Identification of a Heterogeneous Generalized Regression Model with Group Effects

Steven T. Berry
Yale University
Department of Economics
Cowles Foundation
and NBER

Philip A. Haile
Yale University
Department of Economics
Cowles Foundation
and NBER

September 25, 2009
Preliminary

Abstract

We consider identification in a "generalized regression model" (Han, 1987) for panel settings in which each observation can be associated with a "group" whose members are subject to a common unobserved shock. Common examples of groups include markets, schools or cities. The model is fully nonparametric and allows for the endogeneity of group-specific observables, which might include prices, policies, and/or treatments. The model features heterogeneous responses to observables and unobservables, and arbitrary heteroskedasticity. We provide sufficient conditions for full identification of the model, as well as weaker conditions sufficient for identification of the the latent group effects and the distribution of outcomes conditional on covariates and the group effect.

Keywords: nonparametric identification, binary choice, threshold crossing, censored regression, proportional hazard model

1 Introduction

In the generalized regression model of Han (1987), an outcome Y_i for "individual" i is determined by the nonseparable nonlinear model

$$\begin{aligned} Y_i &= D\left(Y_i^*\right) \\ Y_i^* &= F\left(X_i\beta, E_i\right) \end{aligned} \tag{1}$$

where D is a known nondegenerate weakly increasing function, X_i is a vector of covariates associated with individual i, $E_i \in \mathbb{R}$ is an unobserved shock, and the index Y_i^* is latent. This formulation nests several common models of limited dependent variables, each arising from a particular choice of D. Examples include models of binary choice or threshold crossing, censored regression, and the proportional hazard model.

We consider the identifiability of a nonparametric heterogeneous generalization of this model for "large-N large-T" panel settings in which each individual i is associated with a "group" t.[1] We are motivated by a variety of applications in which group members are subject to a common unobserved shock that may be correlated with group-level observables affecting outcomes. Examples of "groups" include schools, hospitals, firms, neighborhoods, retailers, and markets. In applications, one is often interested in counterfactuals involving exogenous changes in covariates that affect individual outcomes. We propose a model that generalizes (1) in a way that makes it possible to define such counterfactuals while still allowing for rich heterogeneity in individual responses to observables and unobservables. We then consider identification of this model for environments in which one observes outcomes for many individuals from each of many groups.

Our model takes the form

[1] It would be more precise to let t_i represent the group to which i belongs. We use the simpler notation but emphasize that we do not assume an environment in which an individual appears in different groups.

$$
\begin{aligned}
Y_{it} &= D\left(Y_{it}^{*}\right) \\
Y_{it}^{*} &= F_i\left(S_i, X_t, U_t\left(S_i\right)\right).
\end{aligned}
\tag{2}
$$

Here S_i and X_t are, respectively, individual- and group-specific covariates, $U_t(S_i)$ is an unobserved scalar unobservable, and F_i is a random function on $\chi \equiv \operatorname{supp}(S_i, X_t, U_t(S_i))$. We will refer to $U_t(S_i)$ as a "group effect" even though it is permitted to vary with both S_i and t.[2] The nonparametric random function F_i, defined formally below, generalizes traditional random coefficients models to allow rich heterogeneity across individuals.

Two examples illustrate the types of applications that motivate our study.

Example 1 (Binary Choice Demand). *Each consumer i chooses between two substitute goods.*[3] *A consumer falls into a market t, defined by time or geography. The covariates X_t describe observable characteristics of the goods or markets; S_i denotes observed consumer characteristics; $U_t(s)$ represents unobserved taste among consumers in market t with characteristics $S_i = s$, and/or the value "type" s consumers place on unobserved characteristics of the goods in market t. The latent index Y_{it}^{*} represents the difference between the consumer's conditional indirect random utilities of the two goods; i.e.,*

$$F_i\left(S_i, X_t, U_t\left(S_i\right)\right) = v_{i1t}\left(S_i, X_t, U_t\left(S_i\right)\right) - v_{i0t}\left(S_i, X_t, U_t\left(S_i\right)\right)$$

where v_{ijt} is a random function representing the conditional indirect utility of consumer i

[2] The "group effect" in this model has characteristics of both fixed effects and random effects, although the literature does not always agree on definitions. We avoid this terminology altogether.

[3] Obviously this is only one example of binary choice. Others that may be suited to our setup include a 1) firms' decisions regarding entry or technology adoption, with market-specific unobservables and unobserved heterogeneity in firms' profit functions; 2) a student's decision to drop out or stay in school, with school-specific unobservables and heterogeneity in students' returns to education, labor market productivity, etc; 3) voters' choices of presidential candidate, with heterogeneous preferences and local unobservables; 4) workforce participation decisions, with region-specific unobserved labor market conditions and individual heterogeneity in labor market opportunities and labor-leisure tradeoff; 5) consumer purchase decisions on repeated shopping trips, with persistent consumer-specific effects.

for good j in market t. The transformation D maps the sign of this utility difference into a zero-one purchase decision. At least one of the X_t characteristics, price, will typically be correlated with the market demand shock, $U_t(S_i)$. A possible instrument for price would be a market-level cost shifter.

