models have been the focus of interest in the literature for more than 20 years (see Arellano and Honoré, 2001, for a recent survey). Most papers use "random effect" models introduced by Heckman (1978), where the distribution function of disturbances, including individual effects, is assumed to be independent of covariates. Moreover, most authors parametrically specify this distribution function (where I include cases where the individual effect is assumed to have a discrete and *finite* support). Very few papers use non-parametric methods as Chen, Heckman and Vytlacil, 1999, does.

Turning to the 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), Lewbel regressions (Honoré and Lewbel, 2002) 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. The maximum score approach is based on the weakest assumptions since individual-and-period specific shocks are only assumed stationary, conditional on covariates. This assumption is implied by strict exogeneity but is far weaker. For a cross-section, 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, a strict exogeneity assumption and stronger assumptions are maintained (Lee, 1999). In contrast, the adaptation of Lewbel (2000) estimation method to binary variables panel data do not require strict exogeneity but requires that a "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 not normal but smooth maximum score is "almost" root-n consistent and asymptotically normal (Horowitz, 1992). We do not know much yet, to my knowledge, about the efficiency of maximum rank correlation and Lewbel's 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, 2002, Woutersen, 2002). Two important results related to conditional logit were shown by Chamberlain (1992). Under the assumption of independence between individual-and-period specific shocks across time, 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 if and only if the distribution of specific shocks is logistic because the semi-parametric efficiency bound is equal to zero in all other cases. We generalize the latter result of Chamberlain (1992). If covariates are unbounded, then consistent estimation at a $\sqrt{n}$ rate is possible if and only if the sum of the binary variables is a sufficient statistic for the individual effect. It gives another motivation to the paper since it characterizes all models (where variables are strictly exogenous) for which it is possible to construct $\sqrt{n}-$consistent estimators. Therefore, the absence of "links" between the parametric part and covariates on the one hand, and the non-parametric individual effect on the other hand, in a random effect setting, not only leads to an efficiency gain as usual but also to a gain in the convergence rate.[2] Finally, we can reconcile these results with Chamberlain's. 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.

Because of the simplicity of the conditional likelihood approach, we believe that it could be the first step to attempt when fitting models to the data before considering random effect approaches. An empirical illustration is performed on a labor market participation example and shows the interest of the method. Section 2 illustrates and defines 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 a sufficient statistic is the sum of binary variables and we derive the conditions on the primitives of the problem. We study estimation and give examples in section 4 and we report the results of the empirical illustration in section 5. Section 6 is devoted to

[2] In a random effect setting, results on the speed of convergence are given by Chen, Heckman and Vytlacil (1999)

$\sqrt{n}-$consistency and tests of the property of sufficiency. Section 7 concludes.

2. The Conditional Logit Method

We first concentrate on how the conditional logit approach works and we set up the model we will be talking about in the rest of the paper. 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 random shocks (u_{i1}, u_{i2}) are independent of the covariates at the two periods and of the individual effect, $(z_{i1}, z_{i2}, \varepsilon_i)$:

The method of Conditional Logit described by Chamberlain (1984) for instance, 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, thus 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 not equal to 1 only if $K = 1$). By definition, the statistic $\sum_{t=1}^2 y_t = K$ is a sufficient statistic for the individual effect (Barndorff-Nielsen, 1978).

This approach has been criticized because the assumption of specific shocks that are independently and logistically distributed is seen as overly strong. The principle of sufficiency might however be used for any distribution such that:

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

is independent of ϵ.

