Instrumental Variable Estimation with Discrete Endogenous Regressors*

Tiemen Woutersen†

October 25, 2007

Abstract

This paper develops a new objective function that can be used to estimate models with discrete endogenous regressors. The basis for the objective function is the objective function that one would have used if all covariates were assigned randomly. The paper than proposes a weighting function that uses instruments and 'removes' the endogeneity. Methods that assume that the endogenous regressor or the instrument are continuously distributed already exist and this paper deals with endogenous variables that are discrete.

Very preliminary.

1 Introduction

Recently the scope for using instrumental variables or exclusion restrictions to reduce the bias in the estimated effects of endogenous covariates on an outcome has been greatly expanded. Older methods like Two-Stage Least Squares (2SLS) required that the endogenous regressor entered linearly and that the relation had an additive error. Later methods based on the Generalized Method of Moments (GMM) (Hansen, 1982) allowed that the endogenous regressors enter nonlinearly, but retained the assumption of an additive random error. Recent contributions (Blundell and Powell, 2003a, Imbens and Newey, 2003; see also Blundell and Powell, 2003b, for a survey of the various approaches), consider the estimation of models in which the model is nonlinear in the endogenous covariates and has a random error that is correlated with the endogenous covariates that is nonseparable.

*I have received helpful comments from Geert Ridder, Robert Moffitt, Daniel Scharfstein and Constantine Frangakis.

†Comments are welcome, woutersen@jhu.edu.

An important example of such a model is the binary response model with endogenous covariates. It should be noted that most popular models with limited dependent variables have a nonseparable error.

The estimators proposed by Blundell and Powell (2003a) and Imbens and Newey (2003) make assumptions that ensure that given the (possibly nonseparable) random error in the first-stage relation between the endogenous covariate and the exogenous covariates (including some that do not enter in the structural relation of interest), the (nonseparable) random error in the structural equation is independent of the endogenous covariate. Solving for the error of the first-stage relation one obtains a control function that upon inclusion in the structural relation transforms the model in one with exogenous covariates. This approach only works if the endogenous variables are continuous. Other recent papers include Chesher (2005), Chesher (2007), Vytlacil and Yildiz (2007) and Horowitz and Lee (2007).

In this paper we consider the case that the endogenous variable is discrete and the random error is not separable. Instead of a control function we consider the use of weights that re-weight a sample statistic that depends on the endogenous (discrete) covariate to a (not observed) population of compliers in which that covariate is randomly assigned (given the other exogenous covariates). Abadie (2000) and Abadie, Angrist, and Imbens (2002) also use weights and their weights are a special case of ours. Our contribution is that we explore the application of weighting estimators in models that currently are outside the scope of IV estimation. In particular, we consider estimation of the effect of a 0-1 intervention, i.e. a dummy endogenous covariate, on a limited-dependent or transformed outcome or on a (censored) duration. These outcomes may be observed once, or we may have a sequence of outcomes, i.e. the estimation is on a panel or, more generally, longitudinal data set. We are particularly interested in longitudinal outcomes, i.e. a time-series of dummy dependent variables or a possibly censored duration. As we shall see, an important advantage of the Weighted Objective Function Instrumental Variable (WOF-IV) estimator is that we need to assume very little on the selection equation, or in the context of treatment effects, the treatment assignment mechanism.

Examples of estimators that we consider are the maximum score estimator, maximum rank correlation estimator, maximum likelihood estimator, some random effects estimators and several fixed effect estimators. We distinguish between estimators that estimate the heterogeneity distribution, such as random effects estimators, and those that do not, such as the fixed effects estimators we consider and the estimator of the transformation model. Examples that do not estimate the heterogeneity distribution are the estimator of Horowitz (1996) of the transformation model and Han's (1987) Maximum Rank Correlation estimator for transformation models.

There is an important difference between these two cases. Insofar as the endogeneity is due to dependence of the dichotomous variable on the individual effect, the estimators that do not depend on the individual effect are consistent. Hence the WOF-IV estimator in this case allows for dependence of the dichotomous variable on the idiosyncratic error. For the case that the estimator depends on the distribution of the individual effect, we show that the distribution of that effect in the population defined by the weights, depends on the selection model, be it is a specific way. We also show that given the estimates of the parameters of the model, we can recover the population distribution of the individual effect, so that we can estimate the full counterfactual outcome distributions. Note that for instance in a duration model with unobserved heterogeneity, the effect of the treatment on the average duration or the hazard is not the same for all individuals[1]. Hence, the usual problem that estimates of the counterfactual outcome distributions if treatments have an effect that differs between members of the population does not apply to models that have non-separable individual effects. As far as we know the WOF-IV is the only estimator for mixture models with a discrete endogenous variable and a non-separable individual effect.

The plan of the paper is as follows. In section 2 we explain the role of the weights and we introduce the WOF-IV estimator. Section 3 discusses the case that the estimation criterion and hence the estimator, does not depend on the individual effect. In section 4 we consider the case that the estimator does depend on the individual effect. Section 5 gives an application for the second case and section 6 concludes.

2 Endogenous binary variable

Consider the following model

$$Y_i = G(X_i, v_i, W_i, \varepsilon_i) \tag{1}$$

where Y_i is a vector of outcomes, X_i is an observed binary variable, v_i is a scalar error term, W_i is a vector of observed regressors, and ε_i is a vector

[1] In a (Mixed) Proportional Hazard model the relative effect is the same for all individuals, but there is no reason to restrict attention to such models, for instance by having a different individual effect for treatments and controls that are independent of other exogenous covariates.

of unobserved error terms. For example, Y_i could denote a duration and the censoring indicator or a number of realizations in a panel data model. Let the regressor X_i be endogenous in the sense that it depends on v_i. In particular, the dichotomous variable X is a function of R, v, W, η, and we express this by writing $X(R, v, W, \eta)$. In the case that X is an indicator of an intervention this specifies the treatment assignment mechanism. We make the following assumption on this function.

Assumption 1 *$X(R, v, W, \eta)$ is non-decreasing and not constant in v on the support $\mathcal{V}$ where η is a random variable. Moreover,*

$$X(1, v, W, \eta) \geq X(0, v, W, \eta)$$

for all $v \in \mathcal{V}$ with a strict inequality on a interval in $\mathcal{V}$.

