Deconvoluting preferences and errors: a model for binomial panel data

Fosgerau, Mogens and Nielsen, Søren Feodor

Technical University of Denmark

2010

DECONVOLUTING PREFERENCES AND ERRORS: A MODEL FOR BINOMIAL PANEL DATA*

Mogens Fosgerau, mf@transport.dtu.dk
Technical University of Denmark

Søren Feodor Nielsen, feodor@stat.ku.dk
Department of Mathematical Sciences
University of Copenhagen

November 3, 2009

Abstract

In many stated choice experiments researchers observe the random variables V_t, X_t, and $Y_t = 1\{U + \delta^\top X_t + \epsilon_t < V_t\}, t \leq T$, where δ is an unknown parameter, and U and ϵ_t are unobservable random variables. We show that under weak assumptions the distributions of U and ϵ_t as well as the unknown parameter δ can be consistently estimated using a sieved maximum likelihood estimation procedure.

KEYWORDS: semi-nonparametric, nonparametric, method of sieves, binomial panel, discrete choice, consistent estimation
JEL codes: C14, C23, C25, D12, Q51, R41

1 INTRODUCTION

Observe a sequence $Y_i = \{Y_{i,t}\}_{t=1,\ldots,T}$ of binary choices for individual $i = 1, \ldots, N$ generated by the model

$$Y_{i,t} = 1\{\delta^\top X_{i,t} + U_i + \epsilon_{i,t} < V_{i,t}\} \quad t = 0, \ldots, T, i = 1, \ldots, N \tag{1}$$

where $\delta^\top X_{i,t} + U_i$ is a preference parameter consisting of a systematic part $\delta^\top X_{i,t}$ which may vary over choices and a random effect U_i representing individual heterogeneity, considered to be constant across the choices of each individual; $Y_{i,t}$, $X_{i,t}$ and $V_{i,t}$ are observed and $\epsilon_{i,t}$ is an observation specific error. We are interested in the situation where $T > 1$ is fixed and $N \to \infty$. The objective of this paper is to show that the distributions of U_i and $\epsilon_{i,t}$ along with the parameter δ are identified from the data $(Y_{i,t}, X_{i,t}, V_{i,t})_{i,t}$ and can be consistently estimated under weak assumptions.

*We are grateful to Bo Honoré, the referees and the co-editor Jinyong Hahn for helpful comments. Mogens Fosgerau has received support from the Danish Social Science Research Council.

A number of approaches are available for the binary model without the panel data dimension, i.e. $Y_i = 1\{\delta^\top X_i + U_i < V_i\}$, see the review in Li and Racine (2007). E.g., Lewbel (2000) shows identification of δ and the distribution of U_i. Honoré and Lewbel (2002) show identifiability of the Euclidean parameter δ in the binary model (1) and give a root-N consistent estimator for this. They do not, however, consider identifiability nor estimation of the unknown distributions of U_i and $\epsilon_{i,t}$. Relative to their paper, we make stronger independence assumptions in order to obtain our results. On the other hand, we are able to relax their support condition for the range of the independent variables, and we do not need instruments.

The paper is structured as follows. The model specification is set out in section 2 and identification of the model is shown in section 3. Consistency of the sieved (seminonparametric) maximum likelihood estimator is established in section 4 with some additional restrictions on the parameter space. Longer proofs are deferred to the appendix. A working paper version of this paper (Fosgerau and Nielsen, 2007) presents some applications of the model to simulated and actual data and the estimator seems to work well.