Consider $x_i = -z_i\beta - \epsilon$ and assume as before that shocks (u_1, u_2) are independent of z_1, z_2 and ε. Then:

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

Rewriting (2.1) for $y_1 = 1$, $y_2 = 0$, sufficiency is verified if the following expression is independent of ε:

$$\frac{\Pr(u_1 > x_1, u_2 \leq x_2)}{\Pr(u_1 > x_1, u_2 \leq x_2) + \Pr(u_1 \leq x_1, u_2 > x_2)} = \frac{r(x_1, x_2)}{1 + r(x_1, x_2)}$$

where:

$$r(x_1, x_2) = \frac{\Pr(u_1 > x_1, u_2 \leq x_2)}{\Pr(u_1 \leq x_1, u_2 > x_2)}$$

The expression $r(x_1, x_2)$ is independent of ε if it depends on the only combination of (x_1, x_2) that does not depend on ε, that is the difference $x_1 - x_2$. If x_1 and x_2 are sufficiently continuously varying (in a sense made precise below), it is indeed a necessary and sufficient condition. Therefore, the variable $\sum_{t=1}^{2} y_t$ is sufficient for the individual effect if:

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

The aim of this paper is to find characterizations of function c and of the joint distribution function (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) either by maximizing the conditional likelihood function or by using the generalized method of moments. It is however by no means obvious that the function c is unconstrained and can be set as we want. It is the object of this section to derive these conditions. We shall first derive what are the general implications of the property of sufficiency on the joint d.f. of (u_1, u_2) and function $c(.)$. We will then turn to the necessary and sufficient conditions that shall be imposed on function $c(.)$.

3.1. The Joint Distribution Function

We first limit ourselves to the class of sufficiently smooth 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. Second in contrast to Manski (1987), idiosyncratic shocks are supposed to be independent of all variables but we shall return to this point in the conclusion. Finally, we assume that the individual effect is varying over the whole real line. These assumptions are gathered into:

Assumption A1 : i./ *Random shocks* (u_1, u_2) *have a strictly positive and bounded density function with respect to the Lebesgue measure and are independent of* z_1, z_2 and ε.

ii./ *The marginal distribution function of* u_1 *is denoted* $F(.)$ *and its density function,* $f(.)$. *It is such that* $F(0) = \frac{1}{2}$.

iii./ The individual effect ε *varies continuously over the whole real line* $\mathbb{R}$ *(conditionnally on all other covariates)*

Assumption ii./ on the median is the more common normalization assumption. It suffices to include a constant among variables z_i. Assumption iii/ makes indices x_1 and x_2 vary over the whole real line but this property could be obtained by stating an assumption on the variation of the covariates instead (see Assumption A3).

If we also include a period-specific variable in z_1 and not in z_2, we will see than we need the following normalization concerning function $c(.)$ defined by (2.2):

Assumption A2: $c(0) = 1$

Finally identification restrictions related to covariates in this type of model are derived from Manski (1987). A sufficient identification condition is for instance (see Manski, 1987, Assumption 2, p.358):

Assumption A3: *i./ The difference between the first covariates across periods,* $z_1^{(1)} - z_2^{(1)}$, *varies continuously over the whole real line* R *(conditionnally on all other covariates) and its coefficient* $\beta^{(1)} = 1$.

ii./ The support of $z_1^{(1)} - z_2^{(1)}$ *is not contained in any proper linear subspace of* R^K *where* K *is the dimension of* z*.*

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

Theorem 3.1. *Assume A1-A3 and :*

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

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:*

$$\Pr(u_1 \leq x_1, u_2 \leq x_2) = \frac{F(x_2) - c(x_1 - x_2)F(x_1)}{1 - c(x_1 - x_2)} \text{ when } x_1 \neq x_2 \qquad \text{(a)} \quad (3.1)$$

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

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 (3.1a) is technically undefined when $x_1 - x_2 = 0$. We consider the limit of equation (3.1a) 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 (3.1b). It is only defined when $c'(0) \neq 0$ and therefore $c'(0) < 0$ (point 4). As the joint density function is bounded by assumption (A1), we can differentiate (3.1b) 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 interesting. 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.

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 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. Before that, we begin with an interesting (and known) application of this theorem when u_1 and u_2 are independent.