This assumption implies that there are $\underline{v}(W, \eta)$ and $\overline{v}(W, \eta)$ as in figure 1. Note that this implies that

$$\begin{aligned} X(R, v, W, \eta) &\equiv 0 \qquad \text{if } v < \underline{v}(W, \eta) & (2) \\ &\equiv R \qquad \text{if } \underline{v}(W, \eta) \leq v \leq \overline{v}(W, \eta) & (3) \\ &\equiv 1 \qquad \text{if } v > \overline{v}(W, \eta) & (4) \end{aligned}$$

i.e. we can identify 'never takers'[2], 'compliers' and 'always takers' by intervals of values of the individual effect. Note that this assumption holds if $X(R, v, W, \eta) = I(\delta_0 + \delta_1 R + v - \eta \geq 0$ with $\delta_1 > 0$. It also holds if δ_1 is a random variable provide that it is a positive random variable[3]. We do not need to specify a model for X beyond the qualitative assumptions made here. This is a major advantage of our procedure.

Consider the hypothetical case in which $\underline{v}(W_i, \eta_i) \leq v_i \leq \overline{v}(W_i, \eta_i)$ for all individuals. That is, all individuals are 'compliers'. In this case, we do not have an endogeneity problem. Suppose that we have an objective function for this case. Define $S_i = \{Y_i, X_i, W_i\}$ and

$$Q_N^*(\theta) = f\{\frac{1}{N} \sum_{i: v_i \in [\underline{v}(W_i, \eta_i), \overline{v}(W_i, \eta_i)]} q(S_i; \theta)\}.$$

That is, the objective function $Q_N^*(\theta)$ is a function of an average of the data and let $f(.)$ be a known continuous function. Maximum likelihood, general

[2] See Angrist, Imbens and Rubin (1996) for this terminology.
[3] We can also have that $\delta_1 < 0$ or that it is a negative random variable

method of moments, maximum score and other objective function satisfy this. We assume that maximizing $Q^*_N(\theta)$ yields a consistent estimate of θ. In particular,

Assumption 2 *Suppose that we have a random sample of size N. Let $\theta \in \Theta$, which is compact. Let $Q^*_N(\theta) = \frac{1}{N}\sum_{i:v_i\in[\underline{v}(W_i,\eta_i),\overline{v}(W_i,\eta_i)]} q(S_i;\theta)$ or $Q^*_N(\theta) = \{\frac{1}{N}\sum_{i:v_i\in[\underline{v}(W_i,\eta_i),\overline{v}(W_i,\eta_i)]} q(S_i;\theta)\}'M\cdot\frac{1}{N}\sum_{i:v_i\in[\underline{v}(W_i,\eta_i),\overline{v}(W_i,\eta_i)]} q(S_i;\theta)$ where the matrix M can be consistently estimated. Let $Q^*_N(\theta)$ converge uniformly in probability to $Q_0(\theta)$ and let $Q_0(\theta)$ be continuous and uniquely maximized at $\theta = \theta_0$.*

Assumption 2 is quite general. However, we usually cannot observe whether $i : v_i \in [\underline{v}(W_i,\eta_i),\overline{v}(W_i,\eta_i)]$. Suppose, however, that we have an instrument R that is correlated with X and independent of v_i and ε_i. For now, we assume that R is binary. We base our inference on the following objective function

$$Q_N(\theta) = \frac{1}{N}\sum_i [\{1(R_i = X_i) - 1(R_i \neq X_i)(\frac{\hat{p}(w)}{1-\hat{p}(w)})^{1-2R_i}\}q(S_i;\theta)] \quad (5)$$

or, in case $Q^*_N(\theta)$ is a method of moment estimator,

$$Q_N(\theta) = \{\frac{1}{N}\sum_i [\{1(R_i = X_i) - 1(R_i \neq X_i)(\frac{\hat{p}(w)}{1-\hat{p}(w)})^{1-2R_i}\}q(S_i;\theta)]\}'\hat{M}\cdot$$

$$\cdot\{\frac{1}{N}\sum_i [\{1(R_i = X_i) - 1(R_i \neq X_i)(\frac{\hat{p}(w)}{1-\hat{p}(w)})^{1-2R_i}\}$$

where $\hat{p}(w)$ is a consistent estimator for the probability $P(R = 1|w)$ and $R_i \in \{0,1\}$. If W is a discrete random variable, then $\hat{p}(w) = \sum_i\{R_i \cdot 1(W_i = w)\}/\sum_i\{1(W_i = w)\}$. If R is assumed to be independent of W then $\hat{p}(w) = \hat{p} = \sum_i R_i/N$ is used. The matrix $\hat{M}$ is a consistent estimator of M.

Note that we use the data on all individuals in equation (5), therefore, we also need assumptions on the individuals for which $v_i \notin [\underline{v}(W_i,\eta_i),\overline{v}(W_i,\eta_i)]$.

Assumption 3 *Let (i) $q(S_i,\theta)$ be continuous at each $\theta \in \Theta$ with probability one, and there is $d(S_i)$ with $|q(S_i,\theta)| \leq d(S)$ for all $\theta \in \Theta$ and $E\{d(S)\} < \infty$ or (ii) $q(S_i,\theta)$ be stochastically equicontinuous at each $\theta \in \Theta$ and $E|q(S_i,\theta)| < \infty$ at each $\theta \in \Theta$.*

Assumption 4 *Let one of the following three conditions hold, (i)* $R \perp v, \varepsilon, W$ *and* $X \perp W, \varepsilon | v$ *or (ii)* $R \perp v, \varepsilon | W$ *and* W *is discrete or (iii) [*$R \perp v, \varepsilon | W$, $\widehat{p(w)} = p(w) + o_p(1)$, *and* $p(w) > \delta$ *for all* w *for some* $\delta > 0$. *to be completed if we want it]*

Assumption 4 is satisfied if either (i) R is randomly assigned and X is a function of Z and v or (ii) R is randomly assigned conditional on W and X is a function of R, v, η, and W.

Assumption 5 *Let* $Cov(R, X) > 0$.

Theorem 1
Let assumption 1-5 hold. Then $\hat{\theta} = \arg\max Q_N(\theta)$ converges in probability to θ_0.