2 MODEL SPECIFICATION

We parametrise the model in terms of the unknown parameter (δ, f, h) with true values (δ^*, f^*, h^*), where f is the density of $\epsilon_{i,t}$ and h is the density of U_i. We make the following assumptions:

a) $(X_{i,t}, V_{i,t}) \in \mathbb{R}^{d+1}$ are i.i.d., independent of the unobservable random variables $(\epsilon_{i,t}, U_i)$, and with $E[\|X_{i,t}\|] < \infty$ and $E[|V_{i,t}|] < \infty$.

b) $\epsilon_{i,t} \in \mathbb{R}$ are i.i.d. with bounded support and $E[\epsilon_{i,t}] = 0$. When $T = 2$, the $\epsilon_{i,t}$s are also required to be symmetric.

c) The $U_i \in \mathbb{R}$ are i.i.d., independent of $\epsilon_{i,t}$ and with bounded support.

d) The support of $U_i + \epsilon_{i,t}$ is contained in the support of $V_{i,t} - \delta^{*\top} X_{i,t}$.

e) There exists a point (x, v) in the support of $(X_{i,t}, V_{i,t})$ such that the distribution function of $U_i + \epsilon_{i,t}$ is strictly increasing in an open interval containing $v + \delta^{*\top} x$ and such that the set of vectors of the form $x - x'$, where x' is such that $(x, v) + (w, \delta^{*\top} w)$ is contained in the support of $(X_{i,t}, V_{i,t})$, spans $\mathbb{R}^d$.

Assumption c) is a random effects assumption. Assumption d) is weaker than the assumption in Honoré and Lewbel (2002), who require that the support of $U_i + \delta^{*\top} X_{i,t} + \epsilon_{i,t}$ is contained in the support of $V_{i,t}$. Their requirement may be hard to satisfy in practice and it may hence be important to only have the present weaker requirement. Assumption e) is fulfilled if the joint distribution of $(X_{i,t}, V_{i,t})$ has a strictly positive density in a neighbourhood of (x, v) with respect to lebesgue measure but also allows for the case where some components of $X_{i,t}$ are discrete (as when dummy regressors are used). It ensures that $X_{i,t}$ does not contain an intercept term which is necessary for the identifiability of δ and the support of U_i.

The unknown parameters lie in the parameter space $\Delta \times \Phi \times \Gamma$ where Δ is a subset of $\mathbb{R}^d$, Φ is a set of densities with bounded support and mean zero, and Γ is a set of density functions with bounded support. In the case $T = 2$, Φ is a set of symmetric densities with bounded support.

Using (1) the conditional distribution for one individual can be expressed as

$$P(Y_i|V_i,X_i,\delta,f,h)=\int h(u)\prod_{t=1}^{T}\Big[(2Y_{i,t}-1)F(V_{i,t}-\delta^\top X_{i,t}-u)+(1-Y_t)\Big]du, \qquad (2)$$

where F is the distribution function corresponding to the density f.

3 IDENTIFICATION

We start by showing that the model is identified.

Theorem 1. *Under assumptions a)-e), the parameters of the model are identified: If $P(Y|V,X,\delta,f,h)=P(Y|V,X,\delta^*,f^*,h^*)$ then $\delta=\delta^*$, and $(h,f)=(h^*,f^*)$ almost everywhere.*

Proof Letting $y=(1,\dots,1)$ in (2) we obtain

$$P(U+\epsilon_t\le v_t-\delta^\top x_t, t=1,\dots,T)=P(U^*+\epsilon_t^*\le v_t-{\delta^*}^\top x_t, t=1,\dots,T) \qquad \text{for all } v,x$$

where $U\sim h$, $U^*\sim h^*$ and $\epsilon_t\sim f$, $\epsilon_t^*\sim f^*$. Let G denote the distribution function of $U+\epsilon_t$ and G^* the distribution function of $U^*+\epsilon_t^*$. Then we have

$$G(v-\delta^\top x)=G^*(v-{\delta^*}^\top x) \qquad \text{for all } v,x.$$

Now apply assumption e) to pick v and x such that G^* is strictly increasing in an open interval containing $v-{\delta^*}^\top x$. Then G is strictly increasing around $v-\delta^\top x$. For any vector $w\in\mathbb{R}^d$, we put $v_w=v+{\delta^*}^\top w$. Then

$$G(v-\delta^\top x)=G^*(v-{\delta^*}^\top x)=G^*(v_w-{\delta^*}^\top(x+w))=G(v_w-\delta^\top(x+w))=G(v-\delta^\top x-(\delta-\delta^*)^\top w)$$

which implies that $(\delta-\delta^*)^\top w=0$. By assumption e), this implies that $\delta=\delta^*$.

