Deconvoluting preferences and errors: a model for binomial panel data

Fosgerau, Mogens and Nielsen, Søren Feodor

Technical University of Denmark

9 July 2007

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

Mogens Fosgerau, mf@dtf.dk
Technical University of Denmark

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

July 9, 2007

Abstract

Let U be an unobserved random variable with compact support and let ϵ_t be unobserved i.i.d. random errors also with compact support. Observe the random variables V_t, X_t, and $Y_t = 1\{U + \delta X_t + \epsilon_t < V_t\}, t \leq T$, where δ is an unknown parameter. This type of model is relevant for many stated choice experiments. It is shown that under weak assumptions on the support of $U + \epsilon_t$, 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. The model is applied to simulated data and to actual data designed for assessing the willingness-to-pay for travel time savings.

KEYWORDS: semi-nonparametric, nonparametric, method of sieves, binomial panel, willingness-to-pay, value of time
JEL codes: C14, C23, C25, D12, Q51, R41

*We are grateful to Bo Honoré for taking time to discuss this work. Mogens Fosgerau has received support from the Danish Social Science Research Council.

1 INTRODUCTION

Observe a sequence $Y_n = \{Y_{nt}\}_{t=1,\dots,T}$ of binary choices for individual $n = 1, \dots, N$ generated by the model

$$Y_{nt} = 1\{\delta X_{nt} + U_n + \epsilon_{nt} < V_{nt}\} \quad t = 0, \dots, T, n = 1, \dots, N \tag{1}$$

where $\delta X_{nt} + U_n$ is a preference parameter consisting of a systematic part δX_{nt} which may vary over choices, $x \in \mathbb{R}^d$ and a random part $U_n \in \mathbb{R}$ representing individual heterogeneity, considered to be constant across the choices of each individual; V_{nt} is set by design for each choice situation t and ϵ_{nt} 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_n and ϵ_{nt} can be consistently estimated using semi-nonparametric methods, given some not very restrictive conditions.

Horowitz and Markatou (1996) provide nonparametric estimates of the densities of U_n and ϵ_{nt} in the linear model $Y_{nt} = \delta X_{nt} + U_n + \epsilon_{nt}$. 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, however, not consider identifiability nor estimation of the unknown distributions of U_n and ϵ_{nt}. A number of approaches are available for the binary model without the panel data dimension, i.e. $Y_n = 1\{\delta X_n + U_n < V_n\}$. See the review in Li and Racine (2007).

An application for the binary model, with or without panel data, is the estimation of a willingness-to-pay (WTP) distribution from experimental stated choice survey data (e.g. Hanemann and Kanninen 1998). Individuals are often thought not to be able to explicitly state their WTP for a good; instead they are asked whether their WTP is larger or smaller than some bid. Lewbel, Linton and McFadden (2002) consider the nonparametric estimation of moments and quantiles of a WTP distribution from such binomial data. They do not however consider panel data. Lewbel et al. (2002) recognise that the distribution that they measure may be a convolution of the target WTP distribution and psychometric errors (ϵ_{nt} above) but state that the "difficult problem of deconvoluting a target distribution in the presence of psychometric errors is left for future research." In fact it seems inconsistency and hence error is a prominent feature of such data (see Sælensminde 2001, Sælensminde 2002, Rouwendal and de Blaeijj 2004). This paper deals with the situation when variation is created both by heterogenous preferences and by errors.

Gabler, Laisney and Lechner (1993) discuss the application of the method of sieves to binary choice models. Following Gallant and Nychka (1987), such models are often called seminonparametric (SNP). Chen and Randall (1997) estimate a SNP binary response model to obtain the WTP for an environmental quality improvement where both the random and the systematic variation in WTP are estimated by a SNP form. Belluzzo (2004) compares competing semi-parametric methods for recovering the WTP distribution from binomial data. Neither use panel data and so they do not take errors into account.

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. Section 5 presents some examples of applications of the model and the estimator, first to simulated data and second to survey data designed for assessing the distribution of the WTP for travel time savings. Section 6 provides a few concluding remarks. Longer proofs are deferred to the appendix.

2 MODEL SPECIFICATION

We parametrise the model in terms of (δ, f, h) with true values (δ^*, f^*, h^*), where f is the density of ϵ_{nt} and h is the density of U_n. We make the following assumptions:

a) ϵ_{nt} are i.i.d. with bounded support and $E(\epsilon_{nt}) = 0$. When $T = 2$, ϵ_{nt} are also required to be symmetric.

b) The U_n are i.i.d., independent of ϵ_{nt} and with bounded support.

c) The support of $U_n + \epsilon_{nt}$ is contained in the support of $V_{nt} - \delta^* X_{nt}$.

d) (X_{nt}, V_{nt}) are i.i.d. with $E[\|X_{nt}\|] < \infty$ and $E[|V_{nt}|] < \infty$, independent of the unobservable random variables (ϵ_{nt}, U_n). We let p denote this distribution.

e) There exists a set in the support of X_{nt} with positive probability and with non-empty interior of its convex hull.

Assumption c) is weaker than the assumption in Honoré and Lewbel (2002), who require that the support of $U_n + \delta^* X_{nt} + \epsilon_{nt}$ is contained in the support of V_{nt}. This requirement may be hard to satisfy in practice and it may hence be important to only have the present weaker requirement. The distribution of V_{nt} may be chosen to depend on X_{nt} and may be chosen such that assumption c) holds.

Assumption e) may seem somewhat technical. It ensures that $\delta X_{n,t}$ is not constant unless $\delta = 0$. In particular, it ensures that $X_{n,t}$ does not contain an intercept term which is essential for the identifiability of δ and the support of U_n.

The unknown parameters (δ^*, f^*, h^*) 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.

We use the notational shorthand $y_n = \{y_{n1}, \ldots, y_{n,T}\}$, etc. Using (1) the conditional likelihood of an observation can be expressed as

$$\begin{aligned} L_{Y|X,V}(\delta, f, h) &= P(Y|V, X, \delta, f, h) \\ &= \int h(u) \prod_{t=1}^{T} \Big[(2Y_t - 1)F(V_{nt} - \delta X_{nt} - u) + (1 - Y_t) \Big] du, \end{aligned} \tag{2}$$

where F is the distribution function corresponding to the density f. Note that assumption c) implies that the distribution of $(U_n + \epsilon_{nt})_{t=1,\ldots,T}$ is identified from the likelihood (2) as this is equal to the joint distribution function of $(U_n + \epsilon_{nt})_{t=1,\ldots,T}$ when all $Y_{nt} = 1$.

