

An Overview of the Special Regressor Method

Arthur Lewbel*

Boston College

September 2012

Abstract

This chapter provides background for understanding and applying special regressor methods.

This chapter is intended for inclusion in the "Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics," Co-edited by Aman Ullah, Jeffrey Racine, and Liangjun Su, to be published by Oxford University Press

1 Introduction

The goal of this chapter is to provide some background for understanding how and why special regressor methods work, and their application to identification and estimation of latent variable moments and parameters. Other related surveys include Dong and Lewbel (2012), who describe the simplest estimators for applying special regressor methods to binary choice problems (particularly those involving endogenous regressors), and Lewbel, Dong, and Yang (2012) who provide a comparison of special regressor methods

*I would like to thank Thomas (Tao) Yang for helpful research assistance. Corresponding address: Arthur Lewbel, Dept of Economics, Boston College, 140 Commonwealth Ave., Chestnut Hill, MA, 02467, USA. (617)-552-3678, lewbel@bc.edu, <http://www2.bc.edu/~lewbel/>

to other types of estimators, specifically, to control functions, maximum likelihood, and linear probability models for binary choice model estimation.

A special regressor is an observed covariate with properties that facilitate identification and estimation of a latent variable model. For example, suppose an observed binary variable D satisfies $D = I(V + W^* \geq 0)$, where V is the observed special regressor, and W^* is an unobserved latent variable. Such a W^* will exist as long as the probability that $D = 1$ is increasing in V . The goal is estimation of the distribution of W^* , or estimation of features of its distribution like a conditional or unconditional mean or median of W^* . Many standard models have this form, e.g, a probit model has $W^* = X'\beta + \varepsilon$ with ε normal, and estimating β would correspond to estimating the conditional mean of W^* given X . A simple probit doesn't require a special regressor, but the special regressor would be useful here if ε is heteroscedastic with an unknown distribution, or if some of the regressors in X are endogenous. Special regressor methods work by exploiting the fact that if V is independent of W^* (either unconditionally or after conditioning on covariates) then variation in V changes the probability that $D = 1$ in a way that traces out the distribution of W^* (either the unconditional distribution or the distribution conditional on covariates).

The term special regressor was first used in Lewbel (1998), but the commonest class of applications for the special regressor is to binary choice models as described in Lewbel (2000). The special regressor method has been employed in a wide variety of limited dependent variable models including binary, ordered, and multinomial choice as well as censored regression, selection and treatment models (Lewbel 1998, 2000, 2007a, Magnac and Maurin 2007, 2008) truncated regression models (Khan and Lewbel 2007), binary and other nonlinear panel models with fixed effects (Honore and Lewbel 2002, Ai and Gan 2010, Gayle 2012), contingent valuation models (Lewbel, Linton, and McFadden 2011), dynamic choice models (Heckman and Navarro 2007, Abbring and Heckman 2007), market equilibrium models of multinomial discrete choice (Berry and Haile 2009a, 2009b), models of games, including entry games and matching games (Lewbel and Tang 2011, Khan and Nekipelov 2011, Fox and Yang 2012), and a variety of models with (partly) nonseparable errors (Lewbel 2007b, Matzkin 2007, Briesch, Chintagunta, and

Matzkin 2009).

Additional empirical applications of special regressor methods include Anton, Fernandez Sainz, and Rodriguez-Poo (2002), Maurin (2002), Cogneau and Maurin (2002), Goux and Maurin (2005), Stewart (2005), Avelino (2006), Pistolesi (2006), Lewbel and Schennach (2007), and Tiwari, Mohnen, Palm, and van der Loeff (2007). Earlier results that can be reinterpreted as special cases of special regressor based identification methods include Matzkin (1992, 1994) and Lewbel (1997). Vytlacil and Yildiz (2007) describe their estimator as a control function, but their identification of the endogenous regressor coefficient essentially treats the remainder of the latent index as a special regressor. Recent identification and limiting distribution theory involving special regressor models include Jacho-Chávez (2009), Khan and Tamer (2010), and Khan and Nekipelov (2010a, 2010b).

The remainder of this chapter lays out the basic ideas behind special regressor methods. The focus here is on identification and associated construction of estimators, not on limiting distribution theory. Most of the estimators provided here are multistep estimators, where each step takes the form of a standard parametric or nonparametric density or regression estimator.

The next section provides the basic idea of how a special regressor can identify the distribution of a latent variable, specifically the latent index W^* in a threshold crossing binary choice model $D = I(V + W^* \geq 0)$. This is followed by sections that look at estimation of unconditional moments of the latent index such as $E(W^*)$, and later on estimation of conditional moments like $E(W^* | X)$, conditioning on covariates X . This will then lead to estimation of coefficients β in discrete choice models $D = I(V + X'\beta + \varepsilon \geq 0)$ when the latent error ε is heteroskedastic and has unknown distribution. Later sections will consider semiparametric instrumental variables estimators for these models when the regressors X are mismeasured or endogenous. The final sections consider some extensions, including allowing the special regressor V to be discrete, and consideration of other latent variable models.

2 Identifying a Latent Marginal Distribution

To illustrate the main ideas behind special regressor identification and estimation, begin with an example from Lewbel, Linton, and McFadden (2011). Suppose we want to uncover the distribution of people's willingness to pay (wtp) W^* to preserve a wetland habitat. Denote this distribution function as $F_{W^*}(w)$. For simplicity in this exposition, assume the distribution of W^* is continuously distributed.

A survey is performed in which, for each person in the sample, researchers draw a price P from some distribution function chosen by the researcher, and ask the sampled individual if they would be willing to pay P dollars or more to preserve the wetland. Let D denote an individual's response ($D = 1$ for yes or $D = 0$ for no), so $D = I(W^* \geq P)$, where W^* is the individual's unobserved (latent) willingness to pay, and I is the indicator function that equals one if its argument is true and zero otherwise. The experiment is designed so that prices are randomly assigned to individuals, meaning that the distribution that P is drawn from is independent of W^* . Let $E(D | P = p)$ denote the conditional expectation of D , conditioning on the event that the random variable P equals the value p . By construction of the experiment, P is drawn independently of each subject's willingness to pay, so P is distributed independently of W^* . It follows from this independence that $E(D | P = p) = \Pr(W^* \geq p) = 1 - \Pr(W^* < p) = 1 - F_{W^*}(p)$. For example, suppose that among a random sample of people, 70% said they would be *not* be willing to pay more than \$50 to preserve the wetland. In this example $p = 50$ and so 0.7 would be an unbiased estimate of $1 - E(D | P = 50) = F_{W^*}(50)$.

In the data set examined by An (2000) and reconsidered by Lewbel, Linton, and McFadden (2011), P takes on one of 14 different values, so without additional assumptions this data could identify the distribution function $F_{W^*}(w^*)$ only at $w^* = p$ for these 14 values of p . To identify the entire distribution function F_{W^*} , it would be desirable to design the experiment so that P can take on any value that the willingness-to-pay W^* could equal, meaning that each p should be drawn from a continuous distribution with support equal to an interval that is at least as large as the range of possible values of W^* .¹

¹Past empirical practice has not been to design experiments sufficient to nonparametrically identify the willingness to pay distribution by drawing from a continuous distribution. Instead, the custom in the literature on valuing public goods is to

In this application, P is a special regressor that allows us to identify the distribution function $F_{W^*}(w^*)$ at points $w^* = p$ in the distribution of P . If the experiment were designed so that the distribution of P was continuous, then the entire distribution of W^* would be identified as the number of subjects goes to infinity, and could be consistently estimated by a nonparametric regression of $1 - D$ on P , since $E(1 - D | P = p) = F_{W^*}(p)$.