Identifiability of f^* and h^* then follows from Horowitz and Markatou (1996). Fosgerau and Nielsen (2007) present a simpler proof of the latter assertion by showing that $E{\epsilon_1^*}^k=E\epsilon_1^k$ for all k, which implies that $f^*=f$, as the distributions have bounded support. □

4 CONSISTENCY

A standard argument based on Jensen's inequality (see Fosgerau and Nielsen (2007)) shows that (δ^*,f^*,h^*) is the unique maximiser of the expected log-likelihood. In this section we will show that the parameters (δ,f,h) can be consistently estimated by a sieved maximum likelihood estimation procedure, i.e. by maximising the observed conditional log-likelihood

$$l_N(\delta,f,h)=\frac{1}{N}\sum_{i=1}^{N}\log P(y_i|v_i,x_i,\delta,f,h) \qquad (3)$$

over the set $\Delta\times\Phi_N\times\Gamma_N$ where $\Phi_N\subset\Phi$ is chosen so that the closure in L_1-norm of $\cup_N\Phi_N$ is Φ and similarly $\Gamma_N\subset\Gamma$ is chosen so that the closure in L_1-norm of $\cup_N\Gamma_N$ is Γ. See Chen (2006) for an overview of the method of sieves.

For proving consistency, it is useful to fix the supports of the unknown distributions of U_i and $\epsilon_{i,t}$. Multiplying $V_{i,t}$ by a scale parameter γ we can ensure that the smallest interval of the form $[-c; c]$ containing the support of f is the interval [-1;1]; in the case when f is assumed to be a symmetric density we may thus assume that the convex hull of its support is $[-1; 1]$. We include a constant term in the covariate $X_{i,t}$ in order to fix the infimum of the support of h to 0 and introduce a parameter ζ for the maximum of the support such that the convex hull of the support of h is the interval $[0;\zeta]$. Finally, we replace U_i by ζU_i such that the convex hull of the support of U_i is the unit interval. In summary we have

$$Y_{i,t} = 1\{\theta^\top Z_{i,t} > \zeta U_i + \epsilon_{i,t}\} \quad t = 1, \dots, T, i = 1, \dots, N$$

where $Z_{i,t} = (1, X_{i,t}^\top, V_{i,t})^\top$ and $\theta = (\theta_1, -\delta,^\top \gamma)^\top$.

We let $\Theta \subset \mathbb{R}^{d+3}$ denote the parameter set for the Euclidean parameter (θ, ζ). We restrict Γ to consist of densities h with the convex hull of the support equal to the unit interval $[0; 1]$. Similarly, Φ is restricted to densities with convex hull of the support contained in the interval $[-1; 1]$, but not in any shorter interval of the form $[-c; c]$. We equip Θ with the Euclidian norm, while Φ and Γ are equipped with L_1-norms. The whole parameter space $\Sigma = \Theta \times \Gamma \times \Phi$ is equipped with the norm given by the sum of these norms. We let $\sigma = (\theta, f, h)$ denote an element of this parameter space with σ^* denoting the true value and put

$$P(y|z, \sigma) = \int h(u) \prod_{t=1}^{T} \big((2y_t - 1)F(\theta^\top z_t - \rho u) + (1 - y_t)\big) du.$$

We introduce two new assumptions:

f) f^*, h^* are bounded by a given constant K.

g) Θ is a compact subset of $\mathbb{R}^{d+2}$.

Assumption f) bounds the unknown densities to avoid estimators of h and f that are functions of spikes, regardless of the true form. We note that L_1 is a complete metric space and that Γ is closed by construction. As Γ is totally bounded, it is compact. The same argument applies to Φ, in the case $T = 2$ upon noting that the set of symmetric densities is closed. It follows that Σ is compact.

A convenient choice of sieve spaces is obtained by dividing $[0; 1]$ and $[-1; 1]$ into intervals and use densities that are constant on each interval and let the number of intervals increase as $N \to \infty$.