Example 2 (Patient Outcomes). *Consider a model of health outcomes of patients i treated in hospitals t. The latent Y_i^* might represent a continuos measure of actual health status which is transformed into the observed outcome Y_i, which could be a duration (e.g., length of stay or days survived), a censored variable such as a binned measure of post-treatment function, or a binary indicator for survival vs. death. The vector S_i can reflect characteristics of the patient, her diagnosis and/or the treatment she receives. The vector X_t captures observed characteristics of hospitals: treatment protocols, staffing ratios, for-profit status and so forth. The unobserved $U_t(S_i)$ may reflect unobserved characteristics of hospital t and/or unobserved characteristics of patients in hospital t. Although we restrict $U_t(s)$ to be a scalar, we allow it to vary freely with s and t. This permits responses to hospital unobservables (and their interactions with hospital observables) to vary with patient observables. For example, an unobserved hospital characteristic that is helpful for patients with one S_i vector (e.g., age/diagnosis/co-morbidity profile) may be harmful to other patients. In addition, the random variation across patients in the fuction F_i allows for heterogeneous responses, even among observably identical patients.*

Although we use the terms "individual" and "group" to interpret the indices i and t , other interpretations are possible. For example, t could indicate the person whose outcomes are recorded as a large number of observations i. In that case, what we call a "group effect" would be a person-specific effect.

Not all applications will have the structure we consider, but many do. For such settings, our model (2) generalizes (1) in several ways.[4] First, it incorporates the latent group effect

[4]One may wonder whether there is redundancy here, since S_i enters directly as an argument of F_i and as an index on U_{tS_i}. If instead $F_i(s,x,u) = F_i(x,u)$ for all (s,u), this would rule out heterogeneity (across observations with different S_i) in responses to group unobservables or in the interactions between

$U_t(S_i)$. Second, by specifying the latent index Y_{it}^* as a random function, (2) provides a very general representation of heterogeneity across individuals in responses to covariates and the group effect. Finally, we drop the requirement that covariates act through an index, instead allowing a fully nonparametric specification with arbitrary heteroskedasticity.

As already noted, we focus on one type of endogeneity: correlation between group-level observables X_t and unobservables $U_t(S_i)$. This focus enables us to separate the unobserved variation responsible for endogeneity (i.e., $U_t(S_i)$) from that responsible for heterogeneity in responses and heteroskedasticity (i.e., the variation in the index functions F_i). With this structure, our model can characterize responses in outcomes to exogenous changes in the covariates X_t. In such a counterfactual, $U_t(S_i)$ should typically be held fixed while the full distribution of the latent responses Y_{it}^* (not just its mean, e.g.) is permitted to respond to the change in X_t. This is not possible in models like (1) with a single stochastic element.

We consider the question of nonparametric identifiability primarily to better understand what can be learned from common types of data in the context of models nested in the generalized regression framework. As emphasized since Koopmans (1945), identification and estimation are related but distinct. Even when estimation is likely to be carried out using parametric or semiparametric specifications, it is useful to understand whether such *a priori* restrictions are merely useful approximations in finite samples or are essential regardless of sample size.

We provide two types of nonparametric identification results. The first addresses full identification of the model, i.e., identification of the latent group effects and the distribution of $Y_{it}^*|S_i, X_t, U_t(S_i)$. These are the primitives of the model (2). The second, which we refer to as identification of the *structural outcome distribution*, concerns the identifiability of the group effects and the conditional distribution of outcomes, i.e., of $Y_{it}|S_i, X_t, U_t(S_i)$. Knowledge of the structural outcome distribution is sufficient for many (perhaps most) coun-

group observables and unobservables. On the other hand, setting $U_{tS_i} \equiv U_t$ would restrict the effect of unobservables on outcomes to be perfectly correlated across subpopulations in the same group but with different s_i. Further, with the monotonicity property assumed below, this would require the effect of a change in the group effect to have the same sign for all subpopulations.

terfactuals that motivate estimation in practice. For example, it describes how any moment (or quantile) of outcomes would change in response to an exogenous change in S_i or X_t. Thus, identification of average (or quantile) treatment effects is implied. Not surprisingly, the structural outcome distribution is identified under weaker conditions than those required for full identification.

The first challenge to identification is that group-level covariates may be correlated with group unobservables. We show that results from the recent literature on nonparametric instrumental variables regression (Chernozhukov and Hansen (2005)) can be applied to obtain identification of the latent group effects. Identification of the structural outcome distribution then follows. To obtain full identification of the model, we combine this result with additional separability and support conditions that enable us to apply standard arguments to trace out the conditional distribution of the latent index Y_{it}^*. The result does not rely on an "identification at infinity" argument, and the separability and large support assumptions are each testable.

As this sketch suggests, the results are obtained by applying well known ideas from the literatures on limited dependent variables and on nonparametric instrumental variables regression. Our main contribution is showing how these ideas can be combined to deliver new identification results for a rich class of models useful in a wide range of applications. Although our model involves limited dependent variables and permits high-dimensional heterogeneity, the structural outcome distribution is identified under the same instrumental variables conditions used by Chernozhukov and Hansen (2005) to show identification of quantile treatment effects or a nonparametric regression model. Further, for applications where identification of the structural outcome distribution does not suffice, full identification is obtained by adding the same kind of separability and support conditions previously used in even the simplest semiparametric models of limited dependent variables.