3 IDENTIFICATION

We start by showing that the model is identified i.e. that the chosen parametrisation of our model is injective. We do this by showing that the true parameter (δ^*, f^*, h^*) is the unique value of the parameter which maximises the expected conditional log-likelihood

$$\begin{aligned} \bar{l}(\delta, f, h) &= E^*[\log P(Y|V, X, \delta, f, h)] \\ &= \int \sum_{y \in \{0,1\}^T} \log P(y|v, x, \delta, f, h) \cdot P(y|v, x, \delta^*, f^*, h^*) dp(v, x) \end{aligned} \tag{3}$$

with the expectation, E^*, evaluated under the true parameter (δ^*, f^*, h^*).

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 Observe that since $|\log x| \leq 1/x$ for $x \in]0; 1]$ we have

$$\begin{aligned} E^*[|\log P(Y|V, X, \delta^*, f^*, h^*)|] &\leq E^*[1/P(Y|V, X, \delta^*, f^*, h^*)] \\ &= \int \sum_{y \in \{0,1\}^T} 1 \, dp(v, x) = 2^T < \infty \end{aligned}$$

so that the expected conditional log-likelihood is finite at the true value of the unknown parameter.

For any $(\delta, f, h) \in \Delta \times \Phi \times \Gamma$ we get

$$\begin{aligned}
&\bar{l}(\delta, f, h) - \bar{l}(\delta^*, f^*, h^*) \\
&= \int \sum_{y \in \{0,1\}^T} \log \left[\frac{P(y|v, x, \delta, f, h)}{P(y|v, x, \delta^*, f^*, h^*)} \right] P(y|v, x, \delta^*, f^*, h^*) dp(v, x) \\
&\leq \int \log \sum_{y \in \{0,1\}^T} \frac{P(y|v, x, \delta, f, h)}{P(y|v, x, \delta^*, f^*, h^*)} P(y|v, x, \delta^*, f^*, h^*) dp(v, x) \\
&= \int \log \sum_{y \in \{0,1\}^T} P(y|v, x, \delta, f, h) 1\{P(y|v, x, \delta^*, f^*, h^*) > 0\} dp(v, x) \\
&\leq \int \log(1) dp(v, x) = 0
\end{aligned}$$

by Jensen's inequality with strict inequality unless for all y

$$P(y|v, x, \delta, f, h) = P(y|v, x, \delta^*, f^*, h^*), \quad p(v, x) \text{ a.e.}$$

Letting $y = (1, \dots, 1)$ this implies that

$$P(U + \epsilon_t \leq v_t - \delta x_t, t \leq T) = P(U^* + \epsilon_t^* \leq v_t - \delta^* x_t, t \leq T), \ p(v, x) \text{ a.e.}$$

where $U \sim h$, $U^* \sim h^*$ and $\epsilon_t \sim f$, $\epsilon_t^* \sim f^*$. By assumption c) we can identify the distribution of $U^* + \epsilon_t^*$ over the whole of its support for all t. By the above equality we can similarly identify the distribution of $U + \epsilon_t$. We note, however, that due to the presence of δ^* and δ, we do not yet have $U + \epsilon_t \sim U^* + \epsilon_t^*$. Rearranging yields

$$P(U + \delta x_t + \epsilon_t \leq v_t, t \leq T) = P(U^* + \delta^* x_t + \epsilon_t^* \leq v_t, t \leq T), \ p(v, x) \text{ a.e.}$$

and the distributions on both sides of this equation are identified.

As the distributions of $U + \delta x_t + \epsilon_t$ and $U^* + \delta^* x_t + \epsilon_t^*$ are the same, they share the same moments. In particular

$$\delta^* x_t + E^*[U^* + \epsilon_t^*] = \delta x_t + E^*[U + \epsilon_t]$$

Varying x_t shows that $\delta = \delta^*$ (and $E^*[U^* + \epsilon_t^*] = E^*[U + \epsilon_t]$). Consequently $(U + \epsilon_t)_{t=1,\dots,T}$ and $(U^* + \epsilon_t^*)_{t=1,\dots,T}$ have the same distribution.

Identifiability of f^* and h^* then follows from Horowitz and Markatou (1996). We shall 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 $\epsilon_1^* \stackrel{\mathcal{D}}{=} \epsilon_1$, as the distributions have bounded support.

Given identifiability of δ^*, the joint distribution of $(U^* + \epsilon_1^*, \ldots, U^* + \epsilon_T^*)$ is identical to the joint distribution of $(U + \epsilon_1, \ldots, U + \epsilon_T)$. It follows that the distribution of $\epsilon_1 - \epsilon_2 = (U + \epsilon_1) - (U + \epsilon_2)$ equals that of $\epsilon_1^* - \epsilon_2^*$. Denote by m_k the k'th moment of ϵ_{nt}. Consider first the case where $T = 2$ and the distribution of ϵ_1^* is assumed symmetric. In this case we have

$$E(\epsilon_1 - \epsilon_2)^k = m_k(1 + (-1)^k) + q_k(m_1, \ldots, m_{k-1}) \tag{4}$$

for some function q_k with a similar expression for $E(\epsilon_1^* - \epsilon_2^*)^k$. By induction it follows from (4) that $E\epsilon_1^{*k} = E\epsilon_1^k$ when k is even and since all odd moments are zero for a symmetric distribution, $f = f^*$. In the case $T > 2$ we may dispense with the assumption of symmetry. We observe that $\epsilon_1^* + \epsilon_2^* - 2\epsilon_3^* \stackrel{\mathcal{D}}{=} \epsilon_1 + \epsilon_2 - 2\epsilon_3$ and that

$$E(\epsilon_1 + \epsilon_2 - 2\epsilon_3)^k = m_k(2 + (-2)^k) + q(m_1, \ldots, m_{k-1}) \tag{5}$$

By assumption a) $m_1 = 0$ so by induction we may conclude that $E\epsilon_1^{*k} = E\epsilon_1^k$ for all k implying that $f = f^*$.

The identifiability of h^* follows immediately, since by independence of U^* and ϵ_t^* we have:

$$E[(U^* + \epsilon_t^*)^k] = \tilde{m}_k + \tilde{q}(\tilde{m}_{k-1}, \ldots, \tilde{m}_1; m_k, \ldots m_1)$$

where $\tilde{m}_k = EU^k$. By induction, we see that all moments of h^* are identified and hence h^* is identified. $\square$

4 CONSISTENCY

4.1 Estimation

In this section we shall demonstrate that the parameters (δ, f, h) can be consistently estimated by a sieved maximum likelihood estimation procedure. Thus we estimate (δ, f, h) by maximising the observed conditional log-likelihood

$$l_N(\delta, f, h) = \frac{1}{N} \sum_{n=1}^{N} \log P(y_n | v_n, x_n, \delta, f, h) \tag{6}$$

over the set $\Delta \times \Phi_N \times \Gamma_N$ where $\Phi_N \subseteq \Phi$ 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. We define Φ_N and Γ_N in a straight-forward manner using piecewise constant approximations for the densities h and f. One could use other approximations; indeed our proof of consistency below will show consistency of a large number of approximating spaces, subject to some mild conditions. In particular, Φ_N may be replaced by any set of (uniformly bounded) densities corresponding to distributions with mean 0 and, if $T = 2$, that are symmetric. However, the piecewise constant approximations are easy to work with, the necessary conditions are easily imposed and unlike many other approximations –polynomials, splines– they do not, in our point of view, suggest a functional form which is more a function of the approximation than a function of the data. Our estimator has the form of a histogram, which is easy to understand, and which is not easily over-interpreted.