Theorem 1 covers many objective functions but not those with double summations, such as the maximum rank correlation estimator or Chen's (2002) estimator of the transformation model. Let $Q_{2,N}$ denote such an objective function with a double summation. Consider the following objective function that is a function of the observed data,

$$Q_{2,N}(\theta) = \frac{1}{N^2} \sum_i [\{1(R_i = X_i) - 1(R_i \neq X_i)(\frac{\hat{p}(w)}{1 - \hat{p}(w)})^{1-2R_i}\} *$$
$$* \sum_j \{1(R_j = X_j) - 1(R_j \neq X_j)(\frac{\hat{p}(w)}{1 - \hat{p}(w)})^{1-2R_i}\} q(S_i, S_j; \theta)].$$

where $\hat{p}(w)$ is a consistent estimator for the probability $P(R = 1|w)$ and $R_i \in \{0, 1\}$. The maximum rank correlation estimator by Han (1987), the estimator of the transformation model by Chen (2002) and estimator of the duration model by Hausman and Woutersen (2005) have this form.

Assumption 6 *Suppose that we have a random sample of size* N. *Let* $\theta \in \Theta$, *which is compact. Let*

$$Q_N^*(\theta) = \frac{1}{N^2} \sum_{i: v_i \in [\underline{v}(W_i, \eta_i), \overline{v}(W_i, \eta_i)]} \sum_{j: v_j \in [\underline{v}(W_j, \eta_j), \overline{v}(W_j, \eta_j)]} q(S_i, S_j; \theta)$$

converge uniformly in probability to $Q_{2,0}(\theta)$ *and let* $Q_{2,0}(\theta)$ *be continuous and uniquely maximized at* $\theta = \theta_0$.

Assumption 7 *Let $q(S_i, S_j, \theta)$ be stochastically equicontinuous at each $\theta \in \Theta$ and $E|q(S_i, S_j, \theta)| < \infty$ at each $\theta \in \Theta$..*

Theorem 2
Let assumption 1, 4–7 hold. Then $\hat{\theta} = \arg\max Q_{2,N}(\theta)$ converges in probability to θ_0.

Derivative depend on of Q_N or an approximate derivative. influence function.

Proof of theorem 1
Define

$$r_N(\theta) = \frac{1}{N}\sum_i [\{1(R_i = X_i) - 1(R_i \neq X_i)(\frac{\hat{p}(w)}{1-\hat{p}(w)})^{1-2R_i}\}q(S_i;\theta)]$$

and

$$r_{infeasible,N}(\theta) = \frac{1}{N}\sum_{i:v_i \in [\underline{v}(W_i,\eta_i),\overline{v}(W_i,\eta_i)]} q(S_i;\theta)$$

We will show that $r_N(\theta)$ converges uniformly in probability to $r_{infeasible,N}(\theta)$. Let $\underline{v}$ and $\overline{v}$ denote $\underline{v}(W_i,\eta_i)$ and $\overline{v}(W_i,\eta_i)$.Define

$$\begin{aligned}
\Delta r_N(\theta) &= r_N(\theta) - r_{infeasible,N}(\theta)\\
&= \frac{1}{N}\sum_i [\{1(R_i = X_i) - 1(R_i \neq X_i)(\frac{\hat{p}(w)}{1-\hat{p}(w)})^{1-2R_i}\}q(S_i;\theta)]\\
&- \frac{1}{N}\sum_{i:v_i \in [\underline{v},\overline{v}]} q(S_i;\theta)\\
&= \frac{1}{N}\sum_{i:v_i \notin [\underline{v},\overline{v}]} 1(R_i = X_i)\cdot q(S_i;\theta) - \frac{1}{N}\sum_i 1(R_i \neq X_i)(\frac{\hat{p}(w)}{1-\hat{p}(w)})^{1-2R_i} q(S_i;\theta).
\end{aligned}$$

$$\begin{aligned}
1(R_i = X_i) &= R_i \cdot 1(v_i \geq \underline{v}) + (1-R_i)\cdot 1(v_i \leq \overline{v})\\
&= R_i \cdot 1(v_i \in [\underline{v},\overline{v}]) + R_i \cdot 1(v_i > \overline{v}) + (1-R_i)\cdot 1(v_i < \underline{v})\\
1(R_i \neq X_i) &= (1-R_i)\cdot 1(v_i > \overline{v}) + R_i \cdot 1(v_i < \underline{v})
\end{aligned}$$

This yields

$$\Delta r_N(\theta) = \{\frac{1}{N}\sum_i \{R_i \cdot 1(v_i > \overline{v}) + (1 - R_i) \cdot 1(v_i < \underline{v})\} \cdot q(S_i;\theta)$$

$$-\frac{1}{N}\sum_i \{(1 - R_i) \cdot 1(v_i > \overline{v}) + R_i \cdot 1(v_i < \underline{v})\}(\frac{\hat{p}(w)}{1 - \hat{p}(w)})^{1-2R_i}\} \cdot q(S_i;\theta).$$

Note that $\{R_i \cdot 1(v_i > \overline{v}) \cdot q(S_i;\theta)\}$ is either (i) continuous in θ for all $\theta \in \Theta$ with probability one and bounded by the function $d(S)$ of assumption 3 or (ii) stochastically equicontinuous in θ for all $\theta \in \Theta$ with finite expectation. Note that R_i and $1(v_i > \overline{v})$ are bounded and do not depend on θ. Therefore, $\frac{1}{N}\sum_i \{R_i \cdot 1(v_i > \overline{v}) \cdot q(S_i;\theta)\}$ converges uniformly to $E_W[E\{q(S_i;\theta)|v_i > \overline{v}, W_i\}p(v_i > \overline{v}|W_i)]$ by Newey McFadden (1994) lemma 2.4 and lemma 2.8 respectively. Similar reasoning[4] yields that $\frac{1}{N}\sum_i \{(1 - R_i) \cdot 1(v_i > \overline{v})(\frac{\hat{p}(w)}{1-\hat{p}(w)})^{1-2R_i}\} \cdot q(S_i;\theta)\}$ converges uniformly to the same function of θ so that the difference between these two terms converges uniformly to zero. The same reasoning holds for the other two terms so that $r_N(\theta)$ converges uniformly to $r_{infeasible,N}(\theta)$ and $Q_N(\theta)$ to $Q_N^*(\theta)$. Theorem 2.1 of Newey and McFadden (1994) applies and the result follows.
Proof of theorem 2: The proof is very similar to the proof of theorem 1 but now

$$\frac{1}{N}\sum_j \{1(R_j = X_j) - 1(R_j \neq X_j)(\frac{\hat{p}(w)}{1 - \hat{p}(w)})^{1-2R_i}\}q(S_i, S_j;\theta)]$$

converges uniformly to $E[E\{q(S_i, S_j)|S_i, v_j \in [\underline{v}, \overline{v}]\}, W_j)p(v_j \in [\underline{v}, \overline{v}]|W_j)]$ so that $Q_{2,N}(\theta)$ converges uniformly to $Q_{2,0}(\theta)$.