In the following section we discuss related literature. We describe the model and observables in sections 3 and 4. We present the identification results in section 5 and conclude in section 6.

2 Related Literature

Han (1987) introduced the generalized regression model (1), assuming $E_i \perp\!\!\!\perp X_i$. Manski (1987) considered binary choice in a panel setting and was the first to consider identification (and estimation) of nonlinear fixed effects model outside the small class of parametric models to which a traditional "differencing" is easily applied.[5] Abrevaya (2000) considered estimation of β in a fixed-effects version of (1), $Y_{it} = D \circ F(S_i\beta, U_t, E_{it})$ with $E_{it} \in \mathbb{R}$.[6] This is similar to a semiparametric version of our model. In addition to restricting individual unobservables to a scalar E_{it}, Abrevaya's model excludes the group-level covariates and variation in the group-level unobservable with S_i that we permit. We are able to allow these because we consider a different panel structure—a "large-N large-T" environment in contrast to the "small-T large-N" setup studied by Manski (1987) and Abrevaya (2000).

Altonji and Matzkin (2005) considered the model $Y_{it} = g(X_{it}, U_t, E_{it})$, which is similar to our model (2) with $g = D \circ F$.[7] They permit U_t to be a vector and do not require our "large-N large-T" environment. Like us, they rely on an exclusion restriction. Whereas we rely on conventional instrumental variables assumptions for nonseparable models, they assume existence of an excluded variable Z_{it} such that the conditional density restriction $f(U_t, E_{it}|\, X_{it}, Z_{it}) = f(U_t, E_{it}|\, Z_{it})$ holds.[8] An important limitation is that they consider identification only of a local average response.

Honoré and Lewbel (2002) studied a related semiparametric binary choice panel model $Y_{it} = 1\left\{X_{it}^{(1)} + X_{it}^{(2)}\beta + U_t + E_{it} > 0\right\}$. They permit U_t to be correlated with $\left(E_{it}, X_{it}^{(2)}\right)$ but require that $U_t + E_{it}$ be independent of $X_{it}^{(1)}$ conditional an excluded Z_{it}. They focus on identification and estimation of β, using a differencing strategy that relies on the additive

[5] He refers to this as a random effects model. See Chamberlain (1984) and Newey (1994) for related models.

[6] To ease comparisons, we have reversed the indices i and t from Abrevaya's setup, which focuses on individual fixed effects instead of group or time effects.

[7] They consider two types of estimators, one of which rules out qualitative response models and censoring. Our discussion here focuses on their first estimator, which admits such models.

[8] This is closely related to control function approaches in triangular models (e.g., Chesher (2003), Imbens and Newey (2006)). See also Matzkin (2004).

separability of the index function in $(U_t + E_{it})$.

Much prior work has considered identification in binary choice/threshold crossing models that are special cases of (1). Examples include Manski (1985), Manski (1988), Matzkin (1992), Lewbel (2000), Hong and Tamer (2004), Blundell and Powell (2004), Lewbel (2005), and Magnac and Maurin (2007). Work focusing on identification of an average derivative or average treatment effect includes Vytlacil and Yildiz (2007), Shaikh and Vytlacil (2005), and Hoderlein (2008). Ichimura and Thompson (1998) and Gautier and Kitamura (2007) studied binary choice in a linear random coefficients model, a special case of the model we consider, but without group effects.[9] Matzkin (2004) (section 5.1) considers a binary choice model making a distinction between group-specific unobservables and an additive preference shock, but without heteroskedasticity.[10] Blundell and Powell (2004), Matzkin (2004), and Hoderlein (2008) consider binary choice in semiparametric triangular models, leading to the applicability of control function methods or the related idea of "unobserved instruments."

In work simultaneous to our own, Chiappori and Komunjer (2009) study identification and testable restrictions of the cross-sectional model $Y_i = F(g(X_i) + E_i)$, where $F' > 0$. Aside from the panel environment we consider, a second fundamental difference is our distinction between unobservables that give rise to endogeneity and sources of randomness responsible for conditional heterogeneity in outcomes. As discussed above, this is essential for many types of counterfactuals in applications with both endogeneity and heterogeneity.

Finally, our own recent work (Berry and Haile (2009a), Berry and Haile (2009b)) considers identification of multinomial choice demand models, in some cases applying ideas used here as well. However, those results are obtained under more restrictive conditions than those here. For example, the results on identification of structural outcome distributions (i.e., "demand") in Berry and Haile (2009a) and Berry and Haile (2009b) rely on an index restriction on the way market- (group-) level unobservables enter utilities.[11] The results here show that this

[9] Generalizations to multinomial choice have been considered by Briesch, Chintagunta, and Matzkin (2005) and Fox and Gandhi (2009).

[10] See also Matzkin (2007a) and Matzkin (2007b).