In order to estimate the unknown parameters it is useful to fix the supports of the unknown distributions of U_n and ϵ_{nt}. By reparametrising the model in the manner described below, we may restrict Γ to consist of densities h with the convex hull of the support of h equal to the unit interval $[0; 1]$. Similarly, Φ is restricted to densities f with convex hull of the support of f contained in the interval $[-1; 1]$. This is obtained as follows:

Multiplying V_{nt} by a scale parameter γ ensures that 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_{nt} 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;ζ]. Finally, we replace U_n by ζU_n such that the convex hull of the support of U_n is the unit interval.

In summary we have

$$Y_{nt} = 1\{\theta Z_{nt} > \zeta U_n + \epsilon_{nt}\} \quad t = 1, \dots, T, n = 1, \dots, N \tag{7}$$

where $Z_{nt} = (1, X_{nt}, V_{nt})$ and $\theta = (\theta_1, -\delta, \gamma)$. We let Θ denote the parameter set for the Euclidean parameter (θ, ζ).

Given M_h, the number of "parameters" in the approximation of h, we divide the unit interval into $M_h + 1$ intervals $I_{0,M_h}, \dots, I_{M_h,M_h}$ of equal length and assign to each interval a positive mass λ_{m,M_h}. Then the approximation to h is defined by

$$h_{M_h}(u) = (M_h + 1) \sum_{m=0}^{M_h} \lambda_{m,M_h} 1\{u \in I_{m,M_h}\}$$

where λ_{m,M_h}, $m = 0, ..., M_h$ are positive, smaller than $K/(M_h + 1)$ for some

K not depending on M_h, and sum to 1. This approximation is bounded by K. When $M_h = 0$, $h_0(u)$ reduces to the uniform distribution on the unit interval. We let Γ_N be the set of such functions for some M_h depending on N with $M_h \to \infty$.

In Lemma 2 in appendix A.1 it is shown that $\Gamma = cl_{L_1}(\cup_M \Gamma_M)$ with closure in L_1-norm may be identified with the space of densities on the unit interval bounded by K: If h is such a bounded density there exists a sequence $h_m \in \cup_M \Gamma_M$ with $\|h_m - h\|_1 \to 0$.

We use a similar approximation to f:

$$f_{M_f}(x) = \begin{cases} \dfrac{M_f+1}{2} \displaystyle\sum_{m=0}^{M_f} \lambda'_{m,M_f} \Big(1\{x \in I_{m,M_f}\} + 1\{-x \in I_{m,M_f}\} \Big) & \text{for } T = 2 \\ (M_f+1) \displaystyle\sum_{m=0}^{M_f} \lambda'_{m,M_f} 1\{2x - 1 \in I_{m,M_f}\} & \text{for } T > 2 \end{cases}$$

with λ'_{m,M_f} positive, smaller than $K'/(M_f+1)$ for some K' and summing to 1. Without the assumption of symmetry, we further need to impose the condition

$$\frac{1}{M_f+1} \sum_{m=0}^{M_f} \lambda'_{m,M_f} \frac{2m+1}{2} = 1$$

to ensure that the mean of the distribution is 0. For $M_f = 0$ this is the uniform distribution on $[-1;1]$ and $cl_{L_1}(\cup_M \Phi_M)$ may be identified with the space of (symmetric, if $T = 2$) densities on $[-1,1]$ bounded by K'. We define Φ_N to be the set of such densities for some M_f depending on N with $M_f \to \infty$.

This choice of Γ and Φ does restrict the model somewhat compared to what we discussed in section 3 as we have now imposed a bound on the unknown densities. Some sort of regularisation clearly is necessary to ensure consistent estimation. Otherwise one would expect the estimators of h and f to be functions of spikes, regardless of the true form. Note also that we have fixed the support of these distributions but that this is not a new assumption but a result of the re-parametrisation. Finally we need to restrict attention to the case where Θ is a compact set which is an additional restriction on the supports of the distributions of ϵ_{nt} and U_n compared to section 2. Thus we assume

f) f^*, h^* are bounded by given constants K and K'.

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

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. Considering Γ we note that L_1 is a complete metric space and that Γ is closed by construction. As it is a VC hull class of functions (see appendix A.2) it is totally bounded and thus 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.

4.2 Proof of consistency

In order to prove consistency we need two further assumptions.

h) There exists a sequence $\sigma_N \in \Sigma_N$ such that $\sigma_N \to \sigma^*$ and

$$\frac{P(y|z,\sigma^*)}{P(y|z,\sigma_N)} \leq C \quad \text{p-a.e. } z \text{ for some } C. \tag{8}$$

Assumption h) is an assumption on the rate of approximation of Σ by Σ_N. In particular, should $\sigma^* \in \Sigma_N$ for all sufficiently large N, then the assumption holds with $C = 1$. We can now prove the following theorem.

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

$$l_N(\sigma) = \frac{1}{N} \sum_{n=1}^{N} \log P(y_n|z_n, \sigma)$$

over $\Sigma_N = \Theta \times \Gamma_{M(N)} \times \Phi_{M(N)}$ *is consistent.*

The following lemma established Lipschitz continuity of the likelihood function.

Lemma 1. *The likelihood* $P(Y_n|Z_n, \sigma)$ *is Lipschitz continuous.*

$$\begin{aligned}&\left|P(Y_n|Z_n,\sigma) - P(Y_n|Z_n,\tilde{\sigma})\right| \\ &\qquad \leq \|h - \tilde{h}\|_1 + T\|f - \tilde{f}\|_1 + TK'|\zeta - \tilde{\zeta}| + K' \sum_{t=1}^{T} |Z_{nt} \cdot (\theta - \tilde{\theta})|\end{aligned}$$

The proof is given in appendix A.1.