To be consistent with later notation on latent variable models (e.g., the notation of a probit model), let $V = -P$, so $D = I(V + W^* \geq 0)$. Suppose P and hence V is continuously distributed, and let $F_V(v)$ denote the distribution function of the special regressor V . Let $H(v) = E(D | V = v)$ and let $\widehat{H}(v)$ be a uniformly consistent estimator of $H(v)$, such as a Nadaraya-Watson local constant or local linear kernel regression of D on V . Then, replacing p above with $-v$, we have $H(v) = 1 - F_{W^*}(-v)$, and therefore $1 - \widehat{H}(-w^*)$ is a consistent estimator of $F_{W^*}(w^*)$. If the support of V contains the support of $-W^*$, then the entire distribution function $F_{W^*}(w^*)$ would be identified and consistently estimated by $1 - \widehat{H}(-v)$ for values of v in the support of V .

In this example the special regressor V allows us to recover the distribution of the latent variable W^* in a limited dependent variable model. Suppose we wanted to estimate $E(W^*)$, the average willingness to pay in the population. Let w_L and w_U denote lower and upper bounds on the range of values w^* can take on, and let f_{w^*} denote the probability density function of w^* . Then

$$E(W^*) = \int_{w_L}^{w_U} w^* f_{w^*}(w^*) dw^* = \int_{w_L}^{w_U} w^* \frac{\partial F_{W^*}(w^*)}{\partial w^*} dw^* = \int_{w_L}^{w_U} w^* \frac{\partial [1 - H(-w^*)]}{\partial w^*} dw^*. \quad (1)$$

This equation shows that $E(W^*)$ is identified, and could be consistently estimated by replacing H with \widehat{H} in the above integral. The next section will provide a much simpler estimator for $E(W^*)$.

The key features of the special regressor V that allow us to recover the distribution of the latent W^* are independence (later this will be relaxed to conditional independence), additivity, continuity, and large make functional form assumptions that suffice to identify the desired features of the willing to pay distribution from the few points on the F_{W^*} distribution that are revealed by standard experiments. Lewbel, Linton, and McFadden (2011) observe that identification could be obtained by increasing the number of mass points in the distribution of P in the experimental design as the sample size grows, to become asymptotically dense on the line. They supply associated limiting distribution theory for estimation based on that design. See section 14 below, regarding when the special regressor is discrete.

support. In this example V is statistically independent of the latent W^* , V appears added to the latent W^* in the model $D = I(V + W^* \geq 0)$, V is continuously distributed, and the range of values V takes on is at least as great as the range that $-W^*$ takes on. Having one regressor in a model that satisfies conditional independence, additivity, continuity, and large support is the typical set of requirements for special regressor methods, though these features can sometimes be relaxed in various directions.

Note in particular that large support is not required, but without it the distribution of W^* will only be identified at the values $-V$ can take on. This may suffice for some applications. For example, if support for the wetland will be decided by a referendum, then by the median voter model the maximum amount of money that can be raised to protect the wetland will be obtained by having people vote on whether they are willing to pay the median of the distribution of W^* . So in this application we only need the range of $-V$ to include the median of W^* , which would be estimated as $-v$ for the value v that makes $\widehat{H}(v) = 1/2$. However, if we instead wish to estimate the mean of W^* , then large support will generally be required.²

These requirements for special regressors can be restrictive, but many discrete choice estimators make use of similar restrictions like these, though sometimes implicitly. For example, Manski's (1985) maximum score and Horowitz's (1992) smooth maximum score estimator papers assume the presence of a regressor with these properties. Except for large support, standard logit and probit models assume all regressors satisfy these properties.

The above results can be generalized in a variety of directions. For example, suppose we had the more general model $Y = g(V + W^*)$ for some possibly unknown function g . Here V may now be thought of as just a regressor, not necessarily set by experimental design. As long as g is nonconstant and weakly monotonic, then the mean and all other features of the distribution of W^* can still be identified (up to an implicit location normalization) by letting $D = I(Y \geq y_0)$ for any constant y_0 in the support of Y that makes D be nonconstant, and then applying the same methods as above.

Even more generally, if Y depends on two or more latent variables, then their distribution can be

²While estimation of the mean formally requires knowing $F_{w^*}(w^*)$ on the entire support of w^* , if the tails of F_{w^*} are thin, then the bias of failing to estimate those tails will be small. Magnac and Maurin (2007) provide conditions, call tail symmetry, that make this bias not just small but actually zero, permitting consistent estimation of means without large support.

generally identified if we have a separate special regressor for each latent variable.

To summarize, the key result derived in this section is that, in a latent index model of the form $D = I(V + W^* \geq 0)$, variation in the special regressor V can be used to identify and estimate the distribution function of the unobserved latent variable W^* , based on $F_{W^*}(w^*) = 1 - E(D | V = -w^*)$ for any constant w^* in the support of $-V$. This result more generally shows identification of $Y = g(V + W^*)$ where g is weakly monotonic.

3 Unconditional Moments

Identification and hence estimation of the mean willingness to pay $E(W^*)$ based on equation (1) is somewhat elaborate, since this equation involves integrating a function of the derivative of the conditional expectation function H . In this section we derive a simpler expression for $E(W^*)$. Let $f_v(v)$ denote the probability density function of V , which in the experimental design context is determined by, and hence known to, the researcher. Let c be any constant in the interior of the supports of $-W^*$. Define the variable T_1 by

$$T_1 = \frac{D - I(V \geq c)}{f_v(V)} - c. \quad (2)$$

Then it will be shown below that

$$E(T_1) = E(W^*). \quad (3)$$

Using equation (3), the mean of W^* can be consistently estimated just by constructing the variable $T_{1i} = [D_i - I(V_i \geq 0)] / f_v(V_i)$ for each individual i in a sample, and then letting the estimate of $E(W^*)$ be the sample average $\sum_{i=1}^n T_{1i} / n$. Here we have made the simplest choice for the constant c of letting $c = 0$.

The derivation of $E(T_1) = E(W^*)$ has two main components. One component is to show that

$$E(W^*) = -c + \int_{w_L}^{w_U} [H(-w^*) - I(-w^* \geq c)] dw^*. \quad (4)$$

The crux of the argument behind equation (4) is that, by definition of an expectation $E(W^*) = \int w^* dF_{w^*}(w^*)$ and using integration by parts, this integral is equivalent to $-\int F_{w^*}(w^*) dw^*$ plus boundary terms. Since

$F_{w^*}(w^*) = 1 - H(-w^*)$ we get $\int H(-w^*) dw^*$ plus boundary terms. The role of c and of $I(-w^* \geq c)$ in equation (4) is to handle these boundary terms.

The second component to proving equation (3) is to show that $E(T_1)$ equals the right side of equation (4). This is shown by rewriting the integral in equation (4) as $\int_{-w_U}^{-w_L} [H(v) - I(v \geq c)] dv$ and plugging in $H(v) = E(D | V = v)$. The argument inside this integral can then be rewritten as $E(T_1 + c | V = v) f_v(v)$. Notice that the c inside the expectation will be cancelled out with $-c$ before the integral in equation (4), so the right hand side of equation (4) is equivalent to an expectation over V , giving $E(T_1)$.

The remainder of this section writes out these steps of the proof of equation (3) in detail. To ease exposition, technical conditions required for applying integration by parts, and for allowing the limits of integration w_L and w_U to be infinite, will be ignored here.

We begin by deriving equation (4), applying the following integration by parts argument.

$$\begin{aligned}
-c + \int_{w_L}^{w_U} [H(-w^*) - I(-w^* \geq c)] dw^* &= -c - \int_{w_L}^{w_U} [1 - H(-w^*) - I(w^* > -c)] dw^* \\
&= -c - \int_{w_L}^{w_U} [F_{w^*}(w^*) - I(w^* > -c)] \frac{\partial w^*}{\partial w^*} dw^* \\
&= -c - (w^* [F_{w^*}(w^*) - I(w^* > -c)]|_{w_L}^{w_U}) + \int_{w_L}^{w_U} w^* \frac{\partial [F_{w^*}(w^*) - I(w^* > -c)]}{\partial w^*} dw^*
\end{aligned}$$

To evaluate this expression, first consider the boundary related terms, noting that $F_{w^*}(w_L) = 0$ and $F_{w^*}(w_U) = 1$.

$$\begin{aligned}
-c - (w^* [F_{w^*}(w^*) - I(w^* > -c)]|_{w_L}^{w_U}) &= -c - (w^* F_{w^*}(w^*)|_{w_L}^{-c}) - (w^* [F_{w^*}(w^*) - 1]|_{-c}^{w_U}) \\
&= -c - (-c F_{w^*}(-c) - w_L F_{w^*}(w_L)) - (w_U [F_{w^*}(w_U) - 1] - (-c) [F_{w^*}(-c) - 1]) \\
&= -c - (-c F_{w^*}(-c)) + (-c) [F_{w^*}(-c) - 1] = c - c = 0
\end{aligned}$$

Observe that the role of the term $I(w^* > -c)$ is to simplify these boundary terms, since $F_{w^*}(w^*) - I(w^* > -c)$ equals zero for $w^* = w_L$ and for $w^* = w_U$.