[11] In Berry and Haile (2009b) we also require individual-level covariates (permitted but not required here)

restriction can be dropped in the case of binary choice, allowing identification of demand in a rich heterogeneous random utility model with endogeneity without any functional form or distributional assumption. For example, relative to the model of Ichimura and Thompson (1998), the specialization of the results here to binary choice allow us to drop their linearity restriction, add the group-level unobservable, and allow for endogeneity.

3 Model

We consider the generalized regression model (2), where the *index function* F_i of individual i is a specified as a random function on $\chi \equiv \text{supp}(S_i, X_t, U_t(S_i))$. Letting $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space, we define

$$F_i(S_i, X_t, U_t(S_i)) = F(S_i, X_t, U_t(S_i), \omega_i)$$

where $\omega_i \in \Omega$. Thus the model (2) can be written

$$Y_{it} = D \circ F(S_i, X_t, U_t(S_i), \omega_i). \tag{3}$$

Note that the measure $\mathbb{P}$ does not vary with $(S_i, X_t, U_t(S_i))$. This is without loss since arbitrary dependence of the distribution of Y_{it}^* on $(S_i, X_t, U_t(S_i))$ is already permitted through the function F.[12] Note also that there is no variation in the measure $\mathbb{P}$ across groups. This reflects the important assumption, already made, that the scalar $U_t(s)$ captures the effects of all unobservables common to individuals in group t with characteristics $S_i = s$. This generally rules out multidimensional group-specific unobservables, although we do permit $U_t(s)$ to vary freely across s and t.

Aside from the restriction to a scalar group effect, we have placed no restriction on the

and require separability and support conditions even for identification of the structural outcome distribution (not required here). One qualification is that, unlike some results in Berry and Haile (2009b) and Berry and Haile (2009a), in the present context we require fully independent (not just mean independent) instruments.

[12] Below we will discuss our assumptions on observables.

random index functions F_i. Note in particular that ω_i is not a random variable but an elementary event in Ω; any number of random variables can be defined as functions of ω_i. The following example illustrates.

Example 3. *A special case of our model is obtained with the linear random coefficients specification*

$$F\left(S_i, X_t, U_t\left(S_i\right), \omega_i\right) = S_i \gamma + X_t \beta_i + U_t\left(S_i\right) + \mathcal{E}_{it}$$

where the random variables $(\beta_i, \mathcal{E}_{it})$ *can be defined on* $(\Omega, \mathcal{F}, \mathbb{P})$ *as, for example,*

$$\begin{aligned} \beta_i &= \left(\beta^{(1)}\left(S_i, \omega_i\right), \ldots, \beta^{(K)}\left(S_i, \omega_i\right)\right) \\ \mathcal{E}_{it} &= \mathcal{E}(S_i, X_t, \omega_i). \end{aligned}$$

Although this special case of our model involves a strong functional form restriction, it permits an arbitrary joint distribution of $\left(\beta_i^{(1)}, \ldots, \beta_i^{(K)}, \mathcal{E}_{it}\right)$, *dependence of* β_i *on* S_i, *and dependence of* $\mathcal{E}_{it}$ *on* (S_i, X_t).

Because $U_t\left(S_i\right)$ is unobservable and enters as a nonseparable argument of a nonparametric function, it requires a normalization. We will assume for simplicity that $U_t\left(S_i\right)$ is continuously distributed in the population. Then, without further loss, for each s, we let $u = \Pr\left(U_t\left(S_i\right) \leq u | S_i = s\right)$, normalizing $U_t\left(S_i\right)$ to have a uniform [0,1] distribution conditional on each value of S_i.

Throughout our analysis we will maintain the following assumption.

Assumption 1. For each s in the support of S_i, there exists a known $\tilde{y} \in \mathbb{R}$ such that $\Pr\left(D \circ F\left(S_i, X_t, U_t\left(S_i\right), \omega_i\right) \leq \tilde{y} \middle| S_i = s, X_t = x, U_t\left(S_i\right) = u\right)$ is strictly decreasing in u for all x.

If we had assumed $F_i = F$, monotonicity of F in $U_t\left(S_i\right)$ would be without loss, merely defining an order on the latent group effect (for each s_i). Because we allow the index functions F_i to differ across individuals, Assumption 1 is a restriction. A strong sufficient condition is that F strictly increase in $U_t\left(S_i\right)$, as in Example 3.

4 Observables

We assume that the population distribution of Y_{it} is observed in every group t, conditional on the covariates (S_i, X_t) and excluded instruments $\tilde{Z}_t$.

Assumption 2. For all $y \in \text{supp}Y_i$, $\Pr_{\mathbb{P}}\left(Y_i \leq y|t, S_i, X_t, \tilde{Z}_t\right)$ is observed for all t and all $S_i, X_t, \tilde{Z}_t$.

Observation of $\left(t, Y_{it}, S_i, X_t, \tilde{Z}_t\right)$ for all i, t suffices, in which case we can think loosely of observing many individuals from each of many groups ("large N, large T"). However, this is not necessary. For example, in Example 1 it is sufficient to observe X_t and the market share of good 1 by demographic group (value of S_i) in many markets t.