Proof of Theorem 2 Without loss of generality we may assume that the constant C in the assumption h) is larger than 1. Then we have

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

This implies by Lemma 1 that

$$E^*\left|\frac{1}{N}\sum_{n=1}^{N}\log P(Y_n|Z_n,\sigma_N) - \frac{1}{N}\sum_{n=1}^{N}\log P(Y_n|Z_n,\sigma^*)\right| \to 0 \tag{9}$$

as $\sigma_N \to \sigma^*$ since $E\|Z_t\| < \infty$. Hence

$$\frac{1}{N}\sum_{n=1}^{N}\log P(Y_n|Z_n,\sigma_N) \xrightarrow{L_1} E^*[\log P(Y|Z,\sigma^*)] \tag{10}$$

As $\hat{\sigma}_N$ maximises the conditional log-likelihood over Σ_N we have

$$\begin{aligned}
0 \leq &\frac{1}{N}\sum_{n=1}^{N}\log P(Y_n|Z_n,\hat{\sigma}_N) - \frac{1}{N}\sum_{n=1}^{N}\log P(Y_n|Z_n,\sigma_N)\\
= &\frac{1}{N}\sum_{n=1}^{N}\log\frac{P(Y_n|Z_n,\hat{\sigma}_N)}{P(Y_n|Z_n,\sigma^*)}\\
&- \left(\frac{1}{N}\sum_{n=1}^{N}\log P(Y_n|Z_n,\sigma_N) - \frac{1}{N}\sum_{n=1}^{N}\log P(Y_n|Z_n,\sigma^*)\right)\\
= &\frac{1}{N}\sum_{n=1}^{N}\log\frac{P(Y_n|Z_n,\hat{\sigma}_N)}{P(Y_n|Z_n,\sigma^*)} + o_P(1)
\end{aligned}$$

by (9). By the concavity of the logarithm

$$\begin{aligned}
\frac{1}{N}\sum_{n=1}^{N}\log\frac{P(Y_n|Z_n,\hat{\sigma}_N)}{P(Y_n|Z_n,\sigma^*)} \leq &\frac{2}{N}\sum_{n=1}^{N}\log\frac{P(Y_n|Z_n,\hat{\sigma}_N) + P(Y_n|Z_n,\sigma^*)}{2P(Y_n|Z_n,\sigma^*)}\\
= &\frac{2}{N}\sum_{n=1}^{N}\Bigg(\log\frac{P(Y_n|Z_n,\hat{\sigma}_N) + P(Y_n|Z_n,\sigma^*)}{2P(Y_n|Z_n,\sigma^*)}\\
&\qquad - E^*\left[\log\frac{P(Y|Z,\hat{\sigma}_N) + P(Y|Z,\sigma^*)}{2P(Y|Z,\sigma^*)}\right]\Bigg)\\
&+ E^*\left[\log\frac{P(Y|Z,\hat{\sigma}_N) + P(Y|Z,\sigma^*)}{2P(Y|Z,\sigma^*)}\right]
\end{aligned}$$

The first term goes to 0 by the uniform law of large numbers; the proof of this is somewhat involved and we defer it to Appendix A.2. The second term may be bounded as follows:

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

where $h(\hat{\sigma}_N,\sigma^*)$ is the Hellinger distance between $(p(y|z,\sigma)+p(y|z,\sigma^*))/2$ and $p(y|z,\sigma^*)$ given by

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

Thus we obtain the inequality

$$\begin{aligned}
0 \leq& h^2(\hat{\sigma}_N,\sigma^*) \\
=&\frac{1}{2}\int\sum_{y\in\{0,1\}^T}\left(\left(\frac{P(y|z,\hat{\sigma}_N)+P(y|z,\sigma^*)}{2}\right)^{1/2}-P(y|z,\sigma^*)^{1/2}\right)^2 p(z)dz \\
\leq&\frac{2}{N}\sum_{n=1}^{N}\Bigg(\log\frac{P(Y_n|Z_n,\hat{\sigma}_N)+P(Y_n|Z_n,\sigma^*)}{2P(Y_n|Z_n,\sigma^*)} \\
&\qquad\qquad -E^*\left[\log\frac{P(Y|Z,\hat{\sigma}_N)+P(Y|Z,\sigma^*)}{2P(Y|Z,\sigma^*)}\right]\Bigg) \\
&-\left(\frac{1}{N}\sum_{n=1}^{N}\log P(Y_n|Z_n,\sigma_N)-\frac{1}{N}\sum_{n=1}^{N}\log P(Y_n|Z_n,\sigma^*)\right)
\end{aligned}$$

showing that $h(\hat{\sigma}_N,\sigma^*)=o_P(1)$. This implies that

$$\begin{aligned}
\frac{1}{N}\sum_{n=1}^{N}\log P(Y_n|Z_n,\sigma_N) \leq& \frac{1}{N}\sum_{n=1}^{N}\log P(Y_n|Z_n,\hat{\sigma}_N) \\
&\qquad\qquad \leq \frac{1}{N}\sum_{n=1}^{N}\log P(Y_n|Z_n,\sigma^*)+o_P(1)
\end{aligned}$$

which by (10) implies that

$$\frac{1}{N}\sum_{n=1}^{N}\log P(Y_n|Z_n,\hat{\sigma}_N) \to E^*[\log P(Y|Z,\sigma^*)]$$

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

$$\begin{aligned} h^2(\sigma_{N_j},\sigma^*) &\leq \frac{1}{2}\int \sum_{y\in\{0,1\}^T} \left| \frac{P(y|z,\sigma_{N_j})+P(y|z,\sigma^*)}{2} - P(y|z,\sigma^*)\right| p(z)dz \\ &= \frac{1}{4}\int \sum_{y\in\{0,1\}^T} \left|P(y|z,\sigma_{N_j}) - P(y|z,\sigma^*)\right| p(z)dz \to 0 \end{aligned}$$

by the Lipschitz-continuity. Hence we get

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

This implies that

$$\frac{P(y|z,\tilde{\sigma})+P(y|z,\sigma^*)}{2} = P(y|z,\sigma^*)$$

which using the argument proving identifiability in section 3 shows that $\tilde{\sigma} = \sigma^*$. Hence $\hat{\sigma}_N$ is consistent in the norm on Σ. □

5 EXAMPLES

We illustrate the performance of the estimator first on simulated data and then we also apply the estimator to empirical data. Programming is carried out in Ox (see Doornik 2001) and the code is available on request. Evaluation of the likelihood requires integration. Analytical integration seems to be possible in principle. The actual expression for the outer integral is very complicated however and this integral in the likelihood is therefore approximated by simulation using pseudo-random Halton draws (Train 2003); the inner integral ($F(V_{nt}-\delta X_{nt}-u)$) is evaluated analytically.

5.1 Simulated data

We simulate observations from 1000 individuals making 8 choices each. This size of data is in line with our empirical dataset. We generate data for

Table 1: First and second moments of U

	Avg of estimated mean	Avg of estimated second moments
TRUE	-0.333	0.238
(2,1)	-0.330 (0.018)	0.196 (0.015)
(2,2)	-0.330 (0.016)	0.196 (0.014)
(3,2)	-0.331 (0.014)	0.195 (0.014)
(3,3)	-0.332 (0.014)	0.196 (0.013)

Table 2: Estimated ranges of U and ϵ

	Avg of min U	Avg of max U	Avg of scale of ϵ
TRUE	-1	1	1
(2,1)	-0.899 (0.083)	0.591 (0.158)	0.885 (0.079)
(2,2)	-0.879 (0.060)	0.597 (0.171)	0.930 (0.119)
(3,2)	-0.932 (0.101)	0.649 (0.197)	0.884 (0.108)
(3,3)	-0.951 (0.103)	0.621 (0.215)	0.873 (0.097)

U_n from an asymmetric beta(2,4) distribution, then we multiply by 2 and subtract 1 such the generated U_n have support on the interval $[-1; 1]$. We draw errors ϵ_{nt} from a symmetric beta(2,2) distribution and again multiply by 2 and subtract 1. Now $U_n + \epsilon_{nt}$ have support on $[-2; 2]$. We draw random bids from a uniform distribution on this interval. We have generated 100 such datasets for each of four estimated models. In order to obtain the likelihood we integrate analytically over the distribution of ϵ_{nt}, while the integration over the distribution of U_n is performed by simulation using 300 Halton draws.