To finish deriving equation (2), now consider the remaining integral part.

$$\begin{aligned} \int_{w_L}^{w_U} w^* \frac{\partial [F_{w^*}(w^*) - I(w^* > -c)]}{\partial w^*} dw^* &= \int_{w_L}^{-c} w^* \frac{\partial F_{w^*}(w^*)}{\partial w^*} dw^* + \int_{-c}^{w_U} w^* \frac{\partial [F_{w^*}(w^*) - 1]}{\partial w^*} dw^* \\ &= \int_{w_L}^{w_U} w^* \frac{\partial F_{w^*}(w^*)}{\partial w^*} dw^* = \int_{w_L}^{w_U} w^* dF_{w^*}(w^*) = E(W^*) \end{aligned}$$

where the last equality just used the definition of an expectation. We have now shown that equation (4) holds.

Recall that H is defined by $H(v) = E(D | V = v)$. It would be simpler to base an estimator on equation (4) rather than equation (1), because now one would only need to replace H with a nonparametric regression estimator \hat{H} in an integral without taking a derivative of \hat{H} . However, a further simplification is possible, by showing that equation (3) holds, as follows. Start from equation (4), making the change of variables $v = -w^*$ in the integral.

$$\begin{aligned} E(W^*) &= -c + \int_{w_L}^{w_U} [H(-w^*) - I(-w^* \geq c)] dw^* \\ &= -c + \int_{-w_U}^{-w_L} [H(v) - I(v \geq c)] dv \end{aligned}$$

Let Ω_v be the support of V . Then

$$\begin{aligned} E(W^*) &= -c + \int_{-w_U}^{-w_L} [H(v) - I(v \geq c)] dv = -c + \int_{-w_U}^{-w_L} \left[\frac{H(v) - I(v \geq c)}{f_v(v)} \right] f_v(v) dv \\ &= -c + \int_{-w_U}^{-w_L} \left[\frac{H(v) - I(v \geq c)}{f_v(v)} \right] f_v(v) dv \\ &= -c + \int_{-w_U}^{-w_L} \left[\frac{E(D | V = v) - I(v \geq c)}{f_v(v)} \right] f_v(v) dv \\ &= -c + \int_{-w_U}^{-w_L} \left[E \left(\frac{D - I(V \geq c)}{f_v(V)} \mid V = v \right) \right] f_v(v) dv \\ &= -c + \int_{v \in \Omega_v} \left[E \left(\frac{D - I(V \geq c)}{f_v(V)} \mid V = v \right) \right] f_v(v) dv. \end{aligned}$$

The last equality above used the assumption that V can take on any value that $-W^*$ can take on, so the support of $-W^*$, which is the interval from $-w_U$ to $-w_L$, lies inside Ω_v , and also that $D - I(V \geq c) = I(V + W^* \geq 0) - I(V \geq c)$, which equals zero for any value of V that lies outside the interval $-w_U$ to $-w_L$.

Now substitute in the definition of T_1 to get

$$\begin{aligned} E(W^*) &= -c + \int_{v \in \Omega_v} E(T_1 + c \mid V = v) f_v(v) dv = -c + E[E(T_1 + c \mid V)] \\ &= E[E(T_1 \mid V)] = E(T_1) \end{aligned}$$

where we have use the definition of expectation over V , and applied the law of iterated expectations. This completes the derivation of equation (3).

4 An Alternative Derivation

Let $c = 0$ for simplicity. The previous section proved $E(T_1) = E(W^*)$ by applying an integration by parts argument to get that $E(W^*) = \int [H(-w^*) - I(-w^* \geq 0)] dw^*$, or equivalently $E(W^*) = \int [H(v) - I(v \geq 0)] dv$. Multiplying and dividing the expression in this integral by $f_v(v)$ and using $H(v) = E(D \mid V = v)$ then shows that this integral is the same as $E(T_1)$.

The advantage of the derivation in the previous section is that it directly shows how T_1 is obtained from the definition of the mean of a random variable W^* , and hence how it directly follows from identification of H .

A more direct, but perhaps less insightful derivation of the result follows from the proof in Lewbel (2000). Starting from $E(T_1)$ with $c = 0$, follow the steps of the previous section in reverse order to get

$$\begin{aligned} E(T_1) &= \int_{-w_U}^{-w_L} \left[E \left(\frac{D - I(V \geq 0)}{f_v(V)} \mid V = v \right) \right] f_v(v) dv \\ &= \int_{-w_U}^{-w_L} E[D - I(V \geq 0) \mid V = v] dv \\ &= \int_{-w_U}^{-w_L} E[I(V + W^* \geq 0) - I(V \geq 0) \mid V = v] dv \\ &= \int_{-w_U}^{-w_L} E[I(V + W^* \geq 0) - I(V \geq 0) \mid V = v] dv \\ &= E \left[\int_{-w_U}^{-w_L} [I(v + W^* \geq 0) - I(v \geq 0)] dv \right]. \end{aligned}$$

where this last step used the independence of V and W^* to pass the integral into the expectation. Equivalently, the expectation is just over W^* , so we are just changing the order of integration over v and over

W^* . When W^* is positive this expression becomes

$$\begin{aligned} E(T_1) &= E\left(\int_{-w_U}^{-w_L} I(0 \leq v \leq W^*) dv\right) \\ &= E\left(\int_0^{W^*} 1 dv\right) = E(W^*) \end{aligned}$$

and an analogous result holds when W^* is negative.

An advantage of this derivation is that it does not entail direct consideration of the boundary terms associated with the integration by parts. The technical assumptions required for dealing with integration by parts and possibly infinite boundaries w_L or w_U are replaced with the assumption that we can change the order of integration, e.g., Fubini's theorem.

5 Estimating Unconditional Moments

Given that $E(T_1) = E(W^*)$ we can construct the variable $T_{1i} = [D_i - I(V_i \geq 0)]/f_v(V_i)$ for each individual i in a sample, and let the estimate of $E(W^*)$ be the sample average $\bar{T}_1 = \sum_{i=1}^n T_{1i}/n$. Here we have made the simplest choice for the constant c , letting $c = 0$. In applications where the density of V is not already known to the researcher by experimental design, \bar{T} can still be used to estimate $E(W^*)$ by replacing $f_v(V_i)$ in T_{1i} with a uniformly consistent density estimator $\hat{f}_v(V_i)$. For example an ordinary Rosenblatt-Parzen kernel density estimator could be used, or the simpler sorted data estimator described by Lewbel and Schennach (2007).

Although \bar{T}_1 is a sample average, it is possible that this estimator will not converge at rate root n . This is because the density $f_v(V_i)$ may have thin tails, and we are dividing by this density, which means that the distribution of T_1 can have tails that are too thick to satisfy the Lindeberg condition for the central limit theorem. It can be shown that obtaining parametric rates requires either finite support for W^* and V , or that V have infinite variance, or that W^* satisfy a tail symmetry condition as defined by Magnac and Maurin (2007). See Khan and Tamer (2010), Khan and Nekipelov (2010a, 2010b), and Dong and Lewbel (2012) for more discussion of this point, and see Lewbel and Schennach (2007), Jacho-Chávez (2009), and Chu and Jacho-Chávez (2012) for more general limiting distribution theory regarding averages with a

density in the denominator.