Note that in general Assumption 2 rules out selection of S_i on individual-specific unobservables: conditional on any $(s_i, x_t.u_t(s_i))$ the observable distribution of outcomes Y_{it} is the the same as the conditional distribution in the population. As discussed previously, this reflects a choice to focus on endogeneity of group-level characteristics that arises through the group-level unobservable $U_t(S_i)$.

5 Results

5.1 Identification of the Structural Outcome Distribution

We first consider the identifiability of the *structural outcome distribution*, i.e., of the group effects and the distribution of Y_{it} conditional on $S_i, X_t, U_t(S_i)$. Fix $S_i = s$ and let $\tilde{y}$ be the value referred to in Assumption 1. Then , because conditioning on t fixes $U_t(s)$,

$$\begin{aligned} \Pr{}_{\mathbb{P}}(Y_{it} \leq \tilde{y}|t, X_t) &= \Pr{}_{\mathbb{P}}(Y_{it} \leq \tilde{y}|X_t, U_t(s)) \\ &= \rho(X_t, U_t(s)) \end{aligned} \tag{4}$$

where ρ is an unknown function that is strictly decreasing in its last argument by Assumption 1.

Absent endogeneity, identification of the function ρ would follow immediately from standard arguments (see, e.g., Matzkin (2003)). Because we allow for correlation between X_t and $U_t(S_i)$ conditional on S_i, identification of ρ will require instrumental variables. Let Z_t denote the exogenous conditioning variables at the group level. These may include some components of X_t as well as excluded instruments $\tilde{Z}_t$. We will require fully independent instruments:

Assumption 3. $U_t(S_i) \perp\!\!\!\perp Z_t$ conditional on S_i.

Now condition on a value of $W_t \equiv X_t \cap Z_t$, still fixing $S_i = s$. To simplify notation, let X_t now represent only the endogenous group covariates, with U_t representing the random variable $U_t(s)$.

Without loss, we may assume that Y_{it} and X_t have been transformed to have bounded support. We will focus on the case of continuous X_t, assuming X_t has a conditional density function $f_X(\cdot|Z_t)$.[13] Let $f_\rho(\cdot|X_t, Z_t)$ denote the conditional density of the random variable $\rho(X_t, U_t)$.[14] Fix some small positive constants $\epsilon_q, \epsilon_f > 0$ and for each $\tau \in (0,1)$ define $\mathcal{L}(\tau)$ to be the convex hull of functions $m(\cdot, \tau)$ satisfying

(i) for all $z \in \text{supp} Z_t$, $\Pr(\rho(X_t, U_t) \leq m(X_t, \tau) | \tau, Z_t = z) \in [\tau - \epsilon_q, \tau + \epsilon_q]$; and

(ii) for all $x \in \text{supp} X_t$, $m(x, \tau) \in p_x \equiv \{\rho : f_\rho(\rho | x, z) \geq \epsilon_f \ \forall z \text{ with } f_x(x|z) > 0\}$.

Assumption 4. (i) For any $\tau \in (0,1)$, for any bounded function $B(x,\tau) = m(x,\tau) - \rho(x,\tau)$ with $m(\cdot,\tau) \in \mathcal{L}(\tau)$ and $\varepsilon_t \equiv \rho(X_t, U_t) - \rho(x_t, \tau)$, $E[B(X_t,\tau)\psi(X_t, Z_t, \tau)|Z_t] = 0$ a.s. only if $B(X_t, \tau) = 0$ a.s. for $\psi(x,z,\tau) = \int_0^1 f_\varepsilon(\sigma B(x,\tau)|x,z) \, d\sigma > 0$; (ii) the density $\int_0^1 f_\varepsilon(e|x,z)$ of ϵ_t is bounded and continuous in e on $\mathbb{R}$ a.s.; (iii) $\rho(x,\tau) \in p_x$ for all (x,τ).

Assumption 4 is a "bounded completeness" condition ensuring that the instruments induce sufficient variation in the endogenous variables. This condition, which we take from

[13] Discrete endogenous group covariates can be accommodated by appealing below to Theorems 2 and 4 (and their associated rank conditions) in Chernozhukov and Hansen (2005) instead of their Theorem 4.

[14] Chernozhukov and Hansen's "rank invariance" property holds here because the same unobservable U_t determines potential values of $\rho(X_t, U_t)$ for all possible values of X_t.

Chernozhukov and Hansen (2005) (Appendix C), plays the role of the standard rank condition for linear models, but for the nonparametric nonseparable model

$$P_{it} = \rho\left(X_t, U_t\right)$$

obtained from (4). With Assumptions 1–4, identification of $\rho(x, u)$ at each x and u now follows directly from Theorem 4 of Chernozhukov and Hansen (2005). With the function ρ identified, so is each u_t. The argument can be repeated for every value of (W_t, S_i). Then, since the distribution of $Y_{it}|t, S_i, X_t$ is observed and the value of each $U_t(s)$ is now known, we have proved the following result.

Theorem 1. *Suppose Assumptions 1–4 hold. Then (i) the value of $U_t(s_i)$ is identified for each t and s_i; and (ii) the distribution of $Y_{it}|S_i, X_t, U_t(S_i)$ is identified.*

5.2 Full Identification

To obtain full identification we employ two additional assumptions: a separability restriction on the index function F, and a large support condition. Together these conditions enable the observed outcome distribution identify the distribution of the latent index using standard arguments.