Proof of theorem 1: The conditions ensure that $\hat{\theta}$ is a consistent estimator of θ.

3 Compliers

In this section, we use a different notation. In particular, we use the notation of compliers, always takers, and never takers that was introduced by Angrist, Imbens and Rubin (1996). In particular, Angrist, Imbens and Rubin (1996)

[4]Note that $\frac{\hat{p}(w)}{1-\hat{p}(w)}$ converges in probability to $\frac{p(w)}{1-p(w)}$ and that this term does not depend on θ. Moreover, W is discrete so that $\sup_w(|\widehat{p(w)} - p(w)| = o_p(1)$.

define four types of individuals. The type of the individual, together with R determines the value of X. Consider the following types of individuals, $t_i \in \{\text{complier, always taker, never taker, defier}\}$. Moreover,

$R_i \in \{0,1\}$ is randomly assigned, $R_i \perp t_i, |W_i$

X_i can depend on W_i, t_i, ε_i.

$t_i =$ Complier $\Leftrightarrow X_i \equiv R_i$

$t_i =$ Always taker $\Leftrightarrow X_i \equiv 1$

$t_i =$ Never taker $\Leftrightarrow X_i \equiv 0$

$t_i =$ Defier $\Leftrightarrow X_i \equiv (1 - R_i)$.

Consider the following objective function

$$Q_N^*(\theta) = \frac{1}{N} \sum_{i \in Compliers} q(S_i; \theta).$$

where $S_i = \{Y_i, X_i, W_i\}$. Let

$$\hat{\theta} = \arg\max_{\theta} Q_N^*(\theta) \underset{p}{\rightarrow} \theta_0.$$

Latent variable model

$$Q_N^*(\theta) = \frac{1}{N} \sum_{i: v_i \in [\underline{v}(W_i,\eta_i), \overline{v}(W_i,\eta_i)]} q(S_i; \theta).$$

$$\hat{\theta} = \arg\max_{\theta} Q_N^*(\theta) \underset{p}{\rightarrow} \theta_0.$$

Examples: GMM, Likelihood, maximum score, quantile regression.

Consider

$p(W) = P(R = 1|W)$; estimator $\hat{p}(w)$.

$$Q_N(\theta) = \frac{1}{N} \sum_i w_i q(S_i; \theta)$$

where $w_i = \{1(R_i = X_i) - 1(R_i \neq X_i)(\frac{\hat{p}(w)}{1 - \hat{p}(w)})^{1-2R_i}\}$.

$$\begin{aligned} Q_N(\theta) &= \frac{1}{N} \sum_i w_i q(S_i; \theta) \\ &= \frac{1}{N} \sum_{i \in Com} w_i q(S_i; \theta) + \frac{1}{N} \sum_{i \in Alw} w_i q(S_i; \theta) + \frac{1}{N} \sum_{i \in Nev} w_i q(S_i; \theta) \end{aligned}$$

where $w_i = \{1(R_i = X_i) - 1(R_i \neq X_i)(\frac{\hat{p}(w)}{1 - \hat{p}(w)})^{1-2R_i}\}$.

Thus,

$$
\begin{aligned}
Q_N(\theta) = &\frac{1}{N}\sum_{i\in Com} q(S_i;\theta) \\
&+ \frac{1}{N}\sum_{i\in Alw} \{R_i + (1-R_i)\frac{\hat{p}(w)}{1-\hat{p}(w)}\} q(S_i;\theta) \\
&+ \frac{1}{N}\sum_{i\in Nev} \{(1-R_i) - R_i\frac{1-\hat{p}(w)}{\hat{p}(w)}\} q(S_i;\theta).
\end{aligned}
$$

and

$$
\begin{aligned}
Q_N(\theta) &= \frac{1}{N}\sum_{i} w_i q(S_i;\theta) \\
&= \frac{1}{N}\sum_{i\in Com} q(S_i;\theta) \\
&+ \frac{1}{N}\sum_{i\in Alw} \{\frac{R_i}{1-\hat{p}(w)} - \frac{\hat{p}(w)}{1-\hat{p}(w)}\} q(S_i;\theta) \\
&+ \frac{1}{N}\sum_{i\in Nev} \{\frac{(1-R_i)}{\hat{p}(w)} - \frac{1-\hat{p}(w)}{\hat{p}(w)}\} q(S_i;\theta).
\end{aligned}
$$

Assumption A1: $Cov(R_i, q(S_i;\theta)|i \in Alw) = Cov(R_i, q(S_i;\theta)|i \in Nev) = 0$ for all i.
Assumption A2: Let one of the following hold: (i) $P(R = 1|W, type) = p$ and $Cov(R_i, X_i) \neq 0$; or (ii) $P(R = 1|W, type) = P(R = 1|W) = p(W_k)$ where $k = 1, ..., K$ for some finite K, and $Cov(R, X|W) > 0$ for all $k = 1, ..., K$.

Assumption A3: $\hat{\theta} = \arg\max_\theta Q_N^*(\theta) \underset{p}{\rightarrow} \theta_0$.

Assumption A4: Let one of the following hold
(i) $q(S_i;\theta)$ is continuous with probability one; or
(ii) $\frac{1}{N}\sum_{i\in C} q(S_i;\theta)$, $\frac{1}{N}\sum_{i\in A} q(S_i;\theta)$, $\frac{1}{N}\sum_{i\in N} q(S_i;\theta)$, $\frac{1}{N}\sum_{i\in D} q(S_i;\theta)$ are stochastically equicontinuous
.
Assumption A5: $\theta \in \Theta$, which is compact.

Theorem 1
Let assumption A1-A5 hold. Then $\hat{\theta} = \arg\max Q_N(\theta)$ converges in probability to θ_0.