Based on Lewbel, Linton, and McFadden (2011), equation (3) readily extends to estimating other moments of W^* , using the fact that $D = I(V + W^* \geq 0) = I[-h(-V) + h(W^*) \geq 0]$ for any strictly monotonically increasing function h . Therefore, if we let $\tilde{V} = -h(-V)$ and $h'(v) = \partial h(v) / \partial v$ we have

$$E[h(W^*)] = E\left[c + \frac{D - I(\tilde{V} \geq c)}{f_{\tilde{V}}(\tilde{V})}\right] = E\left(c + \frac{[D - I([h(-V) \leq -c])] h'(-V)}{f_v(V)}\right).$$

It follows that, given a function h , to estimate $E[h(W^*)]$ we can construct

$$T_{hi} = c + \frac{[D_i - I([h(-V_i) \leq -c])] h'(-V_i)}{f_v(V_i)}$$

for each observation i , and then take the sample average $\bar{T}_h = \sum_{i=1}^n T_{hi} / n$ as an estimator for $E[h(W^*)]$.

For example, letting $h(W^*) = (W^*)^\lambda$ for integers λ provides direct estimators for second and higher moments of W^* , if W^* is everywhere nonnegative or nonpositive as in the willingness to pay example.

6 Identification With Covariates

Now consider how the previous section's results can be generalized by the inclusion of additional covariates. Continue with the willingness to pay application for now, so we still have $D = I(V + W^* \geq 0)$. Let X be a vector of observed covariates consisting of attributes of the sampled individuals, such as their age and income, which might affect or otherwise correlate with their willingness to pay. The distribution that V is drawn from could depend on X (e.g., quoting higher prices to richer individuals) but by construction it will be the case that $W^* \perp V \mid X$. The symbols \perp and \mid denote statistical independence and conditioning respectively, so $W^* \perp V \mid X$ says that the conditional distribution of W^* given X is independent of the conditional distribution of V given X .

Let $H(v, x) = E(D \mid V = v, X = x)$ so now H could be estimated as a nonparametric regression of D on both X and V . It follows from $W^* \perp V \mid X$ that

$$H(v, x) = 1 - \Pr[W^* \geq -v \mid X = x] = 1 - F_{W^* \mid X}(-v \mid x) \quad (5)$$

where $F_{W^*|X}$ denotes the conditional distribution of W^* given X . Therefore

$$F_{W^*|X}(w | x) = 1 - E(D | V = -w, X = x), \quad (6)$$

so one minus a nonparametric regression of D on both X and V , evaluated at $V = -w$, provides a consistent estimator of the conditional distribution of W^* given X , that is, the distribution of willingness to pay conditional on observed attributes of individuals like their age or income.

An analogous calculation to that of the previous sections can be applied to calculate the conditional mean willingness to pay $E(W^* | X)$. All that is required is to replace every expectation, density, and distribution function in the previous sections with conditional expectations, densities, or distributions, conditioning on X . In place of equation (2), define T_2 by

$$T_2 = \frac{D - I(V \geq 0)}{f_{v|x}(V | X)} \quad (7)$$

where $f_{v|x}$ denotes the conditional probability density function of V given X (for simplicity, we will hereafter let $c = 0$). As before with the pdf f_v in T_1 , when constructing T_2 the pdf $f_{v|x}$ is either known by experimental design or can be estimated using a Rosenblatt-Parzen kernel density estimator. Repeating the derivations of the previous sections, but now conditioning on $X = x$, shows that $E(W^* | X) = E(T_2 | X)$. Therefore the conditional mean willingness to pay $E(W^* | X)$ can be consistently estimated by a nonparametric regression (e.g., a kernel or local linear regression) of T_2 on X , where each observation of T_2 is defined by $T_{2i} = [D_i - I(V_i \geq 0)] / f_{v|x}(V_i | X_i)$ for each individual i in the sample.

The fact that the entire conditional distribution of W^* given X is identified means that any model for W^* that would have been identified if W^* were observable will now be identified via the special regressor even though W^* is latent. The next sections give some examples.

7 Latent Linear Index Models

Consider the standard consumer demand threshold crossing model $D = I(V + X'\beta + \varepsilon \geq 0)$ where D indicates whether a consumer buys a product or not, and $V + X'\beta + \varepsilon$ is the individual's utility from

purchasing the good. This corresponds to the special case of the model in the previous section in which the model $W^* = X'\beta + \varepsilon$ is imposed. The vector X can (and typically would) include a constant term. If V is the negative of the price, then normalizing the coefficient of V to equal one puts the utility in a money metric form, that is, $V + X'\beta + \varepsilon$ is then the consumer's surplus, defined as their reservation price W^* minus their cost of purchasing, which is $-V$. The willingness to pay model is an example where the product is a public good like preserving a wetland and price is determined by experimental design. In the more general demand context, we may not want to take price to be the special regressor since, e.g., real world prices may be endogenous, or not vary sufficiently to give V the necessary large support. But whatever variable that we take to be V can have its coefficient normalized to equal one (by changing the sign of V if necessary).

We will not require ε to be independent of X . So, e.g., the random coefficients model $D = I(V + X'e \geq 0)$ would be permitted, defining $\beta = E(e)$ and $\varepsilon = X'(e - \beta)$. More generally the variance and higher moments of ε can depend on X in arbitrary ways, allowing for any, unknown form of heteroscedasticity with respect to X .

This type of model would typically be estimated by parameterizing ε and doing maximum likelihood, e.g., this would be a probit model if ε were normal and independent of X . However, the special regressor V allows this model to be identified and estimated even if the distribution of ε is unknown, and even if the second and higher moments of that distribution depends on X in unknown ways (such as having heteroscedasticity of unknown form).

If all we know about the latent error ε in the threshold crossing model $D = I(V + X'\beta + \varepsilon \geq 0)$ is that $E(X\varepsilon) = 0$, then it can be shown that β would not be identified. However, here we also have the special regressor conditional independence assumption that $W^* \perp V \mid X$ which with $W^* = X'\beta + \varepsilon$ implies that $\varepsilon \perp V \mid X$. This condition along with $E(X\varepsilon) = 0$ suffices to identify the entire model, as follows.

First note that in the usual linear regression way, having $W^* = X'\beta + \varepsilon$ and $E(X\varepsilon) = 0$ means that

$\beta = E (XX')^{-1} E (XW^*)$, assuming that $E (XX')$ is nonsingular. We do not observe W^* , however,

$$E (XW^*) = E (XE (W^* | X)) = E \left(X \int w^* f_{w^*|X} (w^* | x) dw^* \right) = E \left(X \int w^* \frac{dF_{w^*|X} (w^* | x)}{dw^*} dw^* \right)$$

where the integral is over the support of W^* . So, to identify β , recall that $H (v, x) = E (D | V = v, X = x)$ is identified (and could be estimated by an ordinary nonparametric regression). Then $W^* \perp V | X$ implies that $1 - H (v, x) = F_{W^*|X} (-v | x)$ as before. Plugging this into the above integral shows that

$$\beta = E (XX')^{-1} E (XW^*) = E (XX')^{-1} E \left(X \int v \frac{d[1 - H (-v, x)]}{dv} dv \right)$$

and therefore that β is identified, because all of the terms on the right are identified.

Following the logic of the previous sections, a much simpler estimator for β can be constructed using T_2 . In particular, by the derivation of equations (6) and (7) we have

$$\beta = E (XX')^{-1} E (XW^*) = E (XX')^{-1} E (XT_2),$$

so β is given by a linear ordinary least squares regression of T_2 on X , where T_2 is defined by equation (7). This is one of the estimators proposed in Lewbel (2000).

To implement this estimator we first construct $T_{2i} = [D_i - I (V_i \geq 0)] / f_{v|x} (V_i | X_i)$ for each observation i in the data, and then construct the usual ordinary least squares regression estimator $\hat{\beta} = (\sum_{i=1}^n X_i X_i')^{-1} (\sum_{i=1}^n X_i T_{2i})$. As before, if the density $f_{v|x} (V_i | X_i)$ is unknown, it can be estimated using, e.g., a kernel density estimator.