Corollary 3.2. *Assume A1-A3, condition (S) and that u_1 and u_2 are independent. Then, u_1 and u_2 are logistically distributed. Formally, there exists $\mu > 0$ such that:*

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

Proof. Condition (3.1b) and the independence assumption imply that :

$$(F(x))^2 = F(x) + \frac{f(x)}{c'(0)} \tag{3.2}$$

For any x, $0 < F(x) < 1$ by A1. Denote:

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

(3.2) 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 (3.1) we get the expression of $c(h)$. ■

In the present paper, we do not impose that individual-and-period specific random variables are independent and a broader class of distribution functions will be shown to lead to the property of sufficiency. What we can infer from this corollary is that the intersection of the set of joint distribution functions satisfying the property of sufficiency and the set of independent joint distribution functions is reduced to one point, the logistic distribution function.

We return now to studying whether equations (3.1) define a proper joint distribution function and whether the class of joint distribution functions defined by Theorem 3.1 is not reduced to the logistic case. A necessary and sufficient 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 (3.1a), 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 (3.1) to verify assumption A1(i./) is:

$$\forall (x_1, x_2); \quad 0 < g(x_1, x_2) < +\infty \tag{3.3}$$

Proof. As shown in appendix B, $g(x_1, x_2)$ is the density function of (u_1, u_2) that can be derived from (3.1). Furthermore, if $\Pr(.,.)$ is defined by (3.1), it verifies:

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

and this function is a proper joint distribution function as defined in Assumption A1(i./) ■

It is difficult to directly tackle this necessary and sufficient condition. We shall proceed step by step. The first step shows that, under the conditions of Theorem 3.1, 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 function, $c(.)$. It also yields additional necessary conditions that are imposed on $c(.)$ by the property of sufficiency (S).

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-A3 and condition (S) then :*

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

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 (3.3).

Proposition 3.5. *Let $\phi(h)$ the distribution function of $u_1 - u_2$. Under assumptions A1-A3, and condition (S), we necessarily have:*

$$\int_0^{+\infty} \tau\varphi(\tau)d\tau = \int_0^{-\infty} \tau\varphi(\tau)d\tau < +\infty \qquad \textit{(P(1))}$$

$$\lim_{h\to+\infty} h(1-\phi(h)) = 0 \qquad \lim_{h\to-\infty} h\phi(h) = 0 \qquad \textit{(P(2))}$$

$$\exists\beta_0 > 0; \lim_{h\to+/-\infty}\frac{|h\varphi(h)|}{\int_h^{+\infty}\tau\varphi(\tau)d\tau} > \beta_0 \qquad \textit{(P(3))}$$

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 condition C(1).

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 popular distributions.

The previous 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 :*

$$\lim_{h\to-\infty} h\frac{d}{dh}\left(\frac{h}{1-c}\right)=0 \text{ and } \lim_{h\to+\infty} h\left(1-\frac{d}{dh}\left(\frac{h}{1-c}\right)\right)=0 \qquad (C(2))$$

$$\exists\beta_0>0; \min\left(-\lim_{h\to+\infty}\frac{s'(h)}{s(h)}, \lim_{h\to-\infty}\frac{s'(h)}{s(h)}\right)>\beta_0 \qquad (C(3))$$

Proof. See appendix E. ■

After showing the equivalence between stating the problem in terms of function $c(.)$ or in terms of d.f. $\phi(.)$, we can now turn to the condition that we shall impose on the other structural functional $F(.)$ to be able to solve the necessary and sufficient condition (3.3). Before that, we show that it is necessary and sufficient to study only the case where $x_1 \geq x_2$ (or $h \geq 0$) because the problem is symmetric. At first reading, this subsection might be skipped.

3.3. A Technical Simplification: Exploiting the Symmetry of the Problem

It is helpful to note a fundamental symmetry in the problem in terms of disturbances u_1 and u_2. We will use symmetry to simplify the proofs given in the next subsection. 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 (S) 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 (3.3) is also affected in the same way since it can be equivalently written as (setting $x_1=x+h$, $x_2=x$):

$$\forall(h,x); 0<s(h)(f(x+h)+f(x))+s'(h)(F(x+h)-F(x))<+\infty \qquad (3.4)$$

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:

$$\forall(h,x); 0 < s(-h)(f(x)+f(x+h)) - s'(-h)(F(x+h)-F(x)) < +\infty \tag{3.5}$$

Equations (3.4) and (3.5) 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. $1/c(h)$).

