Panel Binary Variables and Sufficiency: Generalizing Conditional Logit

Thierry Magnac, INRA, Paris-Jourdan and CREST-INSEE*

First version: October 2000
This version: June 2003

Abstract

This paper extends the conditional logit approach used in panel data models of binary variables with correlated fixed effects and strictly exogenous regressors. In a two-period two-state model, necessary and sufficient conditions on the joint distribution function of the individual-and-period specific shocks are given such that the sum of individual binary variables across time is a sufficient statistic for the individual effect. Under these conditions, $\sqrt{n}-$consistent conditional likelihood estimators exist. Moreover, it is shown by extending Chamberlain (1992) that $\sqrt{n}-$consistent regular estimators can be constructed in panel binary models if and only if the property of sufficiency holds. Imposing sufficiency is shown to reduce the dimensionality of the bivariate distribution function of the individual-and-period specific shocks. This setting is much less restrictive than the conditional logit approach (Rasch, Andersen, Chamberlain). In applied work, it amounts to quasi-difference the binary variables as if they were continuous variables and to transform a panel data model into a cross-section model. Semiparametric approaches can then be readily applied.

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

1 Introduction[1]

The elementary brick of panel binary models with correlated fixed effects is a two-period two-state model. 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 parameter β of the index is the parameter of interest. Estimating it by conditional logit 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 stem from the existence of a sufficient statistic for the individual effect which is the individual sum of the binary variables. By definition of S-sufficiency, the conditional likelihood function depends on the parameter of interest only while the marginal likelihood function depends on the parameter of interest and the nuisance parameter, the individual effect (Barndorff-Nielsen, 1978, Lancaster, 2000). Conditional logit is nevertheless seen to be restrictive because of the distributional assumptions. Yet, an intriguing and important result related to conditional logit was shown by Chamberlain (1992). If individual-and-period specific shocks are independent over time and if covariates are unbounded, consistent estimation at a $\sqrt{n}-$rate of the parameter of interest is possible if and only if the distribution of individual-and-period specific shocks is logistic. The semi-parametric efficiency bound is equal to zero in all other cases.

In this paper, I show that the sum of the binary variables is a sufficient statistic for the individual effect, under necessary and sufficient conditions that are much less restrictive than in the conditional logit approach. Moreover, the result of Chamberlain (1992) is generalized. 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. The property of sufficiency thus characterizes all models (when variables are strictly exogenous) for which it is possible to construct $\sqrt{n}-$consistent estimators. The strict exogeneity assumption can be partially relaxed. Conditional Logit is a special case. The only joint distribution function such that 1) the individual-and-period specific shocks are independent 2) the sum of the binary variables is a sufficient statistic, is the logistic distribution.

Panel binary models have been the focus of interest in the literature for more than 25 years (see Arellano and Honoré, 2001, for a recent survey). Most papers use "random effect" models

[1] I thank Andrew Chesher, Miguel Delgado, Jean-Pierre Florens, two anonymous referees, the coeditor and participants at seminars at CREST, Toulouse, Tinbergen, UCLondon, Bonn, CIRANO, Brown and Québec and conferences, EC[2] in Dublin, ESEM in Lausanne, Panel Data in Berlin, CAM in Copenhagen, for helpful comments. The usual disclaimer applies.

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. When individual effects are correlated with covariates and when the time dimension is small, we can group estimation approaches into three groups: The first group comprises maximum score (Manski, 1987) and maximum rank correlation (Lee, 1999), the second, pseudo-regression (Honoré and Lewbel, 2002) and the third, conditional logit. The first two groups include methods first developed for cross-section data. The maximum score approach is based on the weakest assumptions since individual-and-period specific shocks are assumed to be exchangeable only, conditional on covariates and the individual effect. This stationarity assumption is far weaker than strict exogeneity. Panel data applications of smoothed maximum score (Horowitz, 1992) were performed by Charlier, Melenberg and Van Soest (1995) and Kyriazidou (1998). 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 does not require strict exogeneity but requires that a "continuous" regressor be independent of the individual effect and requires a large support assumption. Conditional Logit requires strict exogeneity. 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 smoothed maximum score is "almost" root-n consistent and asymptotically normal (Horowitz, 1992). Honoré and Lewbel (2002) estimators are $\sqrt{n}-$consistent and asymptotically normal.

In this paper, it is shown that imposing the property of sufficiency reduces the dimensionality of the model from the unknown bivariate density function of individual-and-period specific shocks into two univariate density functions satisfying additional tail conditions. The property of sufficiency implies that shocks are exchangeable and this setting is nested into the setting of Manski (1987). The reduction of dimensionality explains why $\sqrt{n}-$consistency holds. Furthermore, maximizing conditional likelihood is shown to be equivalent to using an estimating equation which relates the expectation of the difference between the pair of binary variables and the difference in the linear index of covariates, in the sample where binary variables differ across time. It thus amounts to quasi-difference binary variables and turn panel into cross-section data. Semi-parametric techniques can then be readily applied.

Section 2 presents the set-up (Assumptions R), defines the property of sufficiency (Condition S) and proves the extension of Chamberlain (1992) about $\sqrt{n}$-consistent estimation. Section 3 is the main theoretical section where joint distribution functions that satisfy sufficiency are characterized as functions of two univariate distribution functions. Necessary and sufficient conditions on these functions (Properties P) are given. We study estimation and give a parametric illustration in Section 4. Section 5 proposes extensions and concludes. Proofs are in appendices.