8 Latent Random Coefficients

A number of authors have considered binary choice models with random coefficients, including Berry, Levinsohn, and Pakes (1995), Ichimura and Thompson (1998), Gautier and Kitamura (2009), and Hoderlein (2009). Suppose the latent $W^* = X'e$, where e is a vector of random coefficients. Then we have the random coefficients binary choice model $D = I (V + X'e \geq 0)$, where the scale normalization is imposed by setting the coefficient of V equal to one. It then follows immediately from equation (6) that

$F_{X'e|X}(w|x) = 1 - E(D|V = -w, X = x)$. The conditional distribution $F_{X'e|X}$ is the same information that one could identify from a linear random coefficients model (i.e., a model where one observed W and X with $W = X'e$), so the known nonparametric identification of linear random coefficients models (see, e.g., Hoderlein, Klemelae, and Mammen (2010) and references therein) can be immediately applied to show nonparametric identification of binary choice random coefficients, by means of the special regressor.

9 Latent Partly Linear Regression

Suppose that $W^* = g(X) + \varepsilon$ where $E(\varepsilon|X) = 0$ for some unknown function g of the vector X . This corresponds to the partly linear latent variable regression model $D = I(V + g(X) + \varepsilon \geq 0)$. In this model $g(X) = E(T_2|X)$, which could therefore be estimated by an ordinary nonparametric regression of the constructed variable T_2 on X . In this model, the required conditional independence assumption that $W^* \perp V|X$ will hold if $\varepsilon \perp V|X$, in addition to $E(\varepsilon|X) = 0$. More simply, though stronger than necessary, it could just be assumed that $\varepsilon \perp (V, X)$ to make all the required independence assumptions regarding V , X , and ε hold. The only other requirement would then be that the support of $g(X) + \varepsilon$ equals, or is contained in, the support of $-V$ (or that the tail symmetry condition of Magnac and Maurin 2007 holds).

These results immediately extend to the general partly latent variable model $I(V + X_1'B + g_2(X_2) + \varepsilon \geq 0)$ where X_1 and X_2 are subvectors comprising X . In this model we would have $E(T_2|X) = X_1'B + g_2(X_2)$, so the vector B and the function g_2 could be estimated by applying Robinson (1988), using T_2 as the dependent variable.

10 Latent Nonparametric Instrumental Variables

In the previous section we had the model $D = I(V + g(X) + \varepsilon \geq 0)$, where the unobserved W^* satisfied $W^* = g(X) + \varepsilon$ with $E(\varepsilon|X) = 0$. Suppose now we have the same model, except instead of

$E(\varepsilon | X) = 0$ we assume $E(\varepsilon | Z) = 0$ where Z is a vector of observed instruments. Some elements of X , corresponding to exogenous covariates, may also be in Z . Other elements of X are endogenous, in that they could be correlated with ε .

If W^* were observed, then $g(X)$ would need to be identified and estimated from the conditional moments

$$E[W^* - g(X) | Z] = 0. \quad (8)$$

If W^* were observed this would be the instrumental variables nonparametric regression model of, e.g., Newey and Powell (2003), Hall and Horowitz (2005), Darolles, Fan, Florens, and Renault, (2011) and Chen and Reiss (2011). Assume, as those authors do, that $g(X)$ is identified from the conditional moments in equation (8), and consider how we might construct sample moment analogues to this equation in the case where W^* is not observed.

Replace the definition of T_2 in equation (7) with T_3 defined by

$$T_3 = \frac{D - I(V \geq 0)}{f_{v|Z}(V | Z)}, \quad (9)$$

where $f_{v|Z}$ denotes the conditional probability density function of V given Z . Let $\Omega_{v|z}$ be the support of $V | Z$. Then, by replacing X with Z everywhere in section 6 above, we have that $E(W^* | Z) = E(T_3 | Z)$, assuming that $W^* \perp V | Z$, and assuming that $V | Z$ has sufficiently large support. It follows immediately that $E[T_3 - g(X) | Z] = 0$, so after constructing T_3 , estimators like those of Newey and Powell (2003) or the other authors listed above could be directly applied to estimate $g(X)$ by instrumental variables nonparametric regression.

In a later section we will discuss implications of the conditional independence assumption $W^* \perp V | Z$, and how this restriction can be relaxed.

11 Latent Linear Index Models with Endogenous or Mismeasured Regressors

Return to the linear latent variable model $D = I(V + X'\beta + \varepsilon \geq 0)$. The earlier estimator for β in this model assumed that $V | X$ is continuous with large support, that $\varepsilon \perp V | X$, that $E(X\varepsilon) = 0$, and that $E(XX')$ is nonsingular. Apart from restrictions on V , the only assumptions regarding X and ε were the same as the assumptions required for linear ordinary least squares regression, that is, $E(XX')$ nonsingular and $E(X\varepsilon) = 0$.

We now extend this model to allow for endogeneous or mismeasured regressors, by replacing the linear ordinary least squares estimator with a linear two stage least squares estimator. Let Z be a vector of observed instruments. The vector Z is assumed to include any elements of X that are exogenous (including a constant term), but does not include V . Any elements of X that are endogenous or mismeasured are not included in Z .

Suppose that we still have the model $D = I(V + X'\beta + \varepsilon \geq 0)$, but where some or all of the elements of X are endogenous or mismeasured and so may be correlated with ε . First make the exact same assumptions that would be required for linear model two stage least squares with instruments Z . These are $E(Z\varepsilon) = 0$, the rank of $E(ZX')$ equals the dimension of X , and $E(ZZ')$ is nonsingular. Now add special regressor assumptions regarding V , by assuming that $V | Z$ is continuous with large support and $\varepsilon \perp V | Z$. Notice that the special regressor assumptions involve V , Z , and ε , but not X . The only assumptions regarding the endogenous regressors in X that are required are the same as the minimal assumptions needed for linear model two stage least squares regression.

Letting $W^* = X'\beta + \varepsilon$, it follows immediately from the results in the previous section that if we define T_3 by equation (9) then $E(T_3 | Z) = E(X'\beta + \varepsilon | Z)$ and therefore $E(ZT_3) = E(ZX')\beta$, so

$$\beta = \left[E(XZ') E(ZZ')^{-1} E(ZX') \right]^{-1} E(XZ') E(ZZ')^{-1} E \left(Z \frac{D - I(V \geq 0)}{f_{v|Z}(V | Z)} \right)$$

which is identical to a linear two stage least squares regression of T_3 on X , using instruments Z . More generally, we could estimate β by applying standard GMM estimation to the moment conditions $E[Z(T_3 - X'\beta)] =$

0. This GMM could be more efficient than two stage least squares if errors are heteroskedastic.

A particularly useful feature of this construction is that, apart from restrictions involving V , all that is required regarding the endogenous regressors and instruments is identical to what is required for linear two stage least squares. In particular, the type of restrictions required for control function or maximum likelihood based estimation are not needed. This is particularly useful for cases where some of the endogenous regressors are themselves discrete or limited. See Lewbel, Dong, and Yang (2012) for details.

Finally, it should be noted that only one special regressor is required, regardless of how many other regressors are endogenous. If more than one regressor satisfies the assumptions required to be special, then, based on experiments in Lewbel (2000), the one with the largest spread (e.g., largest variance or interquartile range) should be chosen to minimize finite sample bias.

12 Relaxing The Conditional Independence Assumption

For instrumental variables estimation, in the previous two sections it was assumed that

$$W^* | V, Z = W^* | Z \tag{10}$$

to obtain either $E(W^* | Z) = E(T_3 | Z)$ or, for the linear index model $E(ZW^*) = E(ZT_3)$. Then, given $W^* = g(X) + \varepsilon$, assuming either $E(\varepsilon | Z) = 0$ or just $E(Z\varepsilon) = 0$, the model could be estimated by nonparametric instrumental variables, or in the linear index case where $g(X) = X'\beta$, by ordinary linear two stage least squares, treating T_3 as the dependent variable. All that was required for these estimators to work is that equation (10) hold, that V have sufficiently large support, and that Z have the standard properties of instruments for either nonparametric instrumental variables or for linear two stage least squares.