3.4. Conditions on the Marginal Distribution Function $F(.)$

Because of the simplification derived in the previous subsection, we shall assume from now on that $h \geq 0$ and that $\phi(h)$ can either be interpreted as the proper $\phi(h)$ or $1-\phi(h)$. We are now in a position to prove that there exist distribution functions $F(.)$ such that condition (3.3) is satisfied. To prove the point more easily, we further restrict conditions given in proposition 3.5. We write an additional condition on the d.f. $\phi(.)$:

$$\exists \alpha_0 > 0 \text{ such that } \forall h \geq 0; \frac{\alpha_0 \mathrm{sh}(\alpha_0 h)}{\mathrm{ch}(\alpha_0 h) - 1} - \frac{2}{h} < \frac{h\varphi(h)}{\int_h^{+\infty} \tau\varphi(\tau)d\tau} \tag{P.3'}$$

where sh(.) and ch(.) are the hyperbolic sine and cosine functions. [3]

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 P(3) on the limit behaviour of the term on the RHS, is a consequence of condition P(3'). Second, we show after the main result[4] that under weak conditions, condition P(3) implies condition P(3'). The following proposition gives the conditions on the marginal distribution function $F(.)$.

Proposition 3.7. *For any function $\phi(.)$ satisfying conditions P(1, 2 and 3'), there exist functions $F(.)$ satisfying (3.3). They are necessarily such that :*

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

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

$$\forall x; \frac{f"(x)}{f(x)} < \alpha_0^2 \leq \frac{6\varphi(0)}{\int_0^{+\infty} \tau\varphi(\tau)d\tau} \tag{3.6}$$

where α_0 is defined by condition P(3').

Proof. See appendix F. ■

It means that f should not be "too" convex. It is the consequence of condition P(3') when we solve for equation (3.3). 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 this condition, in particular, the double-exponential or Laplace distribution as shown in appendix F.

We now weaken condition P(3') which in most cases is difficult to verify, to condition P(3) 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 condition P(3)

Then, it satisfies condition P(3').

Proof. : See appendix G. ■

Therefore, for most density distributions that one is willing to consider , the important condition to verify is condition P(3).

We can also write this result in terms of function $c(.)$. Condition P(3') can be rewritten using the one to one relationship between $\phi(.)$ and $c(.)$ as :

$$\exists\alpha_0 > 0; \forall h > 0; \frac{\alpha_0 \text{sh}(\alpha_0 h)}{\text{ch}(\alpha_0 h) - 1} < -\frac{s'(h)}{s(h)} \tag{C(3')}$$

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 conditions C(1, 2 and 3').

and $\mathcal{F}_c$ is the non empty set of functions F defined in proposition 3.7.

Let $c(.) \in \mathcal{C}$ and $F \in \mathcal{F}_c$. Equation (3.1) defines a joint distribution function that verifies assumption A1(i./) and property (S).

Proof. The necessary and sufficient condition (3.3) is verified by using proposition 3.7. ■

In the next section, we will study different examples of specification for function $c(.)$ or its equivalent representation, the distribution function $\phi(.)$. We can however reconsider now the particular case of Conditional Logit.

3.5. 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 general expression of the density function of the difference, $u_1 - u_2$:

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

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 (C(2)-C(3)). More interestingly, the last condition (C(3')) can be written as:

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

$$\Leftrightarrow \exists\alpha_0; \forall h > 0; \alpha_0 \frac{1+\exp(-\alpha_0 h)}{1-\exp(-\alpha_0 h)} < \frac{1+\exp(-h)}{1-\exp(-h)}$$

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

Equation (3.6) is verified 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 $(c(.), F(.))$ yields a proper joint distribution function as expected. It is however far to be the only marginal distribution function which can be associated to conditional logit. Any function such that $\frac{f"(x)}{f(x)} < 1$ would qualify. Independence between shocks across time is however possible if $F(.)$ is logistic only (Corollary 3.2).

3.6. Bounds on the correlation

More generally, one may wonder whether the conditions on $c(.)$ and $F(.)$ bound the correlation between the disturbances u_1 and u_2. The correlation is controled by the relative magnitude of the variance of the individual effect and of the specific shocks. Even if there are only two dates, this relative magnitude might be important as we are not in a linear context.