2 Regularity Conditions, S-Sufficiency and $\sqrt{n}-$ Consistency

Consider y_1, y_2 two binary variables, z_1, z_2 two $K-$dimensional covariates[2]. The latent model is:

$$\text{For } t = 1, 2 : y_t = 1 \text{ if and only if } z_t\beta + \epsilon + u_t > 0$$

We adopt some regularity assumptions:

Assumption R1: *(i). The difference across time of the first covariate (say),* $z_1^{(1)} - z_2^{(1)}$*, continuously varies over the whole real line* $\mathbb{R}$ *(for almost all values of other covariates) and the coefficient of the first covariate,* $\beta^{(1)}$*, is equal to one.*

(ii). The support of $z_1 - z_2$ *is not contained in any proper linear subspace of* $\mathbb{R}^K$*.*

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

(iv). The individual effect ε *continuously varies over the whole real line* $\mathbb{R}$ *(for almost all values of covariates)*

Assumptions R1(i) and R1(ii) are sufficient identification restrictions borrowed from Manski (1987, Assumption 2, p.358). In contrast to Manski (1987), random shocks are supposed to be strictly exogenous and independent of all variables by R1(iii) though we discuss how to weaken this assumption in the last section. R1(iii) also limits the discussion to the class of sufficiently smooth distribution functions in order to use the usual tools of differential calculus. R1*(iv)* assumes complete variation of all probabilities in the "between-individual" dimension. It is similar to R1(i) that assumes complete variation in the "within-individual" dimension.

For Conditional Logit, specific shocks u_1 and u_2 are furthermore assumed to be independent and logistically distributed. It implies that the sum of binary variables is S-sufficient for the incidental parameter, ε, (Barndorff-Nielsen, 1978) that is, for $K = 0, 1$ and 2:

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

[2] We only consider random samples and we do not subscript individual observations by i.

While it trivially holds when $K = 0$ or 2, this expression yields the conditional likelihood of an observation such that $(y_1, y_2) \in \{(0,1),(1,0)\}$ when $K = 1$. The conditional likelihood function does not depend on the incidental parameter and the maximum conditional likelihood estimator is $\sqrt{n}-$consistent.

This approach has been criticized on the ground that the independence and logistic assumptions are overly strong. Observe that these assumptions seem indeed to be overly strong to derive $\sqrt{n}-$consistency since S-sufficiency alone implies that the conditional likelihood function is independent of the incidental parameter. Our first motivation is thus to prove that using S-sufficiency substantially generalizes Conditional Logit. Rewrite property (2.1) as Condition S(ufficiency), prominent in the rest of the paper:

Lemma 2.1 *Under Assumption R1, the sum of binary variables is S-sufficient for the individual effect, ε, if and only if there exists a real function $c(.)$ such that:*

$$\forall (x_1, x_2) \in \mathbb{R}^2; \frac{\Pr(u_1 > x_1, u_2 \leq x_2)}{\Pr(u_1 \leq x_1, u_2 > x_2)} = c(x_1 - x_2) \qquad \text{(Condition S)}$$

Another justification for using and characterizing S-sufficiency is that it is the weaker condition that one can find in the present set-up to construct $\sqrt{n}-$consistent estimators. In an unpublished paper, Chamberlain (1992) proved that if u_1 and u_2 are independent, the semi-parametric efficiency bound of parameter is equal to zero unless the distribution function of random shocks is logistic. It therefore tells us that $\sqrt{n}-$consistent estimators can only be constructed under the logistic assumption. By dropping the independence assumption, Chamberlain's result[3] is extended into:

Theorem 2.2 *Under Assumption R1, the semi-parametric efficiency bound for β is equal to zero unless the sum of binary variables is S-sufficient for the individual effect.*

3 Characterization of S-Sufficiency

Condition S is easily used in estimation (see Section 4). It is however by no means obvious that function $c(.)$ is unconstrained and can be set as wanted. It is the purpose of this section to derive these conditions. We first derive what are the general implications of the property

[3] This theorem also leads to prove the conjecture that the sum of binary variables is the only candidate for a sufficient statistic. 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, this sufficient statistic is the sum of binary variables across time.

In Chamberlain (1992) there is another result about identification when regressors are bounded. It uses a very similar technique of proof. It is a conjecture that a generalization of that result also holds.

of sufficiency on the joint distribution function of (u_1, u_2). We then prove that function $c(.)$ is in a one-to-one relationship with the distribution function of the difference $u_1 - u_2$. Necessary and sufficient conditions for sufficiency are then provided. We conclude by investigating what is the consequence of the additional assumption of independence between u_1 and u_2.

Before, by including two time-dummies among the explanatory variables, we can always adopt the following normalizations:

Assumption R2 : (i). *The marginal distribution function of* u_1 *is denoted* $F(.)$ *and its density function,* $f(.)$. *It is such that* $F(0) = \frac{1}{2}$.

(*ii*). $c(0) = 1$

First, we derive some necessary conditions for Condition S to hold and characterize the expression of the joint distribution function of (u_1, u_2).

Theorem 3.1 *Assume that conditions R1, R2 and condition S hold. Then :*
1. $c(h)$ *is a strictly decreasing function from* $+\infty$ *to* 0 *and is twice continuously 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:*

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

4. $F(.)$ *is three-times continuously differentiable and* $f"$ *is bounded.*

Claim 1 is directly derived from the limit conditions in Condition S. In claim 2, the identity of marginal distributions of u_1 and u_2 is reminiscent of the property of exchangeability at the heart of the score method developed by Manski (1987). This property is here shown to be the consequence of the sufficiency property which is therefore stronger than exchangeability. Claim 3 of Theorem 3.1 gives a characterization of the joint probability function in terms of two functions $F(.)$ and $c(.)$ only. The sufficiency property thus reduces the dimensionality of the problem, at least, from a function of two real arguments to two functions of one real argument. We show in the appendix how to define the joint probability distribution when $x_2 = x_1$ by continuity.

It is easier to continue to work with the distribution function of the difference, $u_1 - u_2$, which is in a one-to-one mapping with function $c(.)$.