Since $W^* = g(X) + \varepsilon$, a sufficient but stronger than necessary condition for equation (10) to hold is that

$$X, \varepsilon | V, Z = X, \varepsilon | Z, \tag{11}$$

meaning that V is exogenous in a standard sense of being conditionally independent of the latent model

error ε , but in addition $X \perp V, Z = X \perp Z$, meaning that V would drop out of a model of endogenous regressors as a function of V and Z . This is a strong restriction on the special regressor V relative to the endogenous regressors, but fortunately it is stronger than necessary.

One way to relax equation (10) is to replace T_3 in equation (9) with T_4 defined by

$$T_4 = \frac{D - I(V \geq 0)}{f_{v|S}(V | S)} \quad (12)$$

where S is a vector containing the union of all the elements of Z and of X . We then have that

$$E(W^* | S) = E(T_4 | S) \quad (13)$$

holds, assuming $W^* \perp V | S$ and that $V | S$ has sufficiently large support. It follows from applying the law of iterated expectations to equation (13) that $E(W^* | Z) = E(T_4 | Z)$, which is what we require to estimate $g(X)$ using $E[T_4 - g(X) | Z] = 0$. Similarly, by applying the law of iterated expectations to equation (13), T_4 can be used in place of T_1, T_2 , or T_3 in all of the estimators described so far.

Requiring $W^* \perp V | S$ is equivalent to $W^* \perp V | X, Z$. With $W^* = g(X) + \varepsilon$, this is in turn equivalent to

$$\varepsilon \perp V | X, Z. \quad (14)$$

Equation (14) relaxes equation (11), and in particular does not impose the condition that $X \perp V, Z = X \perp Z$. Equation (14) will hold if we can write a model $V = M(U, X, Z)$ where M is invertible in U and U is an error term that is independent of X, Z, ε . For example, define $P = E(V | X, Z)$, define $U = V - P$, and suppose that the endogenous elements of X are functions of Z, P , and an error vector e that is independent of Z, P , and U . It is not necessary to actually specify or estimate this or any model for any elements of X . With this construction, X can depend on V by depending on P , and X can be endogenous by e correlating with ε , with equation (14) holding. This construction also does not impose any control function type restrictions on W^*, X , and Z , and so for example still allows X to be discrete or limited.

13 Constructing T_{ji}

Implementing the estimators discussed in the previous sections requires constructing an estimate of T_{ji} for each observation i and for $j = 1, 2, 3$, or 4 . For each observation i , the variable T_{ji} is given by $T_{ji} = -c + [D_i - I(V_i \geq c)] / f_{v|S}(V_i | R_i)$ where R is either empty as in equation (2) where $j = 1$, or $R = X$ in equation (7) where $j = 2$, or $R = Z$ in equation (9) where $j = 3$, or $R = S$ in equation (12) where $j = 4$. The constant c can be any value inside the support of V . A natural choice for the constant c is the mean or median of V . More simply, we will just assume that V is centered (e.g., demeaned prior to the analysis), and just let $c = 0$.

Lewbel and Tang (2012) prove that the term $I(V_i \geq 0)$ in the definition of T_{ji} can be replaced with $M(V_i)$ where M is any mean zero probability distribution function that lies inside the support of V . Choosing M to be a simple differentiable function like

$$M(V) = I(-\sigma \leq V \leq \sigma) \frac{V + \sigma}{2\sigma}$$

where σ is the standard deviation of V (corresponding to a uniform distribution on $c - \sigma$ to $c + \sigma$) can simplify the calculation of limiting distributions and possibly improve the finite sample performance of the estimators. The free substitution of $M(V)$ for $I(V \geq 0)$ can be made for all special regressor estimators.

If V is determined by the researcher as in the willingness to pay examples, then the density function $f_{v|R}$ is known by experimental design. Otherwise, $f_{v|R}$ will need to be estimated. If $f_{v|R}$ is parameterized as $f_{v|R}(V | R, \theta)$ for some parameter vector θ , then efficient estimation could be accomplished by GMM, combining the moments based on the score function for maximum likelihood estimation of θ , i.e., $E[\partial f_{v|R}(V | R, \theta) / \partial \theta] = 0$ with the moments used for estimation of β , such as $E[Z(T - X'\beta)] = 0$ from the previous section.

The function $f_{v|R}$ could also be estimated nonparametrically by, e.g., a Nadayara-Watson kernel density estimator, but this may be very imprecise if the dimension of R is high. Dong and Lewbel (2012) propose some alternative semiparametric estimators for this density. For example, suppose $V = S'\gamma + U$, where S is as defined in equation (12), as the union of all the elements of X and Z . If $U \perp S$ then

$f_{v|S}(V | S) = f_u(U)$. We may then define T by

$$T = \frac{D - I(V \geq 0)}{f_u(U)} \quad (15)$$

and correspondingly construct data $T_i = [D_i - I(V_i \geq 0)] / f_u(U_i)$. By the law of iterated expectations, this T_i can be used in place T_{ji} for all the special regressor estimators. Moreover, this is a special case of the model discussed at the end of previous section, so the required conditional independence assumptions involving the special regressor V will be satisfied if U is independent of X , Z , and W^* .

The advantage of this construction is that each U_i can be estimated as the residuals from an ordinary least squares linear regression of V on S , and the density function f_u of the scalar random variable U can be estimated by a one dimensional kernel density estimator applied to the data U_1, \dots, U_n . Even more simply, the ordered data estimator of Lewbel and Schennach (2007) can be applied to U_1, \dots, U_n , which does not require any choice of kernel or bandwidth. See Dong and Lewbel (2012) for more details and for alternative simple estimators.

One final note regarding construction of T , is that $f_{v|R}$ must have a large support, and so $f_{v|R}(V_i | R_i)$ may be very close to zero for very low and very high values of V_i . Similarly, $f_u(U_i)$ may be very close to zero for large values of $|U_i|$. The corresponding values of T_i may then be extremely large in magnitude. The special regressor estimators that involve estimating either moments of T or regression models using T as the dependent variable can be very sensitive to outlier observations of T . It may therefore be prudent, based on a mean squared error criterion, to Winsorize T , or to trim out data observations i where either T_i or the residuals in the regressions of T_i on X_i take on extreme values.

14 What If the Special Regressor is Discrete?

Suppose we have a model that fits the requirements for special regressor estimation, except that the special regressor V is discretely distributed, and thereby violates the required support restrictions. In this case, it is still sometimes possible to apply the estimators discussed earlier.

One possibility is to assume that the number of values that V can take on grows with the sample size.

For example, in the earlier willingness to pay application, this would mean assuming that the larger the size of the experiment (measured in terms of the number of experimental subjects), the larger would be the number of different proposed willingness to pay values that experimentors would select from to offer to a subject. Analogous to the type of infill asymptotics that is sometimes used in the time series literature, suppose that in the limit as $n \rightarrow \infty$, the number of values V can take on grows to make the support of V become dense on a large interval (larger than the support of W^*). Then Lewbel, Linton, and McFadden (2011) show that, asymptotically, the distribution of W^* can be identified, and they supply associated limiting distribution theory for estimation based on that design.

One implication of the results in Lewbel, Linton, and McFadden (2011) is that if V is discretely distributed, but the number and range of values V can take on grows sufficiently quickly with n , then one can ignore the fact that V is discrete and do special regressor estimation as if V was continuous. For the resulting estimator to perform well in practice, the number of different values that V takes on in the observed data, and the range of those values, will need to be relatively large.³ Li and Racine (2007) give other examples of treating discrete data as if it was continuous in e.g., nonparametric kernel regressions.