For simplicity, we shall from now on assume that all moments of the distributions exist and we write:

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

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

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

where $V(u_1 - u_2)$ is given by the distribution function of the difference, $u_1 - u_2$, or equivalently by $c(.)$, and where Vu_1 is given by the other structural object, $F(.)$. Equation (3.6) imposes a condition on the relationship between these two structural functionals. We derive 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 (3.6) 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 P(3'), it also affects $V(u_1 - u_2)$. This variance is bounded by the following expression.

Lemma 3.10.

$$V(u_1 - u_2) \leq \frac{2\pi^2}{3\alpha_0} \int_0^{+\infty} \tau\varphi(\tau) d\tau$$

Proof. See appendix H. ■

The coefficient of correlation is therefore bounded by :

$$\rho \geq 1 - \frac{\pi^2 \alpha_0}{6} \int_0^{+\infty} \tau\varphi(\tau) d\tau$$

Deriving that $\alpha_0^2 \leq \frac{6\varphi(0)}{\int_0^{+\infty} \tau\varphi(\tau) d\tau}$ by 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 introduced in the next section, this bound is always less than -1.

4. Estimation and Semi-Parametric or Parametric Specifications

We now use these results to derive the estimating equations and give some parametric examples where we use the simple 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 maximizing the likelihood function or using equation (4.2) and the generalized method of moments. It should be noted that we are back to the static case as 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. The other identification conditions are given by Assumption A3 or in Manski, 1988 or Horowitz, 1998. There is at least one continuously distributed regressor and by normalization its coefficient is set to 1. It means 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 (maximum score, maximum correlation, Lewbel's method), other methods (average derivatives) but also an efficient method under our assumptions, that is the estimator developed by Klein and Spady (1993). It should be remarked that conditions (C(1-3)) imposed on $c(.)$ do not affect the results developed in that paper because these conditions relate to the tails of the distribution functions that should not affect the estimator of Klein&Spady.

Finally, borrowing the idea of Manski (1987), we can extend this estimation principle to T periods with $T > 2$, by using the principle of differencing across any two periods in sequence and treating the result as a pseudo-likelihood function.

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 popular distributions in (4.1), for instance the normal d.f. and we call the method Conditional Normal (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 Normal

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. Condition C(1) first translates into:

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

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 :

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

We just consider the case of $h > 0$ as the d.f. is symmetric (see subsection 3.3). 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 P(3) 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 condition P(3) and is continuous, we can apply lemma 3.8 to prove condition P(3'). All the conditions in theorem 3.9 are therefore proven and Conditional Normal is a valid method.

To finish, 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 :

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

which 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 in particular P(1-2). Condition P(3) 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 P(3').

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 Empirical Illustration

We now present an illustration of this estimation method on the prototypical binary model of female participation in the labor market. We chose to use real data instead of a Monte Carlo analysis since conditional likelihood estimation is far from being a new estimation method (Andersen, 1973). It is the way models are specified that is original here. For the empirical application, 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 is defined as employment or the search for a job. Only 5.8% of women in 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. All variables are supposed to be strictly exogenous.

Results of conditional likelihood estimation for various 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.

Those specifications lead to very similar estimates. Whether it be Conditional Logit or Normal, 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 different by any reasonable criterium. 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 chose Age (in months) as the (approximately) continuously distributed variable – 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 conditional likelihood results to the results of standard random effect models (Logit and Probit) where individual effects are supposed to be normally distributed. We also report estimation results of models where 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, 2001). 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 the Age squared, 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 non-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. Convergence Rates and Specification Tests

The first motivation of this paper was to present generalizations of Conditional Logit because it was felt that the assumption of logistic distributions was too strong. This type of distribution functions is however very peculiar in the panel binary choice model as it was proved in a unpublished paper of Chamberlain (1992). We restate his main result that applies to the same setting as ours except that specific shocks are assumed to be independent.