Lemma 1. *There exists a sequence $\sigma_N \in \Sigma_N$ and a constant $C > 1$ such that $\sigma_N \to \sigma^*$ and*

$$\frac{P(y|z, \sigma^*)}{P(y|z, \sigma_N)} \leq C \quad p\text{-a.e. } z \text{ and every } y \text{ for } N \text{ sufficiently large.}$$

See appendix A.1 for a proof. We can now prove consistency.

Theorem 2. *Under assumptions a)-g) the sieved maximum likelihood estimator found by maximising*

$$l_N(\sigma) = \frac{1}{N} \sum_{i=1}^{N} \log P(y_i|z_i, \sigma)$$

over $\Sigma_N = \Theta \times \Gamma_N \times \Phi_N$ is consistent.

Proof Using Lemma 1 we obtain[1]

$$|\log P(y|z,\sigma^*) - \log P(y|z,\sigma_N)| \leq C\frac{|P(y|z,\sigma^*) - P(y|z,\sigma_N)|}{P(y|z,\sigma^*)}.$$

The likelihood $\sigma \to P(Y_i|Z_i,\sigma)$ is Lipschitz continuous by Lemma 2 in appendix A.2. Combining these facts and assumption a) we find that

$$E\left|\frac{1}{N}\sum_{i=1}^{N}\log P(Y_i|Z_i,\sigma_N) - \frac{1}{N}\sum_{i=1}^{N}\log P(Y_i|Z_i,\sigma^*)\right| \to 0 \tag{4}$$

as $\sigma_N \to \sigma^*$. As $\hat{\sigma}_N$ maximises the conditional log-likelihood over Σ_N we have

$$0 \leq \frac{1}{N}\sum_{i=1}^{N}\log P(Y_i|Z_i,\hat{\sigma}_N) - \frac{1}{N}\sum_{i=1}^{N}\log P(Y_i|Z_i,\sigma_N) = \frac{1}{N}\sum_{i=1}^{N}\log\frac{P(Y_i|Z_i,\hat{\sigma}_N)}{P(Y_i|Z_i,\sigma^*)} + o_P(1)$$

by (4). By the concavity of the logarithm

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\log\frac{P(Y_i|Z_i,\hat{\sigma}_N)}{P(Y_i|Z_i,\sigma^*)} &\leq \frac{2}{N}\sum_{i=1}^{N}\log\frac{P(Y_i|Z_i,\hat{\sigma}_N)+P(Y_i|Z_i,\sigma^*)}{2P(Y_i|Z_i,\sigma^*)} \\ &= 2E\left[\log\frac{P(Y|Z,\sigma)+P(Y|Z,\sigma^*)}{2P(Y|Z,\sigma^*)}\right]_{|\sigma=\hat{\sigma}_N} + o_P(1)\end{aligned} \tag{5}$$

by the uniform law of large numbers; the proof of this is somewhat involved and we defer it to Appendix A.3. Furthermore,

$$\begin{aligned}&E\left[\log\frac{P(Y|Z,\sigma)+P(Y|Z,\sigma^*)}{2P(Y|Z,\sigma^*)}\right]_{|\sigma=\hat{\sigma}_N} \\ &\qquad \leq 2E\left[\sqrt{\frac{P(Y|Z,\sigma)+P(Y|Z,\sigma^*)}{2P(Y|Z,\sigma^*)}} - 1\right]_{|\sigma=\hat{\sigma}_N} = -h^2(\hat{\sigma}_N,\sigma^*)\end{aligned}$$

where

$$h^2(\sigma,\sigma^*) = \int\sum_{y\in\{0,1\}^T}\left(\left(\frac{P(y|z,\sigma)+P(y|z,\sigma^*)}{2}\right)^{1/2} - P(y|z,\sigma^*)^{1/2}\right)^2 p(z)dz.$$

Thus $0 \leq h^2(\hat{\sigma}_N,\sigma^*) \leq o_P(1)$. Hence

$$\frac{1}{N}\sum_{i=1}^{N}\log P(Y_i|Z_i,\sigma_N) \leq \frac{1}{N}\sum_{i=1}^{N}\log P(Y_i|Z_i,\hat{\sigma}_N) \leq \frac{1}{N}\sum_{i=1}^{N}\log P(Y_i|Z_i,\sigma^*) + o_P(1)$$

which by (4) and the law of large numbers implies that

$$\frac{1}{N}\sum_{i=1}^{N}\log P(Y_i|Z_i,\hat{\sigma}_N) \to E[\log P(Y|Z,\sigma^*)].$$