Proposition 3.2 *Let $\phi(h)$ (resp. $\varphi(h)$) the distribution (resp. density) function of $u_1 - u_2$. Under conditions R1-R2 and condition S, we necessarily have:*

$$0 < \int_0^{+\infty} \tau\varphi(\tau)d\tau = \int_0^{-\infty} \tau\varphi(\tau)d\tau < +\infty \qquad \text{(Condition P1)}$$

$$\exists\beta_0 > 0; \lim_{h\to+/-\infty} \frac{|h\varphi(h)|}{\int_h^{+\infty} \tau\varphi(\tau)d\tau} > \beta_0 \qquad \text{(Condition P2)}$$

Function $c(.)$ and distribution $\phi(.)$ are in the following one-to-one relationship:

$$c(h) = 1 - \frac{h}{h\phi(h) + \int_h^{+\infty} \tau\varphi(\tau)d\tau} \tag{3.2}$$

$$\phi(h) = \frac{d}{dh}\frac{h}{1-c(h)} \tag{3.3}$$

The first condition tells us that the expectation of $(u_1 - u_2)$ when $u_1 - u_2 \geq 0$ is finite and uses that u_1 and u_2 are identically distributed. Thus, $E(u_1 - u_2) = 0$ even when Eu_1 does not exist. These two regularity conditions P1 and P2 are verified if the distribution of $u_1 - u_2$ is thin tailed and not too hectic at infinity, for instance, if $\phi(.)$ is the normal d.f. They are not if the distribution is Cauchy for instance.

In contrast, the marginal distribution $F(.)$ should have thick tails:

Proposition 3.3 *Under conditions R1-R2, and condition S, we necessarily have:*

$$\forall x; \frac{f''(x)}{f(x)} < \frac{6\varphi(0)}{\int_0^{+\infty} \tau\varphi(\tau)d\tau} \qquad \text{(Condition P3)}$$

For any distribution $\varphi(.)$ such that $\varphi(0) > 0$ and P.1 holds, the set of such distributions is not empty.

The marginal density function f should not be "too" convex and the normal distribution, say, would not qualify for this condition. In contrast, mixtures of normal distributions through a gamma-distributed precision parameter verify this condition as presented next section.

We can now prove that the sufficiency property can be used for a much broader set of joint distributions than the logistic.

Theorem 3.4 *Assume conditions R1(i),R1(ii) and R1(iv). Let $\mathcal{D}$ be the set of pairs (φ, f) of strictly positive, continuous and bounded density functions on $\mathbb{R}$, such that f is twice continuously differentiable and f'' is bounded, and such that R2, P1, P2 and P3 hold. Then, for any $(\varphi, f) \in \mathcal{D}$, equations (3.1) and (3.2) define a joint distribution function that verifies condition (S) and condition R1(iii).*

As said, the sufficiency property can also be interpreted as reducing the dimensionality from a set of bivariate density functions to a set of pairs of univariate density functions (φ, f). This reduction explains why the $\sqrt{n}-$consistency result can hold. The theorem gives additional restrictions P1, P2 and P3 though these conditions do not affect dimensionality. An open question is how restrictive they are for empirical work on top of the dimensionality reduction.

Observe also that the reduction of dimensionality achieved by assuming independence between u_1 and u_2 is of the same magnitude since the joint density function is then given by two univariate marginal density functions. Using both dimensionality reductions at the same time is very restrictive however since it yields only one parametric family, the logistic distribution.

Corollary 3.5 *Assume R1-R2, 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:*

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

4 Semi-Parametric Estimation and a Parametric Example

Under Condition S, the conditional probability that can be used as the estimating equation, can be written as (see equation A.1 in Appendix A):

$$\Pr(y_1 = 1, y_2 = 0 \mid z_1, z_2, \sum_{t=1}^{2} y_t = 1) = \frac{c((z_1 - z_2)\beta)}{1 + c((z_1 - z_2)\beta)} \equiv G((z_1 - z_2)\beta) \qquad (4.4)$$

Observe first that the model is a cross-section binary model. The "transformed" dependent variable can only take two values "Entry" ($y_1 = 0$, $y_2 = 1$, $\Delta y = +1$) or "Exit" ($y_1 = 1$, $y_2 = 0$, $\Delta y = -1$) in the sample of movers ($\sum_{t=1}^{2} y_t = 1$). As the index $(z_1 - z_2)\beta$ is linear in β, it is in this sense that the sufficiency property permits to "quasi-difference" the data. Semiparametric identification is achieved as in Manski (1988) or Horowitz (1998) using Assumption R1*(i,ii)*. Provided that conditions R2(ii), P1 and P2 are verified[4], all semi-parametric estimation methods applicable to binary models can be used: maximum score, maximum rank correlation, Lewbel's method, average derivatives, semi-parametric NLS or ML (Horowitz, 1998). Besides, we can extend this estimation principle to T periods

[4]Normalization R2 translates into $G(0) = 1/2$, condition P1 into $G'(0)$ is positive and bounded and P2 into conditions on the tails of $G(.)$. They are derived using equation (3.2).

with $T > 2$, by borrowing the idea of Manski (1987). Difference across any two periods in sequence and treat the result as a pseudo-likelihood function.

Parametric methods are less attractive than semi-parametric methods since assuming a parametric distribution different from the logistic implicitly assumes away independence. It is nevertheless interesting to investigate special cases to understand the consequences of imposing sufficiency on distribution functions. The simplest parametric example is the known case of logistic distributions when function $c(h) = \exp(-h)$ and the distribution function given by (4.4) is the logistic distribution. There are two routes to depart from this assumption. The first route is to use popular distributions in (4.4), for instance the normal d.f. It however seems to generate quite implausible distribution functions for the difference between u_1 and u_2.[5] The second route is to specify the distribution function of the difference between u_1 and u_2. We now briefly look at that case, when $u_1 - u_2$ is normally distributed with zero-mean and variance equal to σ_0, say. The density function, $\varphi_0(.)$, is symmetric around 0 and $\int_h^{+\infty} \tau\varphi_0(\tau)d\tau = \sigma_0\varphi_0(h)$ is finite when $h = 0$ (Condition P1). Condition P2 is satisfied since:

$$\frac{h\varphi_0(h)}{\int_h^{+\infty} \tau\varphi_0(\tau)d\tau} = \frac{h}{\sigma_0}$$

Looking for compatible marginals, the convexity condition P3 shall be satisfied for any marginals such that $f"/f < 6/\sigma_0$. Consider a mixture of zero-mean normal variates where the precision is gamma-distributed with parameter δ and λ. Then, the density function is:[6]

$$l(x) = \frac{\Gamma(\delta + 1/2)\lambda^\delta}{\sqrt{2\pi}\Gamma(\delta)(\lambda + x^2/2)^{\delta+1/2}}$$

and:

$$\max_x \frac{l"(x)}{l(x)} = \frac{\delta + 1/2}{\lambda(2\delta + 3)}(\delta + 1)^2$$