Theorem 6.1 (Theorem 2, page 7, Chamberlain (1992)). *Suppose that u_1 and u_2 are independent. The semi-parametric efficiency bound (of the parametric vector β), $I_\Lambda = 0$ for all β in Θ unless the distribution function of random shocks F is logistic.*

In this section, we extend this result to our setting by showing that a non-zero semi parametric efficiency bound and the property of sufficiency (S) are equivalent. This result says that $\sqrt{n}$ consistent estimators can only be constructed under the property of sufficiency (S). At the end of this section, we show how the implied specification for the joint distribution function can be tested.

6.1. $\sqrt{n}-$ Consistency

We generalize Chamberlain (1992) by showing that the semiparametric efficiency bound of the parametric part, β, 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:

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

to be able to write:

Lemma 6.2. *The semi-parametric efficiency bound $I_\Lambda = 0$ for all β in Θ unless the distribution function of random shocks is such that:*

$$\begin{aligned} \forall z_1, z_2; \exists\psi &= (\psi_1,\psi_2,\psi_3,\psi_4)\in\mathbb{R}^4 \text{ such that:} \\ \forall\varepsilon &\in \mathbb{R},\ \psi' a(z,\varepsilon,\beta)=0 \end{aligned} \tag{6.1}$$

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. Fix z_1 and z_2. Note first that if $\varepsilon \to +\infty$ then necessarily $\psi_4 = 0$ and that if $\varepsilon \to -\infty$ then necessarily $\psi_1 = 0$ (Chamberlain, 1992). Therefore condition (6.1) is equivalent to:

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

which is equivalent to:

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

where this ratio is independent of ε (but depends on z_1 and z_2). This condition is a slightly stronger version of condition (S) where the assumption of independence w.r.t. the covariates has been abandoned:

$$\frac{\Pr(u_1 > x_1, u_2 \leq x_2)}{\Pr(u_1 \leq x_1, u_2 > x_2)} = c(x_1 - x_2; z_1, z_2) \tag{S'}$$

We can therefore summarize this development as:

Theorem 6.3. *The semi-parametric efficiency bound is equal to zero unless the sum of binary variables is a sufficient statistic for the individual effect.*

We will see in the conclusion how our results could be extended using property (S'). It is immediate to note that property (S) is equivalent to (S') **and** independence of the covariates.

In Chamberlain (1992) there is another result about identification when regressors are bounded. It is a conjecture that the proof of this result can also be derived in the current framework.

The previous theorem can also be used to assert that:

Corollary 6.4. *The sum of the binary variables across time is the only statistic to which can be applied the principle of sufficiency, with respect to individual effects, in the panel binary choice model.*

Proof: Suppose there exists a statistic such that the principle of sufficiency, with respect to individual effects, could be applied to. Then, using conditional likelihood methods, it would be possible to construct a $\sqrt{n}$ consistent estimator by conditioning on this statistic. By the previous theorem, we know that the sufficient statistic is the sum of binary variables across time. ■

6.2. Specification Tests

The set of models verifying the property of sufficiency (S) is a restricted set. Restrictions might therefore be tested and that is what we briefly investigate now. Given (A1-A3) we know that β is identified and we will treat this parameter as known. 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:

$$\begin{cases} \Pr\{y_1 = 0, y_2 = 0 \mid z_1, z_2\} = & E_\varepsilon \Pr\{y_1 = 0, y_2 = 0 \mid z_1, z_2, \varepsilon\} \\ \Pr\{y_1 = 0 \mid z_1, z_2\} = & E_\varepsilon \Pr\{y_1 = 0 \mid z_1, z_2, \varepsilon\} \\ \Pr\{y_2 = 0 \mid z_1, z_2\} = & E_\varepsilon \Pr\{y_2 = 0 \mid z_1, z_2, \varepsilon\} \end{cases}$$