We estimate four models, differing in the number of parameters used to represent the distributions of U_n (M_h) and ϵ_{nt} (M_f). We have estimated with $(M_h, M_f) = (2,1), (2,2), (3,2)$ and $(3,3)$ and we have constrained the ϵ to be symmetric. The average of the first and second moments of the estimated distribution of U are presented in the table 1. We note that the first moment is estimated rather precisely and the average estimate is very close to the true value, most so for the models with many parameters. The second moment is generally estimated lower than the true value, which has to do with the thin tails of the true distribution of U.

We also present the estimated ranges of U and ϵ. We note that the ranges are consistently too small compared to the true ranges, this again has to do with the thin tails of the true distributions and especially the thin right tail of the distribution of U.

Finally, figure 1 shows the pointwise average of the estimated cfd's of U

and ϵ for the case with (3,3) parameters along with pointwise 90 per cent confidence bands and the true distribution. It is evident that the estimator is able to track the true distribution quite closely.

5.2 Empirical application

The model has been applied to a real dataset originating from a survey designed to measure the WTP for travel time savings. Subjects were asked to choose between two alternatives described by travel time τ_i and cost χ_i, such that one alternative is cheap and slow and the other is expensive and fast. Each subject is faced with (up to) eight such choices. There is implicit in each choice a bid trade-off price of time, $-\Delta\chi/\Delta\tau$, ranging between 0.27 and 27 Euro per hour, up to more than three times the ex ante expectation of the mean WTP. The bids are transformed by $v = \log(-\Delta\chi/\Delta\tau)$ (see also Fosgerau 2007). We select a sample of 1204 subjects choosing between different trips by bus. No covariates are used in this application. The number of Halton draws for the numerical integration is set to 300.

The model has been estimated with $M_h = M_f = 3$. Beyond this point, improvements in the likelihood were very small. Confidence intervals have been generated by drawing 100 bootstrap samples with replacement from the original sample.

Figure 2 first shows the estimated cdf of U. It seems the distribution is estimated quite precisely. Figure 2 further indicates the support of V_{nt}. When considering the estimated cdf of ϵ, shown in figure 3, it appears that the support condition c) is not met, such that our proof of identification does not apply. A possible remedy is to introduce an index assumption to parametrise the location of U_n as indicated above. This is done with similar data in Fosgerau (2006) and Fosgerau, Hjort and Lyk-Jensen (2007), where the latter indicates that the support condition is then close to being met.

6 CONCLUDING REMARKS

We have shown that an unknown target distribution can be deconvoluted from an unknown error distribution in a model for binomial panel data assuming bounded support for both distributions. The estimator seems to perform well with both simulated and actual data.

So far, we have managed to establish consistency of the estimator. Further issues to consider are the rate of convergence of the estimator and the asymptotic distribution of particularly the Euclidian parameters. Another