We first consider the case in which the separability is in a group-level covariate. Partition X_t as $\left(X_t^{(1)}, X_t^{(2)}\right)$ with $X_t^{(1)} \in \mathbb{R}$.

Assumption 5. $\mathrm{supp} X_t^{(1)}| \left(S_i, X_t^{(2)}, U_t(S_i)\right) = \mathbb{R}$.

Assumption 6. $F(S_i, X_t, U_t(S_i), \omega_i) = X_t^{(1)} + f\left(S_i, X_t^{(2)}, U_t(S_i), \omega_i\right)$.

Assumption 5 is a "large support" condition, familiar from the literature on identification of binary choice models (e.g., Manski (1985), Matzkin (1992), Lewbel (2000)). This condition provides a standard benchmark for understanding what might be learned under ideal circumstances from the types of data typically available. Intuitively, extreme values of observables are needed to trace out the tails of the distribution of the random index.

Assumption 6 has two parts. The first restricts (3) by requiring additive separability of F in $X_t^{(1)}$. This provides a mapping between observed probabilities over Y_{it} and units of the latent Y_{it}^*.[15] The second part is a requirement of a unit coefficient on $X_t^{(1)}$. If Y_{it} takes on only two distinct values, the generalized regression model is a binary threshold crossing model, for which it is well known that one must choose the scale of the index function. In the case of binary choice, this is a normalization of the utility function and without loss. In a binary model where the units of Y_{it}^* are meaningful, the unit coefficient on $X_t^{(1)}$ is a restriction. Aside from binary models, however, the unit coefficient requirement is without further loss. This is because when Y_{it} takes on more than two values, any coefficient on $X_t^{(1)}$ is identified as long as Assumption 5 holds; thus, by rescaling $X_t^{(1)}$ we can obtain the required unit coefficient. This is shown in the following result, which also demonstrates that Assumption 6 is testable.

Proposition 2. *If Y_{it} takes on at least three distinct values and $F(S_i, X_t, U_t(S_i), \omega_i) = X_t^{(1)}\beta + f\left(S_i, X_t^{(2)}, U_t(S_i), \omega_i\right)$, then under Assumptions 1–5, (i) β is identified and (ii) Assumption 6 is testable.*[16]

Proof. Let $y_0 < y_1 < y_2$ be three distinct values in the support of Y_{it}, and suppose y_0 is the value $\tilde{y}$ referred to in Assumption 1.[17] Define $D^{-1}(y) = \sup\{\tau : D(\tau) \leq y\}$. Because $D(\cdot)$

[15] For binary choice, if $F_i(S_i, X_t, U_t(S_i))$ is assumed to be strictly increasing in $X_t^{(1)}$, then the event $\{F_i(S_i, X_t, U_t(S_i)) > 0\}$ is equivalent to the event $\left\{X_t^{(1)} > F_i^{-1}(0; S_i, U_t(S_i))\right\}$. This leads to an observationally equivalent model with separability in $X_t^{(1)}$. This is well known. Nonetheless, additive separability is not without loss under these assumptions. This is because there may be no monotonic transformation of the original utility function that leads to the separable form. For example, suppose that according to $F_i(S_i, X_t, U_t(S_i))$ the marginal rate of substitution between $X_t^{(1)}$ and S_t varies with $X_t^{(1)}$. This property would be preserved by any monotonic transformation but fails under separability. An implication is that there can be simultaneous changes in $X_t^{(1)}$ and S_i that would raise welfare under one model but lower welfare under the other. Thus, although the separable structure preserves individuals' ordinal rankings of the outside good and any inside good, it need not preserve their ordinal rankings of alternative inside goods. Nonetheless, the observational equivalence demonstrates why it may be difficult to obtain full identification without a restriction like the separability we assume. Note that quasilinearity also provides a cardinal representation of utility, enabling one to make aggregate welfare statements in the binary choice environment.

[16] Additional testable restrictions can be obtained if Y_{it} takes on at least four distinct values.

[17] If instead $\tilde{y} = y_1$ the argument follows the same logic, but starting by setting the observed probability below to α at $y = y_0$ instead of at $y = y_1$. If $\tilde{y} = y_2$ then a similar argument applies, but using the function $D^{-1}(y) = \inf\{\tau : D(\tau) \geq y\}$.

is weakly increasing, we must have $-\infty < D^{-1}(y_0) < D^{-1}(y_1) < \infty$. For all y we observe

$$\begin{aligned} \Pr\left(Y_{it} \leq y | S_i, X_t, U_t\left(S_i\right)\right) &= \Pr\left(D\left(X_t^{(1)}\beta + f\left(S_i, X_t^{(2)}, U_t\left(S_i\right), \omega_i\right)\right) \leq y\right) \\ &= \Pr\left(f\left(S_i, X_t^{(2)}, U_t\left(S_i\right), \omega_i\right) \leq D^{-1}\left(y\right) - X_t^{(1)}\beta\right). \end{aligned} \tag{5}$$