Another situation in which a discrete special regressor can be used is when the true V is continuous with a uniform distribution, and what one observes is a discretized version of V . For example, consider an application like that of Maurin (2002), in which the outcome D is whether a student will be held back in school, and the special regressor V is how old the student is at the date of enrollment in school. Suppose we do not observe the student's birthdates and times, but instead only know if a student is either five or six years old when school starts. So here the observed, discretized version of the special regressor age is just the binary variable indicating whether the student is five or six. By defining c appropriately, $I(V_i \geq c)$ will be one for students who are six when school starts, and zero for students who are five. Assuming that birthdates and times are close to uniform within a year,⁴ $f_{V|R}(V_i | R_i)$ is a constant (equal to one if V is

³In Lewbel, Linton, and McFadden's (2011) empirical application, the sample size was $n = 518$ and the number of willingness to pay bid values V took on was 14, ranging in value from 25 to 375. This range between the lowest and highest value of V in the data likely covered a large portion of the actual range of willingness to pay in the population.

⁴There exists evidence of statistically significant seasonal departures from uniformity in the distribution of births within

measured in years) and so is known despite not observing V_i . Since both $f_{V|R}(V_i | R_i)$ and $I(V_i \geq c)$ are known for each student i , the variable T_{ji} can be constructed for each student i , and so the special regressor estimators can be applied. Special regressor estimation only requires observing T_{ji} , not V_i , for each observation i .

If situations like those described above do not apply, then special regressor methods can still be used when V is discrete, but as described in section 2, in that case the distribution of the latent variable W^* will only be identified at the values $-V$ can take on. This in turn only permits bounds on coefficients and moments to be identified and estimated. See Magnac and Maurin (2008) for details on partial identification with a discrete special regressor.

15 Extensions

This chapter has focused on binary choice model estimation, but the main idea of the special regressor method can be applied to a variety of models. For example, ordered choice models are also identified, including models with random thresholds and endogenous regressors. Suppose for $j = 1, \dots, J$ that Y equals the integer j when $\alpha_{j-1} + \varepsilon_{j-1} < V + g(X) \leq \alpha_j + \varepsilon_j$ for some constants α_j and errors ε_j having unknown distributions. Then let $D_j = I(Y \leq j) = I(V + g(X) - \alpha_j - \varepsilon_j)$ and apply the special regressor estimator to each D_j to identify the conditional distribution of $g(X) - \alpha_j - \varepsilon_j$ given X for each j (or given instruments Z if endogenous regressors are present). See Lewbel (2000) for details.

The special regressor method is convenient for panel data latent variable models with latent fixed effects, because if $D_{it} = I(V_{it} + W_{it}^* \geq 0)$ then we can construct T_{jit} such that $E(T_{jit} | X) = E(W_{it}^* | X)$, and so for example $E(T_{jit} - T_{jit-1} | X) = E(W_{it}^* - W_{it-1}^* | X)$, meaning that we can difference out fixed effects in the latent W_{it}^* . This construction permits, but does not require, the special regressor V_{it} to vary over time. See, e.g., Honore and Lewbel (2002), Ai and Gan (2010), and Gayle (2012). Using the special regressor in this way, it is sometimes possible to estimate panel binary choice models with fixed

a year, but the magnitude of these departures from uniformity is quite small. See, e.g., Beresford (1980). A related use of uniformity to account for discretization in observed age is Dong (2012).

effects that converge at rate root n , even when the error distribution is not only not logit, but not known at all. Here the special regressor conditional independence assumption overcomes Chamberlain's (1993) result that logit errors are required for root n convergence of panel binary choice models with fixed effects.

As briefly noted earlier, if $Y = g(V + W^*)$ for some possibly unknown, weakly monotonic function g , then the conditional distribution of W^* given a vector of covariates X can be identified (up to a location normalization) by letting $D = I(Y \geq y_0)$ for any constant y_0 in the support of Y that makes D be nonconstant, which reduces the problem to an equivalent binary choice problem. Identification is only up to a location normalization because for any constant a one could replace W^* with $W^* + a$ and redefine g accordingly.

The particular location value for W^* that is obtained by applying previous estimators to $D = I(Y \geq y_0)$ will depend on g and on the choice of y_0 . To increase the efficiency of terms other than location, one could combine estimates based on multiple choices of y_0 . This can be also be combined with the above panel data generalization to permit estimation of many nonlinear or nonparametric panel data models with latent fixed effects.

Suppose $Y = g(V_1 + W_1^*, V_2 + W_2^*)$ where V_1 and V_2 are two special regressors and W_1^* and W_2^* are two latent variables. Then generally the joint distribution of W_1^* and W_2^* given X can be identified. Lewbel (2000) provides an example, showing identification of general multinomial choice models. Similarly, Lewbel and Tang (2011), Khan and Nekipelov (2011), and Fox and Yang (2012) use multiple special regressors for identification of games, including entry games and matching games, with semiparametrically specified payoffs, while Berry and Haile (2009a, 2009b) use multiple special regressors to identify multinomial discrete choice market equilibrium models that are semiparametric generalizations of Berry, Levinsohn, and Pakes (1995).

16 Conclusions

The goal here has been to provide an understanding of how special regressor methods work and can be applied to estimate features of latent variable models. The focus was on identification and associated

construction of estimators, not on limiting distribution theory. The estimators are multistep estimators, where each step takes the form of a standard parametric or nonparametric density or regression estimator. Yet despite being comprised of standard estimators, a number of technical issues can arise. In particular, the rates of convergence of these estimators can vary depending upon the thickness of the tails of the distribution of the special regressor. In general, converging at standard parametric root n rates requires either parametric specification of the density of V , or finite support of the model errors, or very thick tails for V (thick enough for the variance to be infinite), or conditions like the tail symmetry of Magnac and Maurin (2007). When special regressor estimators do converge at rate root n , the standard methods used to derive asymptotic distributions of multistep estimators can be applied, and in such cases the basic special regressor estimator has been shown to be semiparametrically efficient. Papers discussing limiting distribution theory for special regressor based estimators include Lewbel (2000, 2007a), Magnac and Maurin (2007, 2008), Jacho-Chávez (2009), Khan and Tamer (2010), Khan and Nekipelov (2010a, 2010b), and Dong and Lewbel (2012).

Lewbel, Dong, and Yang (2012) provide a comparison of special regressor models versus maximum likelihood estimation, control function estimators and linear probability models. They conclude that the greatest weakness of special regressor methods is the extent to which they rely on strong properties of just one regressor, some of which are difficult to verify, and the resulting sensitivity of estimates to this one regressor. But the strength of special regressor methods is that they impose very mild conditions on the relationships between the remaining regressors and on the model errors. This is what makes special regressor methods particularly useful for proving identification of models that have some relatively intractable components, such as their use by Berry and Haile (2009a, 2009b) to deal with endogenous prices in semiparametric generalizations of Berry, Levinsohn, and Pakes (1995) type market models, or their use by Lewbel and Tang (2012) to identify semiparametric payoff functions in discrete games.

The restrictions required for consistency of special regressor estimators are generally quite different from what is required for other estimators. This suggests that for empirical work, they may be particularly useful as robustness checks. If both special regressor estimators and more standard methods provide

similar answers, one may have a greater degree of confidence that the findings are not due to violations of standard modeling assumptions. The simplicity of many special regressor estimators makes this an easy suggestion to follow.

17 References

Abbring, J. H. and J. J. Heckman, (2007) "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation," in: J.J. Heckman & E.E. Leamer (ed.), *Handbook of Econometrics*, edition 1, volume 6B, chapter 72 Elsevier.

Ai, C. and L. Gan, (2010) "An alternative root-n consistent estimator for panel data binary choice models" *Journal of Econometrics*, 157, 93-100

An, M.Y. (2000). "A Semiparametric Distribution for Willingness to Pay and Statistical Inference with Dichotomous Choice CV Data," *American Journal of Agricultural Economics* 82, 487-500.

Avelino, R. R. G. (2006), "Estimation of Dynamic Discrete Choice Models with Flexible Correlation in the Unobservables with an Application to Migration within Brazil," unpublished manuscript, University of Chicago.

Anton, A. A., A. Fernandez Sainz, and J. Rodriguez-Poo, (2002), "Semiparametric Estimation of a Duration Model," *Oxford Bulletin of Economics and Statistics*, 63, 517-533.