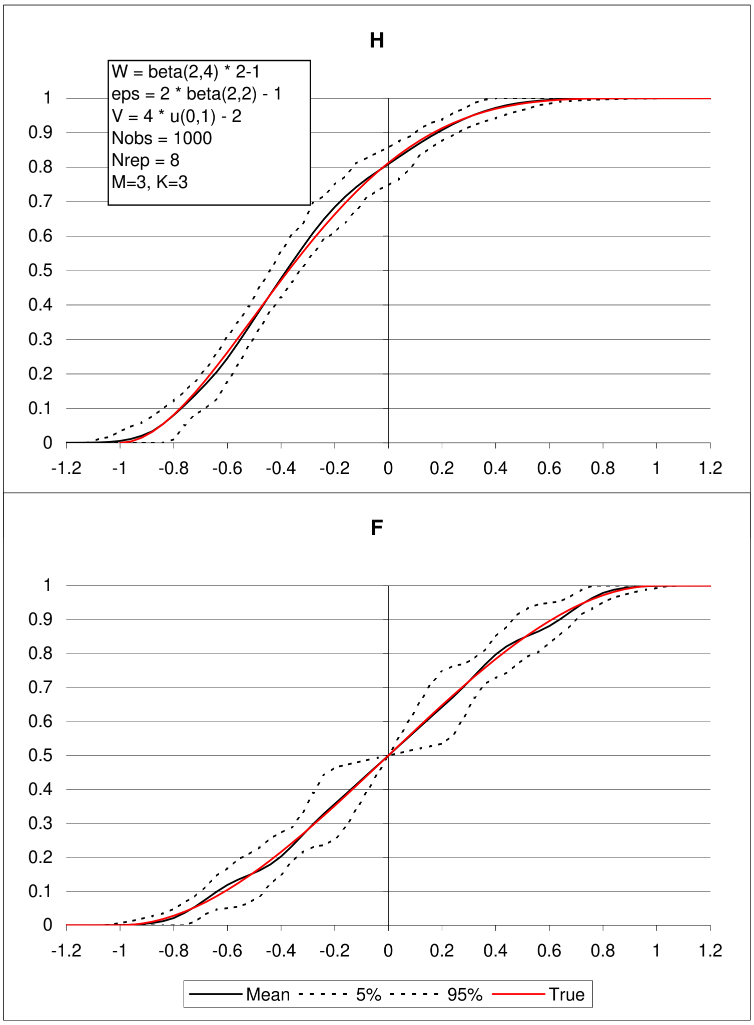

Figure 1: Estimation results, simulation study

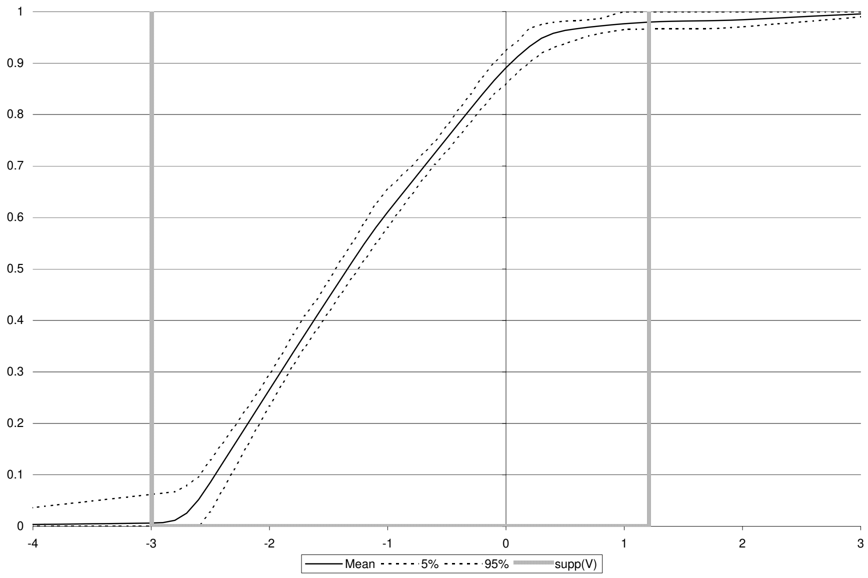

Figure 2: Estimation results, cdf. of U

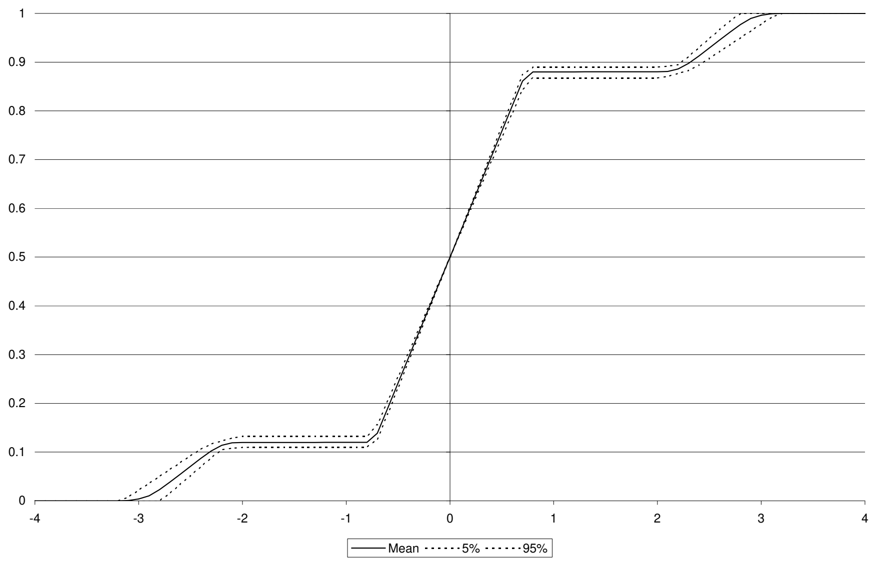

Figure 3: Estimation results, cdf. of ϵ

relevant issue is partial identification when the main support condition c) is not met, which we conjecture is common in applications.

A Appendix

A.1 Proof of Lemma 1 and some additional lemmas

Proof of Lemma 1 We start by noting that for any density $f \in \Phi$, the corresponding distribution function F is Lipschitz with parameter K'. Note that for distribution functions F and $\tilde{F}$ with densities f and $\tilde{f}$

$$\sup_z |F(z) - \tilde{F}(z)| \leq \sup_z \int_{-\infty}^{z} |f(x) - \tilde{f}(x)|dx \leq \|f - \tilde{f}\|_1$$

Consequently, for $f, \tilde{f} \in \Phi$ with corresponding distribution functions F and $\tilde{F}$ we have

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

Thus

$$\begin{aligned} & |F(\theta Z_{nt} - \zeta U_n) - \tilde{F}(\tilde{\theta} Z_{nt} - \tilde{\zeta} U_n)| \\ \leq & K' \Big(|(\theta - \tilde{\theta}) \cdot Z_{nt}| + |\zeta - \tilde{\zeta}| \cdot U_n \Big) + \|f - \tilde{f}\|_1 \end{aligned}$$

Now putting

$$a_{nt} = (2Y_{nt} - 1)F(\theta Z_{nt} - \zeta U_n) + (1 - Y_{nt})$$

and

$$\tilde{a}_{nt} = (2Y_{nt} - 1)\tilde{F}(\tilde{\theta} Z_{nt} - \tilde{\zeta} U_n) + (1 - Y_{nt})$$

and using

$$\prod_{t=1}^{T} a_{nt} - \prod_{t=1}^{T} \tilde{a}_{nt} = \sum_{t=1}^{T} \prod_{s<t} a_{ns} \cdot \Big(a_{nt} - \tilde{a}_{nt} \Big) \cdot \prod_{s>t} \tilde{a}_{ns} \tag{11}$$

we see that since $|a_{nt}| \leq 1$ and $|\tilde{a}_{nt}| \leq 1$

$$\begin{aligned} & \Big| P(Y_n|Z_n, \sigma) - P(Y_n|Z_n, \tilde{\sigma}) \Big| \\ \leq & \int |h(u) - \tilde{h}(u)| \prod_{t=1}^{T} \tilde{a}_{nt} du + \int \left| \prod_{t=1}^{T} a_{nt} - \prod_{t=1}^{T} \tilde{a}_{nt} \right| h(u) du \\ \leq & \|h - \tilde{h}\|_1 + \sum_{t=1}^{T} \int |a_{nt} - \tilde{a}_{nt}| h(u) du \\ \leq & \|h - \tilde{h}\|_1 + T\|f - \tilde{f}\|_1 + TK'|\zeta - \tilde{\zeta}| + K' \sum_{t=1}^{T} |Z_{nt} \cdot (\theta - \tilde{\theta})| \end{aligned}$$

as $E[U_n] \leq 1$. $\square$

Lemma 2. *Consider the set of densities on the unit interval bounded by K:*

$$\mathcal{D} = \{f \in L_1[0;1] : \|f\|_1 = 1, 0 \leq f \leq K\}$$

Let Γ_M be as defined in section 4. Then for any $f \in \mathcal{D}$ there exists a sequence $(h_n)_{n\in\mathbb{N}}$ with $h_n \in \cup_M \Gamma_M$ such that $h_n \to f$ in L_1. In particular, $\mathcal{D}$ may be identified with the L_1-closure of $\cup_M \Gamma_M$.

Proof. $\Gamma \subseteq \mathcal{D}$ follows by definition. To prove the converse, let $f \in \mathcal{D}$ be an arbitrary point. Now the set of continuous functions on the unit interval is dense in L_1 (Rudin 1986, Theorem 3.14), so we can select a sequence of continuous functions $\{f_n\}$ such that $\|f - f_n\|_1 < 2^{-n}$. Since $0 \leq f \leq K$ we have $\|f - (f_n \vee 0) \wedge K\|_1 \leq \|f - f_n\|_1$, so we can ensure that $0 \leq f_n \leq K$ by truncating.