Theorem 1 covers many objective functions but not those with double summations. Let $Q_{2,N}$ denote such an objective function with a double summation. Consider the following function of the observed data,

$$Q_{2,N}(\theta) = \frac{1}{N^2}\sum_i [\{1(R_i = X_i) - 1(R_i \neq X_i)(\frac{\hat{p}(w)}{1-\hat{p}(w)})^{1-2R_i}\} *$$
$$* \sum_j \{1(R_j = X_j) - 1(R_j \neq X_j)(\frac{\hat{p}(w)}{1-\hat{p}(w)})^{1-2R_i}\} q(S_i, S_j; \theta)].$$

where $\hat{p}(w)$ is a consistent estimator for the probability $P(R = 1|w)$ and $R_i \in \{0, 1\}$. The maximum rank correlation estimator by Han (1987); the transformation model by Chen (2002); and duration model by Hausman and Woutersen (2005).

Transformation model

$$H(Y) = X\beta + \varepsilon$$
$$\varepsilon \perp X$$

Single index model

$$Y = G(X\beta) + \varepsilon$$
$$\varepsilon \perp X$$

Some duration models

$$\theta(y|v, x(y)) = v \cdot \exp(x(y)\beta) \cdot \lambda(y).$$

Theorem 2
Let assumption 6-11 hold. Then $\hat{\theta} = \arg\max Q_{2,N}(\theta)$ converges in probability to θ_0.

Derivatives or approximate derivatives (influence functions)

$$Q_N^*(\theta) = \frac{1}{N}\sum_{i \in Com} q(S_i; \theta)$$
$$Q_N(\theta) = \frac{1}{N}\sum_{i \in Com} q(S_i; \theta)$$
$$+ \frac{1}{N}\sum_{i \in Alw} \{\frac{R_i}{1-\hat{p}(w)} - \frac{\hat{p}(w)}{1-\hat{p}(w)}\} q(S_i; \theta)$$
$$+ \frac{1}{N}\sum_{i \in Nev} \{\frac{(1-R_i)}{\hat{p}(w)} - \frac{1-\hat{p}(w)}{\hat{p}(w)}\} q(S_i; \theta).$$

$$
\begin{aligned}
Q^*_{N,\theta}(\theta) &= \frac{1}{N} \sum_{i \in Com} q_\theta(S_i;\theta) \\
Q_{N,\theta}(\theta) &= \frac{1}{N} \sum_{i \in Com} q_\theta(S_i;\theta) \\
&\quad + \frac{1}{N} \sum_{i \in Alw} \{\frac{R_i}{1-\hat{p}(w)} - \frac{\hat{p}(w)}{1-\hat{p}(w)}\} q_\theta(S_i;\theta) \\
&\quad + \frac{1}{N} \sum_{i \in Nev} \{\frac{(1-R_i)}{\hat{p}(w)} - \frac{1-\hat{p}(w)}{\hat{p}(w)}\} q_\theta(S_i;\theta).
\end{aligned}
$$

$$
\begin{aligned}
\sqrt{N}\{Q_{N,\theta}(\theta) - Q^*_{N,\theta}(\theta)\} &= +\frac{1}{\sqrt{N}} \sum_{i \in Alw} \{\frac{R_i}{1-\hat{p}(w)} - \frac{\hat{p}(w)}{1-\hat{p}(w)}\} q_\theta(S_i;\theta) \\
&\quad + \frac{1}{\sqrt{N}} \sum_{i \in Nev} \{\frac{(1-R_i)}{\hat{p}(w)} - \frac{1-\hat{p}(w)}{\hat{p}(w)}\} q_\theta(S_i;\theta).
\end{aligned}
$$

Second derivatives or approximate second derivatives

$$
\begin{aligned}
\{Q_{N,\theta\theta}(\theta) - Q^*_{N,\theta\theta}(\theta)\} &= \frac{1}{N} \sum_{i \in Alw} \{\frac{R_i}{1-\hat{p}(w)} - \frac{\hat{p}(w)}{1-\hat{p}(w)}\} q_{\theta\theta}(S_i;\theta) \\
&\quad + \frac{1}{N} \sum_{i \in Nev} \{\frac{(1-R_i)}{\hat{p}(w)} - \frac{1-\hat{p}(w)}{\hat{p}(w)}\} q_{\theta\theta}(S_i;\theta) \\
&\underset{p}{\rightarrow} 0.
\end{aligned}
$$

4 The Pennsylvania Reemployment Bonus Demonstration

4.1 Data

The Pennsylvania bonus experiment was conducted by the U.S. Department of Labor between July 1988 and October 1989. During the enrollment period, claimants who became unemployed and registered for unemployment benefits in one of the selected local offices throughout the state were randomly assigned either to a control group or one of the six experimental treatment groups. In the control group the existing rules of the unemployment insurance system applied. Individuals in the treatment groups were offered a cash bonus if they became reemployed in a full-time job, working more than 32 hours per week, within a qualification period. In addition, to qualify for the bonus, claimants were required to work in the new job continuously for at least 16 weeks, or they were allowed to change jobs as long as the transition took place within a period of 5 days. The latter requirements were imposed to discourage cases of fraudulent hiring for purposes of obtaining the bonus, and to avoid the possibility of bonus payments to seasonal workers. Two bonus levels were tested. The lower bonus was three times the weekly benefit amount, and the higher bonus was six times the weekly benefit. The low bonus averaged 500 dollars and the high bonus averaged 997 dollars. The two levels were chosen on the basis of both the Illinois and New Jersey experiences. Two qualification periods were considered: a short period of 6 weeks and a longer one of 12 weeks. The long qualification period was close to that studied in Illinois and New Jersey. The choice of the shorter period was intended to test the sensitivity of the treatment effect to alternative specifications of the qualification periods. All of the treatments, except the last one, involved a voluntary option of attending a workshop designed to aid job search. However, less than three percent of eligible participants attended the workshop so we follow the practice established by prior analysts of ignoring the workshop option. Four of the treatments were created by the combination of a bonus amount and a qualification period plus the offer of the workshop. The fifth treatment included an initially high, but declining bonus over the period of 12 weeks plus the optional workshop. The sixth treatment combined the high bonus with the long qualification period without the workshop. The 6 treatment groups are:

- Treatment 1: Low bonus, short fixed qualification period of 6 weeks for all individuals and workshop.
- Treatment 2: Low bonus, long fixed qualification period of 12 weeks for all individuals and workshop
- Treatment 3: High bonus, short fixed qualification period of 6 weeks for all individuals and workshop.
- Treatment 4: High bonus, long fixed qualification period of 12 weeks for all individuals and workshop.
- Treatment 5: Initially high but declining bonus, long fixed qualification period of 12 weeks for all individuals and workshop.
- Treatment 6: High bonus, long fixed qualification period of 12 weeks for all individuals, no workshop.