Recall that by Theorem 1, each $u_t\left(s_i\right)$ is known. Now fix $\left(S_i, U_t\left(S_i\right), X_t^{(2)}\right)$ and choose any $\alpha \in (0,1)$. Let $x\left(y_1\right)$ denote the supremum over the values of $X_t^{(1)}$ setting (5) to α when $y = y_1$. For $y \in \{y_0, y_1\}$, let $x\left(y\right)$ denote the supremum over the values of $X_t^{(1)}$that set (5) equal to α. Then since $D^{-1}(y_0)$ and $D^{-1}(y_1)$ are known, distinct, and finite, β is uniquely determined by

$$\beta = \frac{D^{-1}(y_1) - D^{-1}(y_0)}{x\left(y_1\right) - x\left(y_0\right)}. \tag{6}$$

Part (ii) follows by noting that (6) must hold for any $S_i, U_t\left(S_i\right), X_t^{(2)}, \alpha$. $\square$

The following result demonstrates that the preceding assumptions are sufficient for full identification of the model.

Theorem 3. *Suppose Assumptions 1–6 hold. Then the distribution of $Y_{it}^*|S_i, X_t, U_t\left(S_i\right)$ is identified.*

Proof. The hypotheses of Theorem 1 hold, so each $u_t\left(s_i\right)$ can be treated as known. Let $D^{-1}\left(\tilde{y}\right) = \sup\left\{q | D\left(q\right) \leq \tilde{y}\right\}$. Let $\tilde{y}$ be the value referred to in Assumption 1. Under Assumption 6,

$$\begin{aligned} \Pr\left(Y_{it} \leq \tilde{y} | t, S_i = s, X_t = \left(x^{(1)}, x^{(2)}\right)\right) &= \Pr\left(Y_{it}^* \leq D^{-1}(\tilde{y}) \middle| t, S_i = s, X_t = \left(x^{(1)}, x^{(2)}\right)\right) \\ &= \Pr\left(f\left(s, x^{(2)}, u_t\left(s\right), \omega_i\right) \leq D^{-1}(\tilde{y}) - x^{(1)}\right). \end{aligned}$$

Since the left side is observed and the value $D\left(\tilde{y}\right)$ is known, Assumption 5 ensures that the full CDF of $f\left(s, x^{(2)}, u_t\left(s\right), \omega_i\right)$ is identified for all $\left(s, x^{(2)}, u_t\left(s\right)\right)$. Since $Y_{it}^* = X_g^{(1)} + f\left(S_i, X_t^{(2)}, U_t\left(S_i\right), \omega_i\right)$, this gives the result. $\square$

Essentially the same argument can be applied if the separability is instead in a component of S_i. A qualification is that we must restrict the specification of the group effect. Let $S_i = \left(S_i^{(1)}, S_i^{(2)}\right)$ with $S_i^{(1)} \in \mathbb{R}$ and suppose that $U_t(S_i) = U_t\left(S_i^{(2)}\right)$. With this restriction, we can allow $S_i^{(1)}$ to play the role of the "special regressor." Consider now the following alternatives to Assumptions 5 and 6, noting that the same interpretations apply.

Assumption 7. supp $S_i^{(1)} \left| \left(S_i^{(2)}, X_t, U_t\left(S_i^{(2)}\right)\right) = \mathbb{R}\right.$.

Assumption 8. $F\left(S_i, X_t, U_t\left(S_i\right), \omega_i\right) = S_i^{(1)} + f\left(S_i^{(2)}, X_t, U_t\left(S_i\right), \omega_i\right)$.

Theorem 4. *Suppose $U_t(S_i) = U_t\left(S_i^{(2)}\right)$ and that Assumptions 1–4, 7, and 8 hold. Then the distribution of $Y_{it}^*|S_i, X_t, U_t\left(S_i^{(2)}\right)$ is identified.*

Proof. The hypotheses of Theorem 1 hold, so each $u_t(s_i)$ can be treated as known. Let $\tilde{y}$ be the value referred to in Assumption 1 and let $D^{-1}(\tilde{y}) = \sup\{q|D(q) \leq \tilde{y}\}$. Under Assumption 8

$$\begin{aligned}
\Pr\left(Y_i \leq \tilde{y}|t, S_i = \left(s^{(1)}, s^{(2)}\right), X_t = x\right) &= \Pr\left(Y_i^* \leq D^{-1}(\tilde{y})\middle| g, S_i = \left(s^{(1)}, s^{(2)}\right), X_t = x\right) \\
&= \Pr\left(f\left(s^{(2)}, x, u_t\left(s^{(2)}\right), \omega_i\right) \leq D^{-1}(\tilde{y}) - s^{(1)}\right).
\end{aligned}$$

Since the left side is observed and $D(\tilde{y})$ is known, Assumption 7 ensures that the full CDF of $f\left(s_i^{(2)}, x_t, u_t\left(s_i^{(2)}\right), \omega_i\right)$ is identified for all $\left(s_i^{(2)}, x_t, u_t\left(s_i^{(2)}\right)\right)$. The result then follows. $\square$

It should be clear from the proofs of Theorems 3 and 4 that testing the large support conditions is straightforward. Indeed, when the relevant support condition fails, the result is that the distribution of $Y_{it}^*|S_i, X_t, U_t(S_i)$ will be still be identified on a subset of its support. This also makes clear that the large support condition is required only to identify the tails of the distribution, not for an "identification at infinity" argument.