Choose λ and δ such that this maximum is less than $6/\sigma_0$ and consider the following model (where i now indices individual heterogeneity):

$$\begin{cases} u_{1i} = \sigma_i\xi_i + \sigma_0\xi_i^0/2 \\ u_{2i} = \sigma_i\xi_i - \sigma_0\xi_i^0/2 \end{cases} \tag{4.5}$$

where (ξ_i, ξ_i^0) are two independent zero-mean unit-variance normal variates and $1/\sigma_i^2$ is gamma-distributed with parameter δ and λ. Then, condition P3 is verified because the convolution of a distribution which verifies P3 (*i.e.* $\sigma_i\xi_{0i}$) with any distribution verifies P3.[7]

[5] Such and other examples are studied in the working paper, Magnac (2002) where bounds on correlation coefficients are also derived. In these examples, bounds are not limiting.

[6] These results are shown in an appendix available upon request or on my Web page.

[7] Observe that the original model is not unique since any random variable can be added to $(u_1 + u_2)/2$ and substracted from the individual effect ε. Results are invariant to these renormalizations.

Besides, u_{1i} and u_{2i} are identically distributed because ξ_i is symmetrically distributed. In empirical applications, model (4.5) can be used when there is "a lot of heterogeneity" in the levels of the shocks and less in the difference.

5 Discussion and Extensions

In this paper, we used the principle of sufficiency and conditional inference to derive a generalization of conditional logit. We presented the conditions under which we can quasi-difference binary data as if they were continuous. Cross-section semi-parametric procedures can be used to estimate these models with unrestricted individual effects and their results can be compared 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 linear index property, writing latent variables as $z_t\beta$, is far from necessary. 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.

A more involved extension is to permit the distribution function of specific shocks (u_1, u_2) to depend on (z_1, z_2). As the present model is nested into the flexible setting of Manski (1987), we could get closer to it. It does not seem to be possible however to let the joint distribution function depend on covariates in a completely unspecified way. The reason is that in our proofs, we have to use the continuous variation with respect to the individual effect, ε, and with respect to the difference, $x_1 - x_2$ (see Assumption R1). The closer we can easily get is to make the joint distribution of (u_1, u_2) depend on all covariates except the difference between the first continuous covariate $(z_1^{(1)} - z_2^{(1)})$. We can then repeat the present analysis, conditional on values of $w = (z_1^{(1)} + z_2^{(1)}, z_1^{(-1)}, z_2^{(-1)})$ where $z_1^{(-1)}, z_2^{(-1)}$ include all covariates except the first. All assumptions are written conditional on w and all results apply, conditional on w. It might be reminiscent of the assumption used by Lewbel (2000).

When data is observed over a longer time period $(T > 2)$, periods can be chained two-by-two, as already said. On the other hand, using Condition S in the $T = 3$ case is less interesting. Some tedious investigation revealed that the only possible distribution function that verifies condition S in that case, is the logistic d.f. The idea of the proof is based on the fact that in a three-period setting, the relative probabilities depicted by Condition 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 seem more challenging. 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) and even more difficult in other models such as Tobit-like models.

References

Andersen, E.B., 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.

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 University, 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.

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., and A., Lewbel, 2002, "Semiparametric Binary Choice Panel Data Models without Strict Exogeneity", *Econometrica*, 70:2053-2063.

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.

Kyriazidou, E., 1997, "Estimation of a Panel Data Sample Selection Model", *Econometrica,* 65:1335-64.

Lancaster, T., 2000, "The Incidental Parameter Problem since 1948", *Journal of Econometrics,* 95:391-413.

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.

Magnac, T., 2002, "Panel Binary Variables and Individual Effects: Generalizing Conditional Logit", WP CREST 2002-18, www.crest.fr/doctravail/document/2002-18.pdf.

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.

Appendices

Equations or lemmas starting with A to D refer to the section of the appendix where they are stated. Conditions starting with R, S and P and numerals refer to the text.

A Proofs in Section 2

Proof of Lemma 2.1 Denote $x_t = -z_t\beta - \epsilon$ and use condition R1(iii) to write:

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

By definition, S-Sufficiency is satisfied if and only if, for $(y_1, y_2) \in \{(0,1),(1,0)\}$, equation (2.1) is verified. If it is verified for one of these pairs, say $(1, 0)$, it is satisfied for the other pair since probabilities sum to one. Thus, S-sufficiency is equivalent to the property that

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

is independent of ε, because condition (2.1) says that:

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

is independent of ε.

By R1(iii), $r(x_1, x_2)$ is a smooth function from $\mathbb{R}^2$ to $\mathbb{R}$ and x_1 and x_2 vary over the whole real line by R1(i) and R1(iv). Thus, $r(x_1, x_2)$ is independent of ε if and only if it depends on the only combination of (x_1, x_2) that does not depend on ε, that is the difference $x_1 - x_2$. Thus, there exists a real function $c(.)$ such that $r(x_1, x_2) = c(x_1 - x_2)$. Reciprocally, if such an expression holds, S-sufficiency holds and:

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

Proof of Theorem 2.2 We adapt the proof of Chamberlain (1992), Theorem 2, page 7. 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 A.1 *The semi-parametric efficiency bound $I_\Lambda = 0$ for all β in Θ unless the distribution function of random shocks is such that:*

$$\begin{aligned} \forall z_1, z_2; \exists \psi &= (\psi_1, \psi_2, \psi_3, \psi_4) \in \mathbb{R}^4 \text{ such that:} \\ \forall \varepsilon &\in \mathbb{R},\ \psi' a(z, \varepsilon, \beta) = 0 \end{aligned} \tag{A.2}$$

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

Second, fix z_1 and z_2. By Assumption R1*(iv)*, ε continuously varies over $\mathbb{R}$. Observe that if $\varepsilon \to +\infty$ we must have $\psi_4 = 0$ and that if $\varepsilon \to -\infty$ we must have $\psi_1 = 0$ (Chamberlain, 1992). Therefore equation (A.2) in the Lemma above 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. It is equation (2.1) and by equivalence, Condition S stated in Lemma 2.1. Reciprocally, if condition S holds then the semi-parametric efficiency bound $I_\lambda \neq 0$ since the conditional likelihood estimator is $\sqrt{n}$-consistent.