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 equation (3.1) and using that the marginal distribution functions of u_1 and u_2 are equal to $F(.)$, we can rewrite this system as (where $\Delta z = z_2 - z_1$):

$$\begin{cases} \Pr\{y_1 = 0, y_2 = 0 \mid z_1, z_2\} = & \dfrac{E_\varepsilon(F(z_2\beta + \varepsilon) \mid z_1, z_2) - c(-\Delta z.\beta).E_\varepsilon(F(z_1\beta + \varepsilon) \mid z_1, z_2)}{1 - c(-\Delta z.\beta)} \\ \Pr\{y_1 = 0 \mid z_1, z_2\} = & E_\varepsilon(F(z_1\beta + \varepsilon) \mid z_1, z_2) \\ \Pr\{y_2 = 0 \mid z_1, z_2\} = & E_\varepsilon(F(z_2\beta + \varepsilon) \mid z_1, z_2) \end{cases} \tag{6.2}$$

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

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

the integral being taken over the support of ε. These equations are convolution equations which properties relative to identification, remain to be sought. It implies that some constraints on the observable marginal distribution functions are satisfied. For instance, that for any z_1, z_2 such that:

$$z_1\beta = z_2\beta$$

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 property of sufficiency is to reconsider the first equation of system (6.2) and replace the expressions in the first equation by their value given by the two last equations to obtain:

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

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

7. Conclusion

In this paper, we used the principle of sufficiency and conditioning to derive a generalization of the Conditional Logit method. We presented the conditions under which we can quasi-difference binary data as if they were continuous. Standard statistical packages, such as Probit, or more general semi-parametric procedures, such as Klein&Spady's, can then be used to estimate these models with unrestricted individual effects and compare these results with those that are obtained using random effect specifications. By extending a result by Chamberlain, we also showed that it is under the property of sufficiency only that we can construct $\sqrt{n}$ consistent estimators in the panel binary choice model.

There are some straightforward extensions. The first extension is that the index property, writing latent variables as $z_t\beta$, is far from necessary and we only used it for its simplicity. Deterministic parts of latent variables in each period could be written as $f_t(z_t, \beta_t)$ and the conditional model would become a function of the difference between these non-linear indices provided that the latent models remains additive in the individual effect. Functions f_t could even be partially unknown if the conditions of Matzkin (1992) are fulfilled. It is one of the advantages of our approach to return from panel binary choices to the better known world of cross-section binary choice models.

A more involved extension is to allow the distribution function of specific shocks (u_1, u_2) to depend on (z_1, z_2) as the last section and derived property (S') would suggest.

We could therefore get closer to the flexible framework of Manski (1987). Even under an assumption on conditional medians, it does not seem to be possible to let this d.f. depend on covariates in a completely unspecified way. The reason is that in our proofs, we have to use the continuous variation of $x_1+ x_2$ and $x_1 - x_2$ (see A1 and A3). The closer we can get to a general framework is to make (u_1, u_2) depend on all covariates except the continuous and varying covariates $(z_1^{(1)}, z_2^{(1)})$ (assumption A3). Function $c(.)$ could thus be supposed to be of the form $c(-\Delta z.\beta; z_1^{(-1)}, z_2^{(-1)})$ where $z_1^{(-1)}, z_2^{(-1)}$ include all covariates except the first. It might be reminiscent of the assumption used by Lewbel under a quite general setting.

A third extension would be related to data observed over a longer time period $(T > 2)$. As already said, our results can be applied to the pseudo-likelihood setting where periods are chained two-by-two as suggested by Manski (1987) in another context (maximum score). Applying the property of sufficiency in the $T = 3$ case is not very interesting. Some tedious investigation revealed that the only possible distribution function that verifies (S) in that case, is the logistic d.f. The idea of the proof is based on the fact that with three periods, the relative probabilities depicted by (S) of exchangeable choices between any two periods should not depend on the level of the latent variables in the third period because this variable contains the individual effect. It is where the Independence of Irrelevant Alternatives property comes in and drives us back to the logistic distribution.

Other lines of research seems more difficult to follow. It remains to be seen how such an approach would be applied to other non-linear models. It might be easier to extend this approach in models where we know that the principle of sufficiency can be applied (Weibull, Poisson,...). It seems to be a lot more difficult in dynamic models (Honoré and Kyriazidou, 2000, Hahn, 2001) and even more difficult with other models (Tobit, non-linear continuous etc) but it might be worth trying.