6 Conclusion

We have studied a nonparametric nonseparable generalized regression model in a panel setting with a "group" structure. Our model allows for individual level heterogeneity and for endogeneity through the group effect. We showed that the structural outcome distribution is identified under standard instrumental variables conditions. Because the structural outcome distribution is often the only primitive required for counterfactual simulations, this is an encouraging result. We should not be surprised that instrumental variables conditions are required in the presence of endogeneity. And despite the rich heterogeneity and the limited dependent variables setting, the conditions required are the same as those required in the simpler context of regression with a scalar unobservable.

We also considered full identification of the model and showed that this follows when standard separability and large support conditions are added. Although strong, the type of support condition we consider is commonly used to prove identification of simpler models of limited dependent variables. Further, these additional assumptions are testable, and identification is robust to relaxation of the large support condition in the sense that the tails of the support are required only for identification of tail probabilities.

Although identifiability is a necessary condition for existence of a consistent estimator, additional work would be needed to develop semi-parametric or nonparametric estimation methods for the kinds of models we have considered. Our identification results may suggest new estimation strategies, but this is a topic we leave to future work.

References

Abrevaya, J. (2000): “Rank Estimation of a Gnearlized Fixed Effects Regression Model,” *Journal of Econometrics*, 95, 1–23.

Altonji, J., and R. L. Matzkin (2005): “Cross-Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors,” *Econometrica*, 73, 1053–1102.

Berry, S. T., and P. A. Haile (2009a): “Identification in Differentiated Products Markets Using Market Level Data,” Discussion paper, Yale University.

——— (2009b): “Nonparametric Identification of Multinomial Choice Demand Models with Heterogeneous Consumers,” Discussion paper, Yale University.

Blundell, R. W., and J. L. Powell (2004): “Endogeneity in Semiparametric Binary Response Models,” *Review of Economic Studies*, 71, 655–679.

Briesch, R. A., P. K. Chintagunta, and R. L. Matzkin (2005): “Nonparametric Discrete Choice Models with Unobserved Heterogeneity,” Discussion paper, Northwestern University.

Chamberlain, G. (1984): “Panel Data,” in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. 2. Elsevier Science, Amsterdam.

Chernozhukov, V., and C. Hansen (2005): “An IV Model of Quantile Treatment Effects,” *Econometrica*, 73(1), 245–261.

Chesher, A. (2003): “Identification in Nonseparable Models,” *Econometrica*, 71, 1405–1441.

Chiappori, P.-A., and I. Komunjer (2009): “Correct Specification and Identifiction of Nonparametric Transformation Models,” Discussion paper, University of California San Diego.

Fox, J., and A. Gandhi (2009): "Identifying Heterogeneity in Economic Choice Models," Discussion paper, University of Chicago.

Gautier, E., and Y. Kitamura (2007): "Nonparametric Estimation in Random Coefficients Binary Choice Models," Discussion paper, Yale.

Han, A. K. (1987): "Nonparametric Analysis of a Generalized Regression Model," *Journal of Econometrics*, 35, 303–316.

Hoderlein, S. (2008): "Endogeneity in Semiparametric Binary Random Coefficient Models," Discussion paper, Brown University.

Hong, H., and E. Tamer (2004): "Endogenous Binary Choice Mode with Median Restrictions," *Economics Letters*, pp. 219–224.

Honoré, B. E., and A. Lewbel (2002): "Semiparametric Binary Choice Panel Data Models Without Strictly Exogenous Regressors," *emet*, 70(5), 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(2), 269–95.

Imbens, G., and W. Newey (2006): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," Discussion paper, M.I.T.

Koopmans, T. C. (1945): "Statistical Estimation of Simultaneous Economic Relations," *Journal of the American Statistical Association*, 40, 448–466.

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

——— (2005): "Simple Endogenous Binary Choice and Selection Panel Model Estimators," Discussion paper, Boston College.

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

Manski, C. F. (1985): "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator," *Journal of Econometrics*, 27, 313–333.

——— (1987): "Semiparametric Analysis of Random Effects Linear Models from Binary Panel Data," *Econometrica*, 55, 357–362.

——— (1988): "Identification of Binary Response Models," *Journal of the American Statitical Association*, 83(403), 729–738.

Matzkin, R. L. (1992): "Nonparametric and Distribution-Free Estimation of the Binary Choice and Threshold Crossing Models," *Econometrica*, 60(2).

——— (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71(5), 1339–1375.

——— (2004): "Unobservable Instruments," Discussion paper, Northwestern University.

——— (2007a): "Heterogeneous Choice," in *Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress of the Econometric Society*, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge University Press.

——— (2007b): "Nonparametric Identification," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 6B. Elsevier.

Newey, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349–1382.

Shaikh, A., and E. Vytlacil (2005): "Threshold Crossing Modle and Bounds on Treatment Effects: A Nonparametric Analysis," Discussion paper, Univeristiy of Chicago.

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