B Proofs of Theorem 3.1

Proof of Theorem 3.1 Claim 1 is proven by using monotonicity properties of probability functions, limit conditions and assumption R1*(iii)*.

We now prove claims 2 and 3. The proof proceeds by reparametrizing the problem as:

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

and by observing that by R1(*i*) and R1(*iv*), pairs (Δ, h) span all $\mathbb{R}^2$. Let:

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

and

$$K(h, \Delta) = P(u_1 > \frac{\Delta + h}{2}, u_2 \leq \frac{\Delta - h}{2}) = P(u_2 \leq \frac{\Delta - h}{2}) - M(h, \Delta)$$
$$G(h, \Delta) = P(u_1 \leq \frac{\Delta + h}{2}, u_2 > \frac{\Delta - h}{2}) = P(u_1 \leq \frac{\Delta + h}{2}) - M(h, \Delta) \qquad \text{(B.1)}$$

Using these expressions and condition (S):

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

we can write

$$G(h, \Delta) - K(h, \Delta) = (1 - c(h))G(h, \Delta) = P(u_1 \leq \frac{\Delta + h}{2}) - P(u_2 \leq \frac{\Delta - h}{2}).$$

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

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

and :

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

Using equation (B.1) :

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

which is equation (3.1) in the text.

Observe that equation (3.1) tends to 0/0 when $h = x_1 - x_2 \to 0$. As the numerator and denominator are continuously differentiable, we can do Taylor expansions around points $x_1 = x_2$:

$$\begin{aligned} F(x_2) &= F(x_1) + (x_2 - x_1)f(x_1) + o(x_2 - x_1) \\ c(x_1 - x_2) &= 1 + c'(0)(x_1 - x_2) + o(x_2 - x_1) \end{aligned}$$

where $f(x_1) > 0$ (R1*(iii))*, $c'(0) < 0$ (Claim 1 of this Theorem) and:

$$\lim_{x_1 - x_2 \to 0} o(x_2 - x_1)/(x_2 - x_1) = 0$$

By taking limits of the numerator and denominator of equation (3.1) divided by $(x_1 - x_2)$, we have:

$$\Pr(u_1 \le x_1, u_2 \le x_1) = F(x_1) + \frac{f(x_1)}{c'(0)} \tag{B.2}$$

As the joint density function is continuous and bounded, we can continuously differentiate this last equation two times and the result is continuous and bounded. Therefore, $F(.)$ necessarily is three-times continuously differentiable and f" is bounded (Claim 4).

C Proof of Proposition 3.2

The proof of this proposition proceeds in various steps. We first exhibit a condition of symmetry which substantially simplifies proofs below. We then derive the joint density function of (u_1, u_2) and the distribution of $u_1 - u_2$. We finally derive the expression of function $c(.)$ and conditions P1 and P2 stated in the Proposition.

A Technical Simplification: Exploiting the Symmetry of the Problem There is a fundamental symmetry in the problem with respect to the disturbances u_1 and u_2. Symmetry is a direct consequence of condition (S). If we change u_1 into u_2 and u_2 into u_1, we change $c(h)$ into $1/c(h)$. By Theorem 3.1, we also know that the marginal distributions of u_1 and u_2 are identical and equal to $F(.)$. We can therefore limit the proofs below to the case, $h \geq 0$, **provided that** we verify the conditions bearing on the straight representation $c(h)$ and on the reverse representation $1/c(h)$. This property is summarized quite informally by:

Lemma C.1 *If Condition S holds and if conditions for $c(h)$ and $1/c(h)$ hold for any $h \geq 0$, they globally hold for $c(h)$.*

The Joint Density Function The joint density function is derived by noting that R1*(iii)* allows for differentiating two times equation (3.1). The second cross-derivative, or density function, denoted $g(x_1, x_2)$ is strictly positive by R1*(iii)* and equal to:

$$\forall x_1 \neq x_2; g(x_1, x_2) \equiv s(x_1 - x_2)(f(x_1) + f(x_2)) + s'(x_1 - x_2)(F(x_1) - F(x_2)) > 0 \quad \text{(C.1)}$$

where:

$$s(h) \equiv \frac{c'(h)}{(1 - c(h))^2} = \frac{d}{dh}\frac{1}{1 - c} < 0 \quad \text{(C.2)}$$

is a negative function since $c(.)$ is decreasing. It has a singularity point at $h = 0$. By continuity, we nevertheless can obtain $g(x_1, x_1)$ (see below Proof of Proposition 3.3).

The Distribution of $u_1 - u_2$ Observe that symmetry (Lemma C.1) can be used. Interverting u_1 and u_2 transforms the distribution of $u_1 - u_2$ into the distribution of the opposite $u_2 - u_1$.

Consider then $h > 0$ and use equation (C.1) to write $\Pr(u_1 - u_2 > h)$ as:

$$\int_{-\infty}^{+\infty} \int_{x_1 - x_2 > h} [s(x_1 - x_2)(f(x_2) + f(x_1)) + s'(x_1 - x_2)(F(x_1) - F(x_2))] \, dx_1 dx_2$$

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

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

Observe that:

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

because equation (C.2), because $\lim_{h \to +\infty} \frac{1}{1 - c(h)} = 1$ (Theorem 3.1) and because $\lim_{h \to +\infty} s(h) = 0$ by definition (C.2). Namely, $c(.)$ is decreasing and tending to 0 when h tends to zero so that c' tends to 0.