[1]For $0 < x < C$ $|\log x| \leq C|1-\frac{1}{x}|$ when $C > 1$: For $x > 1$ this follows since $\log x < x - 1$ and $x < C$. For $x \leq 1$ the inequality follows since $\log\frac{1}{x} \leq \frac{1}{x} - 1$.

Now by compactness of Σ, every subsequence of $(\hat{\sigma}_N)_N$ has a further subsequence $(\hat{\sigma}_{N_j})_j$ which converges; let $\tilde{\sigma}$ denote the limit of this subsequence. Then, as $(a^{1/2}-b^{1/2})^2 \leq |a-b|$ for $a, b \geq 0$,

$$h^2(\hat{\sigma}_{N_j}, \sigma^*) \leq \int \sum_{y\in\{0,1\}^T} \left| \frac{P(y|z,\sigma) + P(y|z,\sigma^*)}{2} - P(y|z,\sigma^*) \right| p(z)dz_{\big|\sigma=\hat{\sigma}_{N_j}} \to 0$$

by lemma 2. Hence we get

$$h^2(\tilde{\sigma}, \sigma^*) \leq 2h^2(\tilde{\sigma}, \hat{\sigma}_{N_j}) + 2h^2(\hat{\sigma}_{N_j}, \sigma^*) = o_P(1).$$

By the identifiability (Theorem 1) this implies that $\tilde{\sigma} = \sigma^*$. Hence $\hat{\sigma}_N$ is consistent in the norm on Σ. □

A Appendix

A.1 Proof of Lemma 1

Start by choosing $\theta_N = \theta^*$. Recall that F_N is a piecewise linear function. We choose it so that it is at least as large as F^* when $F^*(x)$ is small and no larger than F^* when $F^*(x)$ is large. To be precise, for some $0 < \alpha < 1/2$ let $q_\alpha = \inf\{x : F^*(x) = \alpha\}$ and $q_{1-\alpha} = \sup\{x : F^*(x) = 1-\alpha\}$ and choose F_N such that $F_N(x) \geq F^*(x)$ for $x \leq q_\alpha$ and $F_N(x) \leq F^*(x)$ for $x \geq q_{1-\alpha}$. Then $F_N(x) \geq F^*(x)\alpha$ and $1 - F_N(x) \geq (1 - F^*(x))\alpha$ for all x. Hence

$$P(y|z, \sigma_N) \geq \alpha^T \cdot P\big(y\big|z, \sigma = (\theta^*, f^*, h_N)\big)$$

Next choose h_N such that $h_N(u) \geq bh^*(u)$ for some constant b. Letting $I_{N,k}$ denote intervals where h_N is constant, we put

$$h_N(u) = \frac{\max_{I_{N,k}} h^*(u) + \min_{I_{N,k}} h^*(u)}{2c_N} \qquad u \in I_{N,k}$$

Here c_N is a constant ensuring that h_N is a density; it is the value of an approximating sum to the integral of h^* and hence converges to 1. It now follows that

$$h^*(u) \leq \max_{I_{N,k}} h^*(u) = h_N(u) \cdot \frac{\max_{I_{N,k}} h^*(u)}{\max_{I_{N,k}} h^*(u) + \min_{I_{N,k}} h^*(u)} \cdot 2c_N \leq h_N(u) \cdot 2 \max_N c_N$$

Hence

$$P(y|z, \sigma_N) \geq \alpha^T \cdot P\big(y\big|z, \sigma = (\theta^*, f^*, h_N)\big) \geq \alpha^T 2 \max_N c_N \cdot P(y|z, \sigma^*).$$

Hence Lemma 1 holds with $C = 1/(\alpha^T 2 \max_N c_N)$. □

A.2 Continuity

Lemma 2. *The likelihood $\sigma \to P(Y_i|Z_i, \sigma)$ is Lipschitz continuous.*