The design consisted of 3,000 control and 10,120 treatment members, allocated to the specific treatments. The sample was drawn randomly from claimants at twelve Job Services (JS) offices located throughout the state of Pennsylvania. The limited selection of sites constituted a compromise between the need to obtain a fairly large sample that could accurately reflect the demographic and occupational characteristics of the state, and the need for an easy monitoring and low operational cost of the study. Twelve local offices were selected which were representative of the insured unemployed population of Pennsylvania. More specifically, the state was divided in eight UI/JS regions. One or more clusters of local offices were formed within each region according to average duration of UI receipt. This process produced twelve clusters of approximately equal size UI caseloads. Finally, one office was selected randomly from each cluster to participate in the demonstration and a random sample of UI claimants from each of the selected offices was selected in a manner which ensured that each eligible claimant in the state had an equal probability of selection into the demonstration sample. Eligibility criteria were imposed to increase the homogeneity of the sample and thus ensure that possible differences in the response could be attributed primarily to variation in treatment. Claimants who filed for a transitional claim were excluded because of the likelihood of a previous job offer. For the same reason there was exclusion from the experiment of individuals who indicated that it was possible they could find a new job through a union channel rather than the market, or if they were waiting for some definite recall within 60 days from their former employer.

The collected sample was the result of fifty-two weekly sub-samples selected in all twelve offices beginning on October 26, 1988. Prior to that date, fifteen weekly sub-samples were drawn from one site for a pilot test of all operations, which are also included in the final collected-sample. Thus, the enrollment period for the experiment started July 1988 and ended October 1989. The design target was to identify and select 13,120 claimants with each site contributing roughly 1,100 individuals in total and a weekly target of 21 claimants per site. However, since some claimants who initially apply for benefits do not return to a local office to file further (they need to wait for one week and return to the local office and file again) , a larger sample was selected to achieve the desired sample size for analysis. Thus, a sample ranging from 22 to 40 claimants was selected at each office per week, depending on the historical experience. Overall, 15,005 individuals were initially selected to participate in the demonstration. A total of 14,086 individuals filed for a week of UI and were included in the study. The attendance in workshop was less than 3 percent, which made the fourth and the sixth treatments indistinguishable. Therefore, as of July 1989, four months before the end of the experiment, individuals who would have been assigned to treatment 6 were assigned to the other treatments. A second change was made because preliminary demonstration results showed that treatment 1 had a larger than expected effect. Initially only a small proportion of the total sample was assigned to this treatment due to its perceived low policy significance. Beginning October 1989, experimenters increased its sample. This change is reflected in the relatively high percentage, 18.3 percent of entries during the last quarter.

The duration of the of UI benefits is measured in weeks. This measure of duration is called "inuidur" in the final report of the experiment. It is worth noting that a large portion of spells end in the first and the twenty seventh week. It should be stressed that the definition of the first spell of UI in the Pennsylvania study includes a waiting week and that the maximum number of uninterruptedly received full weekly benefits is 26. This implies that a total 2491 subjects did not receive any weekly benefit and that most of the claimants received continuously their full, entitled unemployment benefit. A properly made randomization implies that any difference between the length of unemployment insurance of claimants receiving the treatment and those that do not can be attributed exclusively to the treatment effect. Despite the intermediate changes in the rate of entry in the various groups, it is generally considered that the randomization process was effective; see Corson et al. (1992, p 45) and Meyer (1995, p 98).

4.2 Model

Consider the following hazard model $\theta\left(t|v,x\right) = v\phi^{x}\lambda\left(t\right),$ where $\lambda\left(t\right)$ is baseline hazard and is non-parametric, the integrated baseline hazard is defined as $\Lambda\left(t\right)$ and v is the unobserved heterogeneity.
Now, consider the unobserved heterogeneity has 3 discrete points ($v = \{v_1, v_2, v_3\}$), with associated probabilities $p = \{p_1, p_2, p_3 = 1 - p_1 - p_2\}$.
Given the hazard rate $\theta\left(t|v,x\right) = \frac{f(t|v,x)}{1-F(t|v,x)}$, the conditional density function, $f\left(t|v\right),$ is derived as:

$$f\left(t|v,x\right) = \theta\left(t|v,x\right)\left(1 - F\left(t|v,x\right)\right),$$

where $1 - F\left(t|v,x\right) = e^{-ve^{x\beta}\Lambda(t)}$.
The unconditional likelihood for a given $\lambda\left(t\right)$ and for density function of a type "k" individual, $f_k\left(t|v_k,x\right) = v_k e^{x\beta}\lambda\left(t\right)e^{-v_k e^{x\beta}\Lambda(t)}$, with $k = \{1,2,3\}$,

is given by

$$f\left(t|x\right)=\sum_{k=1}^{3}p_{k}f_{k}\left(t|v_{k},x\right)=e^{x\beta}\sum_{k=1}^{3}p_{k}v_{k}\lambda\left(t\right)e^{-v_{k}e^{x\beta}\Lambda(t)},$$

where p_k is the probability of a type k individual.
Define θ as the vector of parameters that need to be estimated (i.e. $\theta = \{p_1, p_2, v_1, v_2, v_3, \beta\}$).
Define now the log-likelihood function

$$\begin{aligned}L(\theta|data) = \ln(l(\theta|data)) &= \sum_{i=1}^{N}\ln(e^{x\beta}\sum_{k=1}^{3}p_{k}v_{k}\lambda(t)e^{-v_{k}e^{x\beta}\Lambda(t)}) \\ &= \sum_{i=1}^{N}x_{i}\beta+\sum_{i=1}^{N}ln(\sum_{k=1}^{3}p_{k}v_{k}\lambda(t)e^{-v_{k}e^{x\beta}\Lambda(t)}).\end{aligned}$$

Suppose that there was not endogeneity. We would then optimize the following objective function,

$$\widehat{\theta_{MLE}} = argmax_{\theta}L(\theta|data)$$

4.3 Likelihood for a piecewise constant $\lambda(t)$ and 5 types heterogeneity

Now, let $\lambda(t)$ be a piecewise constant function. In this case the hazard rate is defined as

$$\theta\left(t|v,x\right)=\{\begin{array}{l}\eta e^{x\beta},\ if\ t\leq 1\\ \eta e^{x\beta}\lambda_{j},\ if\ j-1<t\leq j, j=2,\dots,T\end{array}$$