Using also:

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

we get:

$$\begin{aligned} \Pr(u_1 - u_2 > h) &= \int_{-\infty}^{+\infty} \left[(1 - \frac{1}{1 - c(h)}) f(x_2) - s(h)(F(x_2 + h) - F(x_2)) \right] dx_2 \\ &= 1 - \frac{1}{1 - c(h)} - s(h) \int_{-\infty}^{+\infty} [F(x + h) - F(x)] \, dx \end{aligned}$$

Given that all functions in this expression are well defined,

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

so that differentiation w.r.t h and integration can be permuted:

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

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

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

Replacing the integral in the expression of $\Pr(u_1-u_2>h)$ yields:

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

Denoting $\phi(h)=\Pr(u_1-u_2\leq h)$:

$$\phi(h)=\frac{1}{1-c(h)}+s(h)h=\frac{d}{dh}\frac{h}{1-c(h)} \tag{C.3}$$

using equation (C.2). Symmetry (Lemma C.1) can be checked and this formula applies to $h<0$ and by continuity to $h=0$.

Conditions on the Distribution Function $\phi(.)$: (Conditions P1 and P2) We now seek $c(.)$ as a function of $\phi(.)$ and derive necessary conditions on $\phi(.)$. First, integrate equation (C.3) to get:

$$\frac{h}{1-c}=\int_0^h\phi(\tau)d\tau+A \tag{C.4}$$

where A is a constant of integration to be found. As $\frac{h}{1-c}$ is equal to $-\frac{1}{c'(0)}>0$ when $h=0$, A is equal to $-\frac{1}{c'(0)}>0$. The following lemma determines A.

Lemma C.2

$$A=\int_0^{+\infty}\tau\varphi(\tau)d\tau$$

where $\varphi(h)=\phi'(h)$ is the density function of u_1-u_2.

Proof. Let $h>0$ and consider the joint density function given by equation (C.1):

$$s(h)(f(x+h)+f(x))+s'(h)(F(x+h)-F(x))>0$$

where $s(h)$ is a negative function. Use equation (C.3) to write:

$$s(h)=\frac{1}{h}(\phi(h)-\frac{1}{1-c}),$$

and integrate by parts the integral in equation (C.4):

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

to get:

$$s(h) = \frac{1}{h^2}(\int_0^h \tau\varphi(\tau)d\tau - A). \tag{C.5}$$

As $s(h) < 0$, $\int_0^h \tau\varphi(\tau)d\tau$ is bounded by A. Thus:

$$\lim_{h\to+\infty} h\varphi(h) = 0 \tag{C.6}$$

Rewriting equation (C.1) using $s(h) < 0$ and $F(x+h) - F(x) > 0$ for any $h > 0$:

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

Replacing $s(h)$ by its expression (C.5) implies that:

$$\forall h > 0, \forall x; \quad \frac{f(x+h) + f(x)}{F(x+h) - F(x)} < \frac{2}{h} + \frac{h\varphi(h)}{A - \int_0^h \tau\varphi(\tau)d\tau}$$

Taking limits when $h \to \infty$ yields:

$$\forall x; \quad 0 < \frac{f(x)}{1 - F(x)} \leq \lim_{h\to+\infty} \frac{h\varphi(h)}{A - \int_0^h \tau\varphi(\tau)d\tau}$$

Therefore:

$$\exists \beta_0 > 0; \lim_{h\to+\infty} \frac{h\varphi(h)}{A - \int_0^h \tau\varphi(\tau)d\tau} > \beta_0 \tag{C.7}$$

Because of equation (C.6), the numerator tends to zero and the limit of the denominator is thus necessarily equal to zero. ■

Replacing A by its expression in equation (C.4) and inverting it, yields:

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

To finish the proof and obtain conditions P1 and P2, we use symmetry (Lemma C.1). The reverse representation consists in changing u_1 into u_2 and vice-versa. Observe that if ϕ_r is the distribution of the opposite $(u_2 - u_1)$, we have $\phi_r(h) = 1 - \phi(-h)$ and therefore $\varphi_r(h) = \varphi(-h)$. Apply Lemma C.2 to that distribution to show that $A_r = \int_{-\infty}^0 \tau\varphi(\tau)d\tau$ is finite. As u_1 and u_2 have the same distribution and $E(u_1 - u_2)$ is finite because A and A_r are, it is necessarily equal to zero. Thus, we get Property P1:

$$0 < \int_0^{+\infty} \tau\varphi(\tau)d\tau = \int_0^{-\infty} \tau\varphi(\tau)d\tau < +\infty. \tag{C.8}$$

Second, consider equation (C.7) and apply it to the reverse representation to get Property P2:

$$\exists \beta_0 > 0; \lim_{h\to+/-\infty} \frac{\mid h\varphi(h) \mid}{\int_h^{+\infty} \tau\varphi(\tau)d\tau} > \beta_0. \tag{C.9}$$

We can also summarize the properties of $s(h)$ proven in this section and needed below:

$$s(h) = -\frac{1}{h^2} \int_h^{+\infty} \tau\varphi(\tau)d\tau \tag{C.10}$$

$$-\frac{s'(h)}{s(h)} = \frac{2}{h} + \frac{h\varphi(h)}{\int_h^{+\infty} \tau\varphi(\tau)d\tau} \tag{C.11}$$

D Proofs of Proposition 3.3, Theorem 3.4 and Corollary 3.5

Proof of Proposition 3.3 As $s(h)$ is strictly negative, equation (C.1) for $h > 0$ is equivalent to :

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

and therefore, using equation (C.11):

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

When h tends to zero, we can expand the LHS to the third order as $F(.)$ is continuously differentiable three times (Theorem 3.1):

$$\frac{F(x+h) - F(x)}{h} = f(x) + f'(x)\frac{h}{2} + f"(x)\frac{h^2}{6} + o(h^3)$$

$$f(x+h) + f(x) = 2f(x) + f'(x)h + f"(x)\frac{h^2}{2} + o(h^3)$$

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

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

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

$$h^2\frac{\varphi(0)}{\int_0^{+\infty}\tau\varphi(\tau)d\tau}f(x)$$

As $f(x) > 0$, equation (D.1) therefore implies Property P3:

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

In an appendix available upon request, it is proven that mixtures of zero mean normal variates verify this condition. Precision is the mixing parameter and is Gamma distributed of parameter δ and λ. Then:

$$\max_{x>0}\frac{f"(x)}{f(x)} = \frac{\delta + 1/2}{\lambda(2\delta + 3)}(\delta + 1)^2$$

and for any $\delta > 0$, we can choose $\lambda > 0$ to satisfy property P3.