Proof We start by noting that for any density $f \in \Phi$, the corresponding distribution function F is Lipschitz with parameter K. Hence we have

$$|F(z) - \tilde{F}(\tilde{z})| \leq K|z - \tilde{z}| + \sup_{z \in \mathbb{R}} |F(z) - \tilde{F}(z)| \leq K|z - \tilde{z}| + \|f - \tilde{f}\|_1.$$

Putting $a_{i,t} = (2Y_{i,t} - 1)F(\theta^\top Z_{i,t} - \zeta U_i) + (1 - Y_{i,t})$ and $\tilde{a}_{i,t} = (2Y_{i,t} - 1)\tilde{F}(\tilde{\theta}^\top Z_{i,t} - \tilde{\zeta} U_i) + (1 - Y_{i,t})$ we see that

$$\begin{aligned}\big|P(Y_i|Z_i, \sigma) - P(Y_i|Z_i, \tilde{\sigma})\big| \leq & \int |h(u) - \tilde{h}(u)| \prod_{t=1}^{T} \tilde{a}_{i,t} du + \int \left| \prod_{t=1}^{T} a_{i,t} - \prod_{t=1}^{T} \tilde{a}_{i,t} \right| h(u) du \\ \leq & \|h - \tilde{h}\|_1 + \sum_{t=1}^{T} \int |a_{i,t} - \tilde{a}_{i,t}| h(u) du \\ \leq & \|h - \tilde{h}\|_1 + T\|f - \tilde{f}\|_1 + TK|\zeta - \tilde{\zeta}| + K \sum_{t=1}^{T} |(\theta - \tilde{\theta})^\top Z_{i,t}|\end{aligned}$$

as $0 \leq a_{i,t}, \tilde{a}_{i,t} \leq 1$ and $E[U_i] \leq 1$. □

A.3 Uniform law of large numbers

We wish to show a uniform (in $\sigma \in \Sigma$) law of large number for the right hand side of (5). We will do this by applying Theorem 2.4.3 in van der Vaart and Wellner (1996).

Γ is by construction a subset of a VC-hull class (van der Vaart and Wellner, 1996, Corollary 2.6.12), and it follows from Problem 2.6.14, Lemma 2.6.19 and Lemma 2.6.20 in van der Vaart and Wellner (1996) that the class

$$\big\{(y, z, u) \to h(u) \big[(2y - 1)F(\theta^\top z - \zeta u) + (1 - y)\big] : (\theta, \zeta) \in \Theta, F \in \Phi, h \in \Gamma\big\} \tag{6}$$

is a subset of a VC-hull class. In particular, its covering number is bounded by a constant times a power of $1/\varepsilon$. Repeated use of Lemma 2.6.20 of van der Vaart and Wellner (1996) allows us to extend this class of functions to reflect the fact that $T > 1$ in our model. However, to keep notation simple we do not do this here.

Now consider the function class

$$\mathcal{G} = \left\{(y, z) \to \int h(u)F(\theta^\top z - \zeta u) du : (\theta, \zeta) \in \Theta, F \in \Phi, h \in \Gamma\right\} \tag{7}$$

Let $g_1, \dots, g_k$ be centres for the class (6) corresponding to the L_1-norm with respect to the product of an arbitrary probability measure μ and the Lebesgue measure on $[0; 1]$ for a chosen $\epsilon > 0$. Then for any choice of $h \in \Gamma$, $F \in \Phi$ and $(\theta, \zeta) \in \Theta$ we have

$$\begin{aligned}&\int \Big| \int h(u)F(\theta^\top z - \zeta u) du - \int g_j(v, x, u) du \Big| d\mu(v, x) \\ \leq &\iint |h(u)F(\theta^\top z - \zeta u) - g_j(v, x, u)| du d\mu(v, x)\end{aligned}$$

Hence the covering number of the class $\mathcal{G}$ (7) is at most as large as the covering number of the class (6).