Define the indicator d_j, such that

$$d_1=\{\begin{array}{l}1\text{ if }t\leq 1\\ 0\text{ otherwise}\end{array}, d_j=\{\begin{array}{l}1\text{ if }j-1<t\leq j, j=2,\dots,T\\ 0\text{ otherwise}\end{array}.$$

Then, the integrated baseline hazard is

$$\begin{aligned}\Lambda\left(t\right) &= \int_{0}^{t}\lambda\left(t\right)dt=\left(\int_{0}^{t}dt\right)d_{1}+\sum_{j=1}^{T-1}\left(1+\sum_{s=2}^{j}\int_{s-1}^{s}\lambda_{s-1}dt+\int_{j}^{t}\lambda_{j}dt\right)d_{j+1}\\ &= td_{1}+\sum_{j=1}^{T-1}\left(1+\sum_{s=2}^{j}\lambda_{s-1}+(t-j)\lambda_{j}\right)d_{j+1},\end{aligned}$$

with $\sum_{s=2}^{1}\int_{s-1}^{s}\lambda_{s-1}dt=0$.
In this case the unconditional likelihood is given by

$$f\left(t|x\right)=\sum_{k=1}^{5}p_k f_k\left(t|v_k,x\right)=\sum_{k=1}^{5}p_k v_k e^{x\beta}\lambda(t)e^{-v_k e^{x\beta}\left(td_1+\sum_{j=1}^{T-1}\left(1+\sum_{s=2}^{j}\lambda_{s-1}+(t-j)\lambda_j\right)d_{j+1}\right)},$$

where p_k is the probability of a type $k=\{1,..,5\}$.
Considering again θ as the vector of parameters.
If we use the indicators d_j,with $j=1,2,...,T$, to estimate θ, we need to maximize the following log-likelihood

$$\begin{aligned}L(\theta|data)=\ln(l(\theta|data))=&\sum_{i=1}^{N}x_i\beta\\&+\sum_{i=1}^{N}(\sum_{j=1}^{T}d_j ln(\sum_{k=1}^{5}p_k v_k\lambda_j e^{-v_k e^{x\beta}\left(td_1+\sum_{j=1}^{T-1}\left(1+\sum_{s=2}^{j}\lambda_{s-1}+(t-j)\lambda_j\right)d_{j+1}\right)}).\end{aligned}$$

4.4 Results

For the available data we use a model with three points heterogeneity, ($v=\{v_1,v_2,v_3\}$). To test for the time-varying treatment effect, we conducted several experiments. In the first experiment we generated two dummy variables, the first one has a value of one if individual's duration of unemployment is within the bonus eligibility period and the second dummy has a value of one if individual's duration of unemployment is greater than the bonus eligibility period. We also allow the baseline hazard to change its value at half period of the bonus eligibility (after 6 weeks), at the end of the bonus eligibility period (after 12 weeks), and at the end of the UI eligibility period (after 26 weeks).

$$\theta\left(t|v,x\right)=\begin{array}{l}\eta e^{\beta_1},\quad if\ t\leq 6\\ \eta e^{\beta_1}\lambda_2, if\ 6<t\leq 12\\ \eta e^{\beta_2}\lambda_3, if\ 12<t\leq 18\\ \eta e^{\beta_2}\lambda_4, if\ 18<t\leq 26.\end{array}$$

The results of this experiment are presented in Table 1.

In the second experiment we generated three dummy variables, the first one has a value of one if individual's duration of unemployment is within half the bonus eligibility period, the second dummy has a value of one if individual's duration of unemployment is within the bonus eligibility period and the third dummy has a value of one if individual's duration of unemployment is greater than the bonus eligibility period. Again, we allow the baseline hazard to change its value at half period of the bonus eligibility, at the end of the bonus eligibility period , and at the end of the UI eligibility period.

$$\theta\left(t|v,x\right) = \begin{array}{l} \eta e^{\beta_1}, \quad if\ t \leq 6 \\ \eta e^{\beta_1}\lambda_2, if\ 6<t \leq 12 \\ \eta e^{\beta_2}\lambda_3, if\ 12<t \leq 18 \\ \eta e^{\beta 3}\lambda_4, if\ 18 < t \leq 26. \end{array}$$

The results of this experiment are presented in Table 2.

The results show that bonus is more effective during the first 6 weeks of unemployment.

5 Appendix: The 2SLS and WOF-IV estimator

In this appendix, we show that the 2SLS and WOF-IV estimator are equivalent if one has only one instrument and one an endogenous dummy as explanatory variable.

Model: Let $\{R_i, X_i, \varepsilon_i\}$ be a random sample of size N. For all i, let $R_i \in \{0,1\}$, $X_i \in \{0,1\}$, $Y_i = \alpha + \beta X_i + \varepsilon_i$ where $E\{\varepsilon_i|R_i\} = 0$.

The two stage least squares estimator has the following form,

$$\hat{\beta}_{2SLS} = \frac{\sum_i X_i(R_i - \hat{p})Y_i}{\sum_i X_i(R_i - \hat{p})X_i}$$

where $\hat{p} = \frac{\sum_i R_i}{N}$.

The weighting estimator uses the following objective function,

$$Q(\alpha,\beta) = \frac{1}{N}\sum_i h_i(Y_i - \alpha - X_i\beta)^2$$

where

$$h_i = 1(R_i = X_i) - 1(R_i \neq X_i)(\frac{\hat{p}}{1-\hat{p}})^{1-2R_i}.$$

Minimizing $Q(\alpha, \beta)$ yields the following two first order conditions,

$$\frac{1}{N}\sum_i h_i(Y_i - \alpha - X_i\beta) = 0$$

$$\frac{1}{N}\sum_i h_i X_i(Y_i - \alpha - X_i\beta) = 0$$

This yields

$$\frac{1}{N}\sum_i h_i X_i\{(Y_i - \frac{\sum_i h_i Y_i}{\sum_i h_i}) - (X_i\beta - \frac{\sum_i h_i X_i\beta}{\sum_i h_i})\} = 0$$

and

$$\hat{\beta}_W = \frac{\sum_i\{h_i X_i Y_i - h_i X_i \frac{\sum_i h_i Y_i}{\sum_i h_i}\}}{\sum_i\{h_i X_i - h_i X_i \frac{\sum_i X_i \cdot h_i}{\sum_i h_i}\}}.$$