In conclusion to this proof, and using equations (C.10) and (C.1), the density function on the 45° line can be written by continuity as:

$$g(x,x) = (f(x)\varphi(0) - f"(x)\int_0^{+\infty}\tau\varphi(\tau)d\tau/6) > 0$$

Proof of Theorem 3.4 To prove that equations (3.1) and (3.2) define a joint distribution function that verifies assumption R1 *(ii)* and Condition S we shall prove that the joint density function that these equations define, exists everywhere, is continuous, bounded and positive. Using equation (C.1), it is easy to see that it is defined everywhere (including for $x_1 = x_2$ as proved in the previous sub-section) and is continuous and bounded since $f"$ and φ are.

Let us prove that the joint density is positive. We consider the case $h > 0$ only and rely on symmetry (Lemma C.1) for $h < 0$ and on continuity for $h = 0$.

First it is proven in a Lemma available upon request that if condition P3 is satisfied (i.e. $\forall x; \frac{f"(x)}{f(x)} < \alpha^2$) then :

$$\forall h > 0, \forall x; \frac{\alpha.\mathrm{sh}(\alpha h)}{\mathrm{ch}(\alpha h) - 1} F(x+h) - F(x) > (f(x+h) + f(x)) \tag{D.2}$$

where sh(.) and ch(.) are hyperbolic sine and cosine functions. Second, if the following property holds:

$$\exists \alpha > 0; \forall h > 0; \frac{\alpha \mathrm{sh}(\alpha h)}{\mathrm{ch}(\alpha h) - 1} < \frac{2}{h} + \frac{h\varphi(h)}{\int_h^{+\infty} \tau\varphi(\tau)d\tau} = -\frac{s'(h)}{s(h)} \tag{D.3}$$

then equation (D.2) yields:

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

which proves by equation (C.1) that the joint density is positive. To finish the proof of Theorem 3.4, equation (D.3) shall thus be proven.

Observe first that the limit when $h \to \infty$ of $\frac{\alpha \mathrm{sh}(\alpha h)}{\mathrm{ch}(\alpha h)-1} - \frac{2}{h}$ is equal to α and that:

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

because $\mathrm{ch}(\alpha h) - 1 > (\alpha h)^2/2$ for $h > 0$. Then:

$$\frac{\alpha \mathrm{sh}(\alpha h)}{\mathrm{ch}(\alpha h) - 1} - \frac{2}{h} \leq \alpha$$

Use condition P.2 to define β_0 and M such that:

$$\forall h > M; \frac{h\varphi(h)}{\int_h^{+\infty} \tau\varphi(\tau)d\tau} > \beta_0$$

Set $\alpha \leq \beta_0$ and equation (D.3) is then verified for any $h > M$.

Consider now $h \leq M$. Equation (D.3) can be rewritten as :

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

In a Lemma available upon request, we prove that the expression between brackets on the RHS is positive and less or equal to 1/6. Set β_1 to:

$$\beta_1 = \min_{0 \leq h \leq M} \frac{\varphi(h)}{\int_h^{+\infty} \tau\varphi(\tau)d\tau}$$

As $\varphi(h)$ is positive and continuous and as the minimum is taken over a compact set, $\beta_1 > 0$. Therefore, choose $\alpha \leq (6\beta_1)^{1/2}$ and equation (D.3) is satisfied for $h \leq M$. In conclusion, provided that $\alpha \leq \min(\beta_0, (6\beta_1)^{1/2})$, equation (D.3) is satisfied for any $h > 0$. Using the reverse representation (Lemma C.1), we can prove that it is satisfied for any h. It also proves that if equation (D.3) is verified for α than it is verified for any $\alpha' < \alpha$.

Proof of Corollary 3.5 u_1 and u_2 are assumed to be independent. Then, equation (B.2) implies that :

$$(F(x))^2 = F(x) + \frac{f(x)}{c'(0)} \tag{D.4}$$

For any x, $0 < F(x) < 1$ by condition R1*(iii)*. Denote:

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

Equation (D.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 for $F(x)$. Using equation (3.1) we get the expression for $c(h)$.

Appendix available upon request containing technical lemmas

Proof of the convexity inequality used in the proof of Theorem 3.4

Lemma D.1 *When* $\forall x; \frac{f"(x)}{f(x)} < \alpha^2$, $\alpha > 0$ *then :*

$$\forall h > 0; F(x+h) - F(x) > \frac{ch(\alpha h) - 1}{\alpha.sh(\alpha h)}(f(x+h) + f(x))$$

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

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

$$m(\lambda) = f(x + \lambda h) > 0$$

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

$$\frac{m"(\lambda)}{m(\lambda)} < \alpha^2 h^2$$

Define also function $g(\lambda)$ such that:

$$g(\lambda) = \frac{f(x+h)\mathrm{sh}(\alpha h \lambda) + f(x)\mathrm{sh}(\alpha h(1-\lambda))}{\mathrm{sh}(\alpha h)}$$

Observe that:

$$g(0) = f(x), g(1) = f(x+h)$$
$$\frac{g"(\lambda)}{g(\lambda)} = \alpha^2 h^2 > \frac{m"(\lambda)}{m(\lambda)}$$

As the degree convexity of $g(.)$ is "larger" than the degree of convexity of $m(.)$, we now show that:

$$m(\lambda) > g(\lambda) \text{ for any } \lambda \in]0,1[$$

Let $\Psi(\lambda) = m(\lambda) - g(\lambda)$. Observe that $\Psi(0) = 0$, $\Psi(1) = 0$ and that $\Psi(\lambda)$ is twice differentiable. Thus:

$$\Psi^{"}(\lambda) = m(\lambda)(\frac{m"(\lambda)}{m(\lambda)} - \frac{g"(\lambda)}{g(\lambda)}) + \frac{g"(\lambda)}{g(\lambda)}(m(\lambda) - g(\lambda))$$

Because $m(\lambda) > 0$ and the inequalities above:

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

Assume, by contradiction, $\exists \lambda_0; \Psi(\lambda_0) < 0$. 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 twice differentiable concave function in a interval where it takes value 0 at the end points and is negative in between. Thus, $\Psi(\lambda) \geq 0$. By contradiction assume that $\exists \lambda_0 \in]0,1[; \Psi(\lambda_0) = 0$. It is impossible since Ψ would be concave at that point. Therefore $\Psi(\lambda) > 0$ for any interior point.