As a continuous function on a compact set, each f_n is also uniformly continuous. Hence we can find M_n such that $|x_1 - x_2| < 1/M_n \Rightarrow |f_n(x_1) - f_n(x_2)| < 2^{-n}$.

We shall then approximate f_n by defining

$$h_n(x) = (M_n + 1) \sum_{m=0}^{M_n} \int_{I_{m,M_n}} f_n(z)dz \cdot 1\{u \in I_{m,M_n}\}$$

We have $\|f_n - h_n\|_1 \leq 2^{-n}$ and $\|f_n\|_1 = \|h_n\|_1$. Furthermore, $0 \leq h_n \leq K$ but $\|h_n\|_1$ may be different from 1. We will define a function g_n to make up the difference, i.e. $\|g_n\|_1 = 1 - \|h_n\|_1$ and $h_n + g_n \in \Gamma_{M_n}$. If, say, $\|h_n\|_1 < 1$, then g_n must add some positive mass. This can just be added to intervals $I_{m,M}$ such that we maintain that $0 \leq h_n + g_n \leq K$. Then

$$\begin{aligned}\|f - (h_n + g_n)\|_1 &\leq \|f - f_n\|_1 + \|f_n - h_n\|_1 + \|g_n\|_1 \\ &\leq 2^{-n} + 2^{-n} + 1 - \|h_n\|_1 \to 0,\end{aligned}$$

since $\|f_n\|_1 = \|h_n\|_1 \to 1$. $\square$

Lemma 3. Γ *is contained in the pointwise closure of* $\cup_M \Gamma_M$.

Proof For any $h \in \Gamma$ there exists a sequence of functions $h_n \in \cup_M \Gamma_M$ converging to h in L_1. By Theorem 3.12 in Rudin (1986) we may select a subsequence such that h_m converges pointwise to h a.e. Thus Γ is contained in the pointwise closure of $\cup_M \Gamma_M$. $\square$

A.2 Uniform convergence

We wish to show that

$$\lim_{N\to\infty} \sup_{\sigma\in\Sigma} |s_N(\sigma) - Es_N(\sigma)| = 0 \text{ almost surely} \tag{12}$$

where

$$s_N(\sigma) = \frac{1}{N}\sum_{n=1}^{N} \log \frac{P(Y_n|Z_n,\sigma) + P(Y_n|Z_n,\sigma^*)}{2P(Y_n|Z_n,\sigma^*)}$$

To do this we first review the concepts of the covering number of a class of functions, then calculate the covering number for a specific class of functions, before finally proving the desired uniform result.

In this appendix we will identify Φ with the set of distribution functions, F, corresponding to the densities, f,we are estimating.

A.2.1 Covering numbers

The covering number, $N(\varepsilon, \mathcal{G}, \|\cdot\|)$, of a set of functions, $\mathcal{G}$, is the smallest number, k, of functions $g_1, \ldots, g_k$ such that for any $g \in \mathcal{G}$ there is a $j \in \{1, \ldots, k\}$ such that $\|g - g_j\| < \varepsilon$. The "centres" $g_1, \ldots, g_k$ may depend on the norm and need not be functions in $\mathcal{G}$.

Theorem 2.4.3 of van der Vaart and Wellner (1996) states that if for any $\varepsilon > 0$ and any $M > 0$

$$\log N(\varepsilon, \mathcal{G}_M, \|\cdot\|_N) = o_P(N) \tag{13}$$

where

$$\mathcal{G}_M = \{g \cdot 1\{G \leq M\} : g \in \mathcal{G}\} \tag{14}$$

where G is an integrable envelope for the measurable class $\mathcal{G}$ of functions, i.e. a function such that $|g(x)| \leq G(x)$ for all x and $\int GdP < \infty$, and the norm $\|\cdot\|_N$ is the L_1-norm corresponding to the empirical distribution then

$$\lim_{N\to\infty} \sup_{g\in\mathcal{G}} \left| \frac{1}{N}\sum_{i=1}^{N} g(X_i) - E[g(X_1)] \right| = 0 \tag{15}$$

for iid random variables $X_1, \ldots X_n$.

Thus to obtain (12), we need to bound a suitable covering number. To do this we need to introduce the concept of VC (or Vapnik-Červonenskis) classes of sets, VC hull-classes of functions and VC-major classes of functions. A collection of subsets $\mathcal{C}$ of a set $\mathcal{X}$ shatters a set $A = \{x_1, \ldots, x_n\} \subset \mathcal{X}$ of size n if

$$\{A \cap C : C \in \mathcal{C}\}$$

is the set of all subsets of A. If for some n $\mathcal{C}$ does not shatter any set of size n then $\mathcal{C}$ is a VC class of sets, and the smallest such n is called the VC index of the class. The covering number of the indicator functions for a VC class of sets is bounded by a constant (depending on the VC index but not the norm) times $1/\varepsilon$ raised to the power of VC index minus 1 (van der Vaart and Wellner 1996, Theorem 2.6.4). In particular, it is small compared to the requirements of Theorem 2.4.3 of van der Vaart and Wellner (1996). Examples of VC classes of sets are the set of intervals on the real line (VC index 3) and the set of half-spaces in $\mathbb{R}^d$ (VC index $d+2$); see Example 2.6.1 and Problem 2.6.14 in van der Vaart and Wellner (1996).

A class of functions $\mathcal{G}$ is a VC-hull class of functions if any $f \in \mathcal{G}$ is the pointwise limit of functions of the form

$$\sum_{i=1}^{m} \alpha_i 1_{C_i}(x)$$

with α_i arbitrary subject to the constraint $\sum_1^m |\alpha_i| \leq c$ for some fixed c and the sets C_i chosen from a VC class of sets. Such a class of functions also have covering numbers that are polynomial in $1/\varepsilon$ (van der Vaart and Wellner 1996, Corollary 2.6.12). We note that the set Γ as defined in section 4 is VC-hull, since any function in Γ may be thought of as the pointwise limit of "histogram"-approximations; see Lemma 3. As noted in section 4 this implies that Γ is totally bounded.

Finally, a class of functions $\mathcal{G}$ is a VC-major class if the sets

$$\big\{\{x \in \mathcal{X} : f(x) > t\} : t \in \mathbb{R}, f \in \mathcal{G}\big\}$$

is a VC class of sets. It is easily shown (van der Vaart and Wellner 1996, Lemma 2.6.13) that if $\mathcal{G}$ is a VC-major class with functions that are bounded by some M, then $\mathcal{G}$ is also VC-hull. By van der Vaart and Wellner (1996, Lemma 2.6.19) sets of monotone functions are VC-hull.