Note that

$$h_i X_i = \frac{X_i(R_i - \hat{p})}{1 - \hat{p}}.$$

Our algebra simplifies if we write h_i as follows.

$$\begin{aligned}
h_i &= 1(R_i = X_i) - 1(R_i \neq X_i)\{R_i(\frac{1-\hat{p}}{\hat{p}}) + (1 - R_i)(\frac{\hat{p}}{1-\hat{p}})\} \\
&= R_i X_i + (1 - R_i)(1 - X_i) - R_i(1 - X_i)(\frac{1-\hat{p}}{\hat{p}}) - (1 - R_i)X_i(\frac{\hat{p}}{1-\hat{p}}) \\
&= R_i X_i\{2 + \frac{1-\hat{p}}{\hat{p}} + \frac{\hat{p}}{1-\hat{p}}\} - R_i - R_i(\frac{1-\hat{p}}{\hat{p}}) - X_i - X_i(\frac{\hat{p}}{1-\hat{p}}) + 1 \\
&= R_i X_i\{\frac{1}{\hat{p}(1-\hat{p})}\} - \frac{R_i}{\hat{p}} - \frac{X_i}{1-\hat{p}} + 1 \\
&= \frac{X_i}{\hat{p}(1-\hat{p})}\{R_i - \hat{p}\} - \frac{R_i - \hat{p}}{\hat{p}} \\
&= \frac{h_i X_i}{\hat{p}} - \frac{R_i - \hat{p}}{\hat{p}}.
\end{aligned}$$

Note that

$$\sum_i h_i = \frac{1}{\hat{p}}\sum_i h_i X_i$$

The denominator of $\hat{\beta}_W$ can be written as follows,

$$\begin{aligned}
\sum_i\{h_i X_i - h_i X_i \frac{\sum_i X_i h_i}{\sum_i h_i}\} &= \sum_i h_i X_i - \hat{p}\sum_i h_i X_i \\
&= \sum_i X_i(R_i - \hat{p}).
\end{aligned}$$

Thus, $\hat{\beta}_W$ has the same denominator as $\hat{\beta}_{2SLS.}$ Similarly, the numerator of $\hat{\beta}_W$ equals

$$\sum_i \{h_i X_i Y_i - h_i X_i \frac{\sum_i h_i Y_i}{\sum_i h_i}\} = \sum_i h_i X_i Y_i - \hat{p} \sum_i h_i Y_i.$$

Using $h_i = \frac{h_i X_i}{\hat{p}} - \frac{R_i - \hat{p}}{\hat{p}}$ yields

$$\begin{aligned}\sum_i \{h_i X_i \cdot Y_i - h_i X_i \frac{\sum_i h_i Y_i}{\sum_i h_i}\} &= \sum_i h_i X_i Y_i - \sum_i h_i X_i Y_i + \sum_i (R_i - \hat{p}) Y_i \\ &= \sum_i (R_i - \hat{p}) Y_i.\end{aligned}$$

Thus, $\hat{\beta}_W$ has the same numerator as $\hat{\beta}_{2SLS}$ and the two estimators are equivalent.

References

[1] Angrist, J., G. Imbens, and D. Rubin (1996): "Identification of Causal effects Using Instrumental Variables" *Journal of Econometrics*, Vol. 71, 145-160

[2] Chesher, A. (2005): "Nonparametric identification under discrete variation", *Econometrica*, Vol. 73, No. 5, p. 1525-1550.

[3] Chesher, A. (2007): "Instrumental values", Journal of Econometrics, forthcoming

[4] Chernozhukov, V., H. Hong and E. Tamer (2003): "Parameter Set Inference in a Class of Econometric Models," unpublished manuscript, Department of Economics, Princeton University.

[5] Heckman, J., H. Ichimura, and P. Todd (1997): "Matching as an econometric evaluation estimator: Evidence from Evaluating a Job Training Program," Review of Economic Studies, Vol. 64(4).

[6] Heckman, J., H. Ichimura, and P. Todd (1997): "Matching as an econometric evaluation estimator", Review of Economic Studies, Vol. 65(2).

[7] Heckman, J., H. Ichimura, and P. Todd, and J. Smith (1998): "Characterizing Selection Bias Using Experimental Data", Econometrica Vol. 66.

[8] Horowitz, J. L. (2001): "The Bootstrap" in *Handbook of Econometrics,* Vol. 5, ed. by J. J. Heckman and E. Leamer. Amsterdam: North-Holland.

[9] Horowitz, J. L. and S. Lee (2007): "Nonparametric Instrumental Variables Estimation of a Quantile Regression Model"

[10] Horowitz, J. L., and C. Manski (1998): "Censoring of Outcomes and Regressors due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations," *Journal of Econometrics*, 84, 37-58.

[11] Imbens, G. and W. K. Newey (2002): Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity.

[12] Imbens, G. W. and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters", *Econometrica, 72,* 1845-1857.

[13] Manski, C. (1990): “Nonparametric Bounds on Treatment Effects,” *American Economic Review Papers and Proceedings*, 80, 319-323.

[14] Manski, C. (2003): “*Partial Identification of Probability Distributions*”, New York: Springer-Verlag.

[15] Newey, W. K. (1991): “Uniform Convergence in Probability and Stochastic Equicontinuity”, *Econometrica*, 59, 1161-1167.

[16] Newey, W. K., and D. McFadden (1994): “Large Sample Estimation and Hypothesis Testing,” in *Handbook of Econometrics,* Vol. 4, ed. by R. F. Engle and D. MacFadden. Amsterdam: North-Holland.

[17] Newey, W. K., and J. Powell (2003): “Instrumental Variables Estimation of Nonparametric Models” *Econometrica*, 71.

[18] Powell, J. and R. Blundell (2003): Endogeneity in Nonparametric and Semiparametric Regression Models, in Advances in Economics and Econometrics: Theory and Applications, Eight World Congress, Vol. II, M. Dwatripont, L.P. Hansen and S.J. Turnovsky, eds. Cambridge: Cambridge University Press.

[19] Shaikh, A. M., and E. Vytlacil (2005): “Dummy Endogenous Variables in Weakly Separable Models”, unpublished manuscript

[20] Vytlacil, E. (2002): “Independence, Monotonicity, and Latent Index Models: An Equivalence Result,” *Econometrica*, 70(1): 331-341.

[21] Vytlacil, E. and Nese Yildiz. (2007) Dummy Endogenous Variables in Weakly Separable Models. Econometrica 75:3, 757–779