Returning to the main argument, we therefore have :

$$F(x+h) - F(x) = \int_x^{x+h} f(u)du = \int_0^1 f(x+h\lambda)hd\lambda = h\int_0^1 m(\lambda)d\lambda$$
$$\geq h\int_0^1 g(\lambda)d\lambda = h(\int_0^1 \frac{f(x+h)\mathrm{sh}(\alpha h\lambda) + f(x)\mathrm{sh}(\alpha h(1-\lambda))}{\mathrm{sh}(\alpha h)}d\lambda)$$

using the definition of $g(.)$. Thus, by symmetry :

$$F(x+h) - F(x) \geq h(\int_0^1 \frac{\mathrm{sh}(\alpha h\lambda)}{\mathrm{sh}(\alpha h)}d\lambda)((f(x+h) + f(x))$$

and the proof finishes by integrating the RHS. ■

Mixtures of normals Assume that $f(x)$ is a mixture of zero mean normal variates such that the precision is heterogenous:

$$f(x) = \int_0^\infty \frac{1}{\sqrt{2\pi}}\theta^{1/2}\exp(-\theta x^2/2)\mu(\theta)d\theta$$

where $\mu(\theta)$ is a Gamma density function of parameter δ and λ:

$$\mu(\theta) = \frac{\lambda^\delta\theta^{\delta-1}\exp(-\lambda\theta)}{\Gamma(\delta)}$$

First, we get:

$$f"(x) = \int_0^\infty \frac{1}{\sqrt{2\pi}}\theta^{1/2}(-\theta+\theta^2x^2)\exp(-\theta x^2/2)\mu(\theta)d\theta$$

Second:

$$\begin{aligned} f(x) &= \int_0^\infty \frac{1}{\sqrt{2\pi}}\theta^{1/2}\exp(-\theta x^2/2)\frac{\lambda^\delta\theta^{\delta-1}\exp(-\lambda\theta)}{\Gamma(\delta)}d\theta \\ &= \frac{\Gamma(\delta+1/2)\lambda^\delta}{\sqrt{2\pi}\Gamma(\delta)(\lambda+x^2/2)^{\delta+1/2}}\int_0^\infty \frac{(\lambda+x^2/2)^{\delta+1/2}\theta^{\delta+1/2-1}\exp(-(\lambda+x^2/2)\theta)}{\Gamma(\delta+1/2)}d\theta \\ &= \frac{\Gamma(\delta+1/2)\lambda^\delta}{\sqrt{2\pi}\Gamma(\delta)(\lambda+x^2/2)^{\delta+1/2}} \end{aligned}$$

because the expression under the integral sign is the density of a Gamma distribution of parameter $\delta+1/2$ and $\lambda+x^2/2$. Therefore:

$$f"(x) = f(x)(-E\tilde{\theta} + x^2E(\tilde{\theta}^2))$$

when $\tilde{\theta}$ is distributed Gamma with parameter $\delta+1/2$ and $\lambda+x^2/2$. Thus:

$$\begin{aligned} \frac{f"(x)}{f(x)} &= -\frac{\delta+1/2}{\lambda+x^2/2} + x^2\frac{(\delta+1/2)(\delta+3/2)}{(\lambda+x^2/2)^2} \\ &= \frac{\delta+1/2}{(\lambda+x^2/2)^2}(-(\lambda+x^2/2) + x^2(\delta+3/2)) \\ &= \frac{\delta+1/2}{(\lambda+x^2/2)^2}(-\lambda + x^2(\delta+1)) \end{aligned}$$

It is a symmetric function. When $x = 0$ it is negative and when x tends to ∞ it is equal to 0. Its derivative is of the same sign that:

$$\begin{aligned} & 2x(\delta+1)(\lambda+x^2/2) - (-\lambda + x^2(\delta+1))2x \\ = & (\delta+1)(2x\lambda + x^3 - 2x^3) + 2x\lambda \\ = & x(2\lambda(\delta+2) - (\delta+1)x^2) \end{aligned}$$

which is equal to zero only once for positive x at $x^2 = 2\lambda(\delta+2)/(\delta+1)$. At that point, $\frac{f''(x)}{f(x)}$ is maximum and equal to:

$$\max_{x>0} \frac{f''(x)}{f(x)} = \frac{\delta + 1/2}{\lambda(2\delta+3)}(\delta+1)^2$$

Futhermore, consider the convolution of that distribution with any symmetric around 0 distribution:

$$f_y(y) = \int f(y-\eta) f_\eta(\eta) d\eta$$

Then if $f''/f < \alpha$ then $f_y''/f_y < \alpha$. Property P3 is not affected by taking convolutions of f.

Proof of a Technical Lemma used in the Proof of Theorem 3.4.

Lemma D.2 *For any $y > 0$*

$$\Psi(y) = \frac{sh(y)}{y(ch(y)-1)} - \frac{2}{y^2} \in]0, 1/6]$$

Proof. First, $\lim_{y\to\infty} \Psi(y) = 0$. We now prove that $\Psi(0) = 1/6$ and that $\Psi'(y) \leq 0$.
i). When $y \to 0$, we can replace hyperbolic functions by their expansions:

$$\text{sh}(y) \approx y + y^3/6 \qquad \text{ch}(y) \approx 1 + y^2/2 + y^4/24$$

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 $\frac{d}{dy}\frac{\text{sh}y}{(\text{ch}y - 1)} = -\frac{1}{\text{ch}y - 1}$:

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

We use that for any $y > 0$:

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

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

and therefore $\Psi'(y) \leq 0$ ∎