Let $P_j(y|z)$ denote the centres for $\mathcal{G}$ corresponding to the covering of size ε for the norm

$$\|g\| = \frac{\frac{1}{N}\sum_{i=1}^N \frac{|g(Y_i, Z_i)|}{P(Y_i|Z_i, \sigma^*)}}{\frac{1}{N}\sum_{i=1}^N 1/P(Y_i|Z_i, \sigma^*)}.$$

and consider the class

$$\mathcal{G}' = \left\{ (y, z) \to \log\left(\frac{P(y|z, \sigma) + P(y|z, \sigma^*)}{2P(y|z, \sigma^*)}\right) : \sigma \in \Sigma \right\} \tag{8}$$

Now

$$\left|\log\left(\frac{P(y|z, \sigma) + P(y|z, \sigma^*)}{2P(y|z, \sigma^*)}\right) - \log\left(\frac{P_j(y|z) + P(y|z, \sigma^*)}{2P(y|z, \sigma^*)}\right)\right| \leq \frac{|P(y|z, \sigma) - P_j(y|z)|}{P(y|z, \sigma^*)}$$

To show that the covering number for (8) is polynomial in ε, we bound the relevant distance by:

$$\frac{1}{N}\sum_{i=1}^N \frac{|P(Y_i|Z_i, \sigma) - P_j(Y_i|Z_i, \sigma^*)|}{P(Y_i|Z_i, \sigma^*)} \leq \varepsilon \cdot \frac{1}{N}\sum_{i=1}^N \frac{1}{P(Y_i|Z_i, \sigma^*)}$$

Noting that

$$\frac{1}{N}\sum_{i=1}^N \frac{1}{P(Y_i|Z_i, \sigma^*)} \to \int \sum_{y \in \{0,1\}^T} \frac{P(y|z, \sigma^*)}{P(y|z, \sigma^*)} p(z) dz = 2^T \quad \text{almost surely}$$

it follows that random entropy condition of Theorem 2.4.3 of van der Vaart and Wellner (1996) is satisfied.

To verify the required envelope condition, we note that

$$\log\frac{1}{2} \leq \log\left(\frac{P(y|z, \sigma) + P(y|z, \sigma^*)}{2P(y|z, \sigma^*)}\right) \leq -\log P(y|z, \sigma^*)$$

which provides us with the integrable envelope $G(y, z) = \log 2 - \log P(y|z, \sigma^*)$ for $\mathcal{G}'$ given by (8) as

$$E[|\log P(Y|V, X, \sigma^*)|] \leq E[1/P(Y|V, X, \sigma^*)] = 2^T < \infty.$$

What now remains for the application of Theorem 2.4.3 in van der Vaart and Wellner (1996) is to argue that the class $\mathcal{G}$ is measurable (van der Vaart and Wellner, 1996, Definition 2.3.3). However this follows from the fact that functions in $\mathcal{G}$ may be approximated pointwise by functions from a countable subset of $\mathcal{G}$ constructed by considering functions obtained when (θ, ζ) lies in a countable dense subset of Θ, h and f are given by piecewise constant densities with rational values and rational discontinuity points.

References

Chen, X. (2006) Large Sample Sieve Estimation of Semi-Nonparametric Models *Handbook of Econometrics* forthcoming edn.

Fosgerau, M. and Nielsen, S. F. (2007) Deconvoluting preferences and errors: a model for binomial panel data *Munich Personal RePEc Archive* (3950).

Honoré, B. E. and Lewbel, A. (2002) Semiparametric binary choice panel data models without strictly exogenous regressors *Econometrica* **70**(5), 2053–2063.

Horowitz, J. and Markatou, M. (1996) Semiparametric Estimation of Regression Models for Panel Data *The Review of Economic Studies* **63**(1), 145–168.

Lewbel, A. (2000) Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables *Journal of Econometrics* **97**(1), 145–177.

Li, Q. and Racine, J. S. (2007) *Nonparametric Econometrics: Theory and Practice* Princeton University Press Princeton and Oxford.

van der Vaart, A. and Wellner, J. (1996) *Weak convergence and empirical processes* Springer Series in Statistics 1 edn Springer.