Beresford, G. C. (1980), "The Uniformity Assumption in the Birthday Problem," *Mathematics Magazine*, 53, 286-288.

Berry, S. T., and P. A. Haile (2009a), "Identification in Differentiated Products Markets Using Market Level Data," Unpublished Manuscript.

Berry, S. T., and P. A. Haile (2009b), "Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers," Unpublished Manuscript.

Berry, S., J. Levinsohn, and A. Pakes (1995) "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841-890.

Briesch, R., P. Chintagunta, and R.L. Matzkin (2010) "Nonparametric Discrete Choice Models with

Unobserved Heterogeneity," *Journal of Business and Economic Statistics*, 28, 291-307.

Chamberlain, G. (1993) "Feedback in Panel Data Models," unpublished manuscript, Department of Economics, Harvard University.

Chen, X. and M. Reiss (2011), "On Rate Optimality for Ill-Posed Inverse Problems in Econometrics," *Econometric Theory*, 27 497-521.

Cogneau, D. and E. Maurin (2002), "Parental Income and School Attendance in a Low-Income Country: A Semiparametric Analysis," Unpublished Manuscript.

Chu, B. and D. T. Jacho-Chávez, (2012), "k-Nearest Neighbor Estimation of Inverse-Density-Weighted Expectations with Dependent Data," *Econometric Theory*, 28, 769-803.

Darolles, S., Fan, Y., Florens, J. P. and Renault, E. (2011), "Nonparametric Instrumental Regression," *Econometrica*, 79, 1541–1565.

Dong, Y. (2012), "Regression Discontinuity Applications with Rounding Errors in the Running Variable," University of California Irvine unpublished manuscript.

Dong, Y. and A. Lewbel (2012), "A Simple Estimator for Binary Choice Models with Endogenous Regressors," *Econometrics Reviews*, forthcoming.

Fox, J. and C. Yang (2012), "Unobserved Heterogeneity in Matching Games," unpublished manuscript.

Gautier, E., and Y. Kitamura (2009) "Nonparametric Estimation in Random Coefficients Binary Choice Models," unpublished manuscript.

Gayle, W.-R. (2012) "Identification and \sqrt{N} -Consistent Estimation of a Nonlinear Panel Data Model with Correlated Unobserved Effects," unpublished manuscript, U. of Virginia.

Goux, D. and E. Maurin (2005), "The effect of overcrowded housing on children's performance at school, *Journal of Public Economics*, 89, 797-819.

Hall, P. and J. L. Horowitz (2005), "Nonparametric methods for inference in the presence of instrumental variables," *Annals of Statistics* 33, 2904–2929.

Heckman, J. J. and Navarro, S. (2007), "Dynamic discrete choice and dynamic treatment effects," *Journal of Econometrics*, 136, 341-396.

Hoderlein, S. (2009), "Endogeneity in Semiparametric Binary Random Coefficient Models," unpublished manuscript.

Hoderlein, S., Klemelae, J. and Mammen, E. (2010), "Analyzing the Random Coefficient Model Non-parametrically," *Econometric Theory*, 26, 804-837.

Horowitz, J. L. (1992), "A Smoothed Maximum Score Estimator for the Binary Response Model," *Econometrica*, 60, 505-532.

Honore, B. and A. Lewbel, (2002) "Semiparametric Binary Choice Panel Data Models Without Strictly Exogenous Regressors," *Econometrica*, 70, 2053-2063.

Ichimura, H., and T. S. Thompson (1998): "Maximum Likelihood Estimation of a Binary Choice Model with Random Coefficients of Unknown Distribution," *Journal of Econometrics*, 86, 269–295.

Jacho-Chávez, D. T., (2009), "Efficiency Bounds For Semiparametric Estimation Of Inverse Conditional-Density-Weighted Functions," *Econometric Theory*, 25, 847-855.

Khan, S. and A. Lewbel (2007) "Weighted and Two Stage Least Squares Estimation of Semiparametric Truncated Regression Models," *Econometric Theory*, 23, 309-347.

Khan, S. and E. Tamer (2010), "Irregular Identification, Support Conditions, and Inverse Weight Estimation," *Econometrica*, 78, 2021–2042.

Khan, S. and D. Nekipelov (2010a), "Semiparametric Efficiency in Irregularly Identified Models," unpublished working paper.

Khan, S. and D. Nekipelov (2010b), "Information Bounds for Discrete Triangular Systems," unpublished working paper.

Khan, S. and D. Nekipelov (2011), "Information Structure and Statistical Information in Discrete Response Models," unpublished working paper.

Lewbel, A. (1997), "Semiparametric Estimation of Location and Other Discrete Choice Moments," *Econometric Theory*, 13, 32-51.

Lewbel, A. (1998), "Semiparametric Latent Variable Model Estimation With Endogenous or Mismeasured Regressors," *Econometrica*, 66, 105–121.

Lewbel, A. (2000), "Semiparametric Qualitative Response Model Estimation With Unknown Heteroscedasticity or Instrumental Variables," *Journal of Econometrics*, 97, 145-177.

Lewbel, A. (2007a), "Endogenous Selection or Treatment Model Estimation," *Journal of Econometrics*, 141, 777-806.

Lewbel, A. (2007b), "Modeling Heterogeneity," in *Advances in Economics and Econometrics: Theory and Applications*, Ninth World Congress (Econometric Society Monographs), Richard Blundell, Whitney K. Newey, and Torsten Persson, editors, Cambridge: Cambridge University Press, Vol. III, Chapter 5, 111-121.

Lewbel, A., Dong, Y., and T. Yang (2012), "Comparing Features of Convenient Estimators for Binary Choice Models With Endogenous Regressors," forthcoming, *Canadian Journal of Economics*.

Lewbel, A. and S. Schennach (2007), "A Simple Ordered Data Estimator for Inverse Density Weighted Functions," *Journal of Econometrics*, 136, 189-211.

Lewbel, A. and X. Tang (2011), "Identification and Estimation of Games with Incomplete Information using Excluded Regressors," unpublished manuscript.

Lewbel, A., O. Linton, and D. McFadden (2011), "Estimating Features of a Distribution From Binomial Data," *Journal of Econometrics*, 162, 170-188.

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

Magnac, T. and E. Maurin (2007), "Identification and Information in Monotone Binary Models," *Journal of Econometrics*, 139, 76-104.

Magnac, T. and E. Maurin (2008), "Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data," *Review of Economic Studies*, 75, 835-864.

Manski, C. F. (1985), "Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator," *Journal of Econometrics*, 27, 313-333.

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

Matzkin, R.L. (1994) "Restrictions of Economic Theory in Nonparametric Methods," in *Handbook of Econometrics*, Vol. IV, R.F. Engel and D.L. McFadden, eds, Amsterdam: Elsevier, Ch. 42, 2524-2554.

Matzkin, R. (2007), "Heterogeneous Choice," in *Advances in Economics and Econometrics: Theory and Applications*, Ninth World Congress (Econometric Society Monographs), Richard Blundell, Whitney K. Newey, and Torsten Persson, editors, Cambridge: Cambridge University Press, Vol. III, Chapter 4, 75-110.

Maurin, E. (2002), "The impact of parental income on early schooling transitions A re-examination using data over three generations," *Journal of Public Economics*, 85, 301–332.

Newey, W. K. and Powell, J. L. (2003), "Instrumental Variable Estimation of Nonparametric Models. *Econometrica*," 71, 1565–1578

Pistolesi, N. (2006), "The performance at school of young Americans, with individual and family endowments," unpublished manuscript.

Stewart, M. B. (2005), "A comparison of semiparametric estimators for the ordered response model," *Computational Statistics and Data Analysis*, 49, 555-573.

Tiwari, A. K., P. Mohnen, F. C. Palm, S. S. van der Loeff, (2007), "Financial Constraint and R&D Investment: Evidence from CIS," United Nations University, Maastricht Economic and social Research and training centre on Innovation and Technology (UNU-MERIT) Working Paper 011.

Vytlacil, E. and N. Yildiz (2007), "Dummy Endogenous Variables in Weakly Separable Models," *Econometrica*, 75, 757-779.