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, 2001, "Panel Data Models: Some Recent Developments", in eds E. Leamer and J.J. Heckman, *Handbook of Econometrics*, 5(53):3229-96, 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 Z. Griliches and M. Intriligator, *Handbook of Econometrics*, 2(22):1248-1313, North Holland: Amsterdam.

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.

Hahn, J., 2001, "The Information Bound of a Dynamic Panel Logit Model with Fixed Effects", *Econometric Theory,* 17:913-932.

Heckman, J.J., 1978, "Simple Statistical Models for Discrete Panel Data Developed and Applied to Tests of the Hypothesis of True State Dependence against the Hypothesis of Spurious State Dependence", *Annales de l'INSEE*, 30-31:227-69.

Honoré, B.E. and E., Kyriazidou, 2000, "Panel Data Discrete Choice Models with Lagged Dependent Variables", *Econometrica*, 68:839-74.

Honoré, B.E. and A. Lewbel, 2002, "Semiparametric Binary Choice Panel Data Models without Strictly Exogeneous Regressors", WP, Boston College, forthcoming

Econometrica.

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., 2002, "Orthogonal Parameters and Panel Data", Brown University, forthcoming in *Review of Economic Studies.*

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.

Matzkin, R., 1992, "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and The Binary Choice Models", *Econometrica*, 60:239-270.

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

Woutersen, T., 2002, "Robustness against incidental parameters and mixing distributions", University of Western Ontario.

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:

$$K(h,\Delta) = P(u_1 > \frac{\Delta + h}{2}, u_2 \leq \frac{\Delta - h}{2})$$
$$G(h,\Delta) = P(u_1 \leq \frac{\Delta + h}{2}, u_2 > \frac{\Delta - h}{2})$$
$$M(h,\Delta) = P(u_1 \leq \frac{\Delta + h}{2}, u_2 \leq \frac{\Delta - h}{2})$$

By the main property (S):

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

and by construction :

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

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

Therefore :

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

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

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

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 (3.1). 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 (3.1):

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

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.■

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 :

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

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:

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

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 (3.3). 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 (3.3) 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 (3.3) 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 (3.3) 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 P(3') 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 P(3'), 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.■

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 P(3') 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 $y > 0$, $\Psi(y) \le 1/6$

Proof : We prove first that, by extension, $\Psi(0) = 1/6$, second that $\Psi'(y) \le 0$.
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 $\frac{d}{dy}\frac{\mathrm{sh}y}{(\mathrm{ch}y-1)} = -\frac{1}{\mathrm{ch}y-1}$:

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

$$= \frac{-(\mathrm{sh}y + y)}{y^2(\mathrm{ch}y-1)} + \frac{4}{y^3}$$

We use that for any $y > 0$:

$$\begin{cases} \mathrm{sh}y \ge & y \\ \mathrm{ch}y - 1 \ge & y^2/2 \end{cases}$$

$$\Rightarrow \begin{cases} -(\mathrm{sh}y + y) \le & -2y \\ \dfrac{1}{y^2(\mathrm{ch}y-1)} \le & 2/y^4 \end{cases} \Rightarrow \frac{-(\mathrm{sh}y + y)}{y^2(\mathrm{ch}y-1)} \le -\frac{4}{y^3}$$

and therefore $\Psi'(y) \le 0$■

Returning to the proof of lemma 3.8, we consider a density function φ which verifies condition P(3) 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 P(3') 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):

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

because $\mathrm{ch}(\alpha_0 h) - 1 \geq (\alpha_0 h)^2/2$. Therefore consider $\alpha_0 \leq \beta_0$ and condition P(3') is verified for any $h > A$.

Consider now $h \leq A$. Condition P(3') can be rewritten as :

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

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

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 condition P(3') is satisfied for $h \leq A$. In conclusion, provided that $\alpha_0 \leq \min(\beta_0, (6\alpha_1)^{1/2})$, condition P(3') is satisfied. It proves also that if condition P(3') 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 P(3') as :

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

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 Normal** [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