A.2.2 Calculation of covering numbers

Using the facts of the previous subsection, it now follows that the set of functions

$$\big\{(w,u) \to \theta w - \zeta u : (\theta, \zeta) \in \Theta\big\}$$

is a VC-major class as the corresponding set of sets are half-spaces and thus VC. Moreover, as Φ is a class of bounded monotone functions

$$\big\{(w,u) \to F(\theta w - \zeta u) : (\theta, \zeta) \in \Theta, F \in \Phi\big\}$$

is VC-major (van der Vaart and Wellner 1996, Lemma 2.6.19) and bounded. It follows from Lemma 2.6.20 in van der Vaart and Wellner (1996), that the class

$$\left\{(y,w,u) \to [(2y-1)F(\theta w - \zeta u) + (1-y)] : (\theta,\zeta) \in \Theta, F \in \Phi\right\} \tag{16}$$

is also VC-hull. 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.

Γ is by construction a subset of a VC-hull class and it follows by Lemma 2.6.20 in van der Vaart and Wellner (1996) that the class

$$\begin{aligned}\big\{(y,w,u) \to h(u)\,[(2y-1)F(\theta w - \zeta u) + (1-y)] :& \\ (\theta,\zeta) \in \Theta, F \in \Phi, h \in \Gamma&\big\}\end{aligned} \tag{17}$$

is a subset of a VC-hull class. In particular, its covering number is bounded by a constant times a power of $1/\varepsilon$ with neither the constant nor the power depending on the norm.

Now consider a L_1-norm corresponding to the product of a probability measure μ and the Lebesgue measure on $[0;1]$ and let $f_1, \dots, f_k$ be the corresponding centres. Now consider the function class

$$\mathcal{G} = \left\{(y,w) \to \int h(u)F(\theta w - \zeta u)du : (\theta,\zeta) \in \Theta, F \in \Phi, h \in \Gamma\right\} \tag{18}$$

Then for any choice of $h \in \Gamma$, $F \in \Phi$ and $(\theta,\zeta) \in \Theta$ with corresponding centre g_j we have

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

Hence the covering number of the class $\mathcal{G}$ (18) is at most as large as the covering number of the class (17).

A.2.3 Uniform convergence

To summarise, we have now shown that the class

$$\{(y,z) \to P(y|z,\sigma) : \sigma \in \Sigma\}$$

has a covering number which is polynomial in ε. Let $P_j(y|z)$ denote the centres corresponding to the covering of size ε for the norm

$$\|g\| = \frac{\frac{1}{N}\sum_{n=1}^{N} \dfrac{|g(Y_n, Z_n)|}{P(Y_n|Z_n, \sigma^*)}}{\frac{1}{N}\sum_{n=1}^{N} 1/P(Y_n|Z_n, \sigma^*)}.$$

Consider now the class

$$\mathcal{G} = \left\{ (y, z) \to \log\left(\frac{P(y|z, \sigma) + P(y|z, \sigma^*)}{2P(y|z, \sigma^*)}\right) : \sigma \in \Sigma \right\} \tag{19}$$

Now

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

Recalling that what we need is the covering number for (19) under the random norm $\|\cdot\|_n$ (see (13)), we bound the relevant distance as follows:

$$\begin{aligned} &\frac{1}{N}\sum_{n=1}^{N} \frac{|P(Y_n|Z_n, \sigma) - P_j(Y_n|Z_n, \sigma^*)|}{P(Y_n|Z_n, \sigma^*)} \\ &\quad \leq \frac{\frac{1}{N}\sum_{n=1}^{N} \dfrac{|P(Y_n|Z_n, \sigma) - P_j(Y_n|Z_n, \sigma^*)|}{P(Y_n|Z_n, \sigma^*)}}{\frac{1}{N}\sum_{n=1}^{N} 1/P(Y_n|Z_n, \sigma^*)} \cdot \frac{1}{N}\sum_{n=1}^{N} \frac{1}{P(Y_n|Z_n, \sigma^*)} \\ &\qquad\qquad\qquad\qquad\qquad\qquad \leq \varepsilon \cdot \frac{1}{N}\sum_{n=1}^{N} \frac{1}{P(Y_n|Z_n, \sigma^*)} \end{aligned}$$

Noting that

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

it follows that (13) is satisfied.

To verify the envelope condition (14), 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 (19).

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) is in a countable dense subset of Θ, h and h are given by rational α_is and intervals with rational endpoints.

Remark: In the proof of the uniform convergence, we do not need restrictions on Φ nor on the set Θ which may be the entire $\mathbb{R}^d$. Also Γ may be all of $\mathcal{H}$ (or even more as long as Γ is a VC-hull class). However, some restrictions are necessary in order to ensure the compactness and continuity required in proof of consistency.

References

Belluzzo, W. (2004) Semiparametric Approaches to Welfare Evaluations in Binary Response Models *Journal of Business and Economic Statistics* **22**(3), 322–330.

Chen, H. Z. and Randall, A. (1997) Semi-nonparametric estimation of binary response models with an application to natural resource valuation *Journal of Econometrics* **76**(1-2), 323–340.

Doornik, J. A. (2001) *Ox: An Object-Oriented Matrix Language* Timberlake Consultants Press London.

Fosgerau, M. (2006) Investigating the distribution of the value of travel time savings *Transportation Research Part B: Methodological* **40**(8), 688–707.

Fosgerau, M. (2007) Using nonparametrics to specify a model to measure the value of travel time *Transportation Research Part A: Policy and Practice* **Forthcoming**.

Fosgerau, M., Hjort, K. and Lyk-Jensen, S. V. (2007) An approach to the estimation of the distribution of marginal valuations from discrete choice data *Working Paper* .

Gabler, S., Laisney, F. and Lechner, M. (1993) Seminonparametric Estimation of Binary-Choice Models With an Application to Labor-Force Participation *Journal of Business and Economic Statistics* **11**(1), 61–80.

Gallant, A. R. and Nychka, D. W. (1987) Semi-nonparametric maximum likelihood estimation *Econometrica* **55**(2), 363–390.

Hanemann, W. M. and Kanninen, B. (1998) The Statistical Analysis of Discrete-Response CV Data *in* I. J. Bateman and K. G. Willis (eds), *Valuing Environmental Preferences. Theory and Practice of the Contingent Valuation Method in the US, EC and Developing Countries* Oxford University Press (forthcoming).

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., Linton, O. and McFadden, D. (2002) Estimating features of a distribution from binomial data. unpublished work.

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

Rouwendal, J. and de Blaeijj, A. T. (2004) Inconsistent and Lexicographic Choices in Stated Preference Analysis *Tinbergen Institute Discussion Papers* **04-038/3**.

Rudin, W. (1986) *Real and Complex Analysis* McGraw-Hill Singapore.

Sælensminde, K. (2001) Inconsistent choices in stated choice data. Use of the logit scaling approach to handle resulting variance increases *Transportation* **28**(3), 269–296.

Sælensminde, K. (2002) The Impact of Choice Inconsistencies in Stated Choice Studies *Environmental and Resource Economics* **23**(4), 403–420.

Train, K. (2003) *Discrete Choice Methods with Simulation* Cambridge University Press Cambridge, MA.

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