cemmap
centre for microdata methods and practice

Identification of a class of index models: a topological approach

Mogens Fosgerau
Dennis Kristensen

The Institute for Fiscal Studies
Department of Economics,
UCL

cemmap working paper CWP52/19

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
An ESRC Research Centre

Identification of a class of index models: A topological approach*

Mogens Fosgerau[†] Dennis Kristensen[‡]

October 7, 2019

Abstract

We establish nonparametric identification in a class of so-called index models using a novel approach that relies on general topological results. Our proof strategy imposes very weak smoothness conditions on the functions to be identified and does not require any large support conditions on the regressors in our model. We apply the general identification result to additive random utility and competing risk models.

KEYWORDS: nonparametric identification, discrete choice, competing risks, index
JEL codes: C14, C35, C36, C41

1 Introduction

We develop a novel nonparametric identification result for the following class of models,

$$\Pi\left(w, x, z\right) = \Lambda\left(g\left(w\right) + h\left(x\right), z\right), \tag{1}$$

*This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 740369).
[†]Department of Economics, University of Copenhagen. E-mail: mogens.fosgerau@econ.ku.dk
[‡]Department of Economics, University College London. E-mail: d.kristensen@ucl.ac.uk

where $\Lambda : \mathbb{R}^J \times \mathbb{R}^{d_Z} \mapsto \mathbb{R}^J$, $g : \mathbb{R}^{d_W} \mapsto \mathbb{R}^J$, and $h : \mathbb{R}^{d_X} \mapsto \mathbb{R}^J$ are all vector-valued functions of dimension $J \geq 1$. The arguments $w \in \mathbb{R}^{d_J}$ and $x \in \mathbb{R}^{d_X}$ represent the values of two sets of regressors, W and X, while $z \in \mathbb{R}^{d_Z}$ corresponds to values of a set of control variables, Z. We take as high-level assumption that we know (have observed from data) the function $\Pi(w, x, z)$ for $(w, x, z) \in \text{supp}(W, X, Z)$ from which we then wish to identify the unknown functions $\Lambda(a, z)$ and $h(x)$, while we treat the function $g(w)$ as being known. Here, and in the following, $\text{supp}(A)$ denotes the support of any given random variable A. We refer to this class of models as index models since W and X are restricted to enter the model through $g(W)$ and $h(X)$, respectively.

We make three major contributions relative to the existing literature: First, we do not impose any large support conditions on any of the regressors in our model, which is in contrast to most existing results on identification of this class of models. Second, we impose very weak smoothness conditions on the functions of interest; in particular, we do not require continuity or differentiability in order to obtain identification of Λ and h while most existing results as a minimum require the underlying functions to be differentiable. Third, we show how the presence of the controls Z can help to achieve identification in a nontrivial way: We first show local identification at each value of the control Z. Suitable variation in Z then allows us to piece the locally identified components together across different values of Z to achieve global identification.

Our proof strategy relies on arguments from general topology that, to our knowledge, are completely new to the literature on nonparametric identification. These should be of general interest since they can be used for identification in other settings. A key element of our approach is the notion of relative identification: We say that a function $a(w, x)$ is *relatively identified* on a set $\mathcal{M}$ if there exists x such that for all $(w', x') \in \mathcal{M}$ there exists w with $(w, x) \in \mathcal{M}$ and $a(w, x) = a(w', x')$. If a is indeed relatively identified on $\mathcal{M}$, then for any point $(w', x') \in \mathcal{M}$ we can use injectivity of Λ to find w such that $a(w, x) = a(w', x')$. We will apply this concept to $a(w, x) := g(w) + h(x)$ in the above model. Since $g(w)$ is treated as known, we can therefore identify the difference $h(x) - h(x')$. Importantly, we will not have to require any continuity of h or the domain of x to employ relative identification.

Relative to "identification at infinity", as discussed below, we only require relative identification on small sets. Extending the identification result so to hold on larger set is achieved through the second main ingredient of our argument, which is the topological notion of a connected set. By definition, a connected set cannot be contained in the union of two non-empty disjoint open sets while having non-empty intersection with both. In other words, it is not possible to split a connected set into disjoint subsets that are separated by being contained within disjoint open sets. We will then require the image of $a(W, X)$ to be covered by open sets, within each of which we have relative identification. The image being connected then ensures that local identification extends to all of the image. Like us, Berry and Haile (2018) and Evdokimov (2010), among others, rely on connectness to achieve global identification but in these papers the restriction is imposed directly on the support of the covariates thereby implicity restricting the covariates to be continuous. In contrast, we impose connectedness on the image of $a(W, X)$ and so allow for both X and W to contain discrete components.

Two leading examples that fall within our general framework are nonparametric versions of additive discrete choice models and competing risk models as shown in the next section. There is a large literature on identification and estimation of semiparametric multinomial choice models (see,e.g., Manski, 1975; Lewbel et al., 2000). In contrast, the literature on nonparametric identification is quite thin with few results having been developed since the seminal work of Matzkin (1993). In terms of modelling, Theorem 2 in Matzkin (1993) is probably the most related to our result, but the assumptions and identification strategy of this theorem are very different from ours. Our and her set of assumptions are not clearly ranked with some of our assumptions being stronger while others weaker compared to hers. One key feature of her proof strategy is the introduction of assumptions that ensure the multinomial model may be converted to a binary choice problem and then employment of an "identification at infinity" argument. This assumes availability of a set of special regressors with full support; identification is then achieved by sending each of these special regressors off to infinity. This is an example of what Khan and Tamer (2010) call "thin set identification" which they show leads to irregularly behaved estimators. In contrast, we achieve identification as long as $g(w) + h(x)$ exhibits sufficient, but potentially bounded, variation. More recently, Allen and Rehbeck (2019) provide conditions under which one can identify how regressors alter the desirability of alternatives using only average demands. Their conditions are weaker than ours but on the other hand they are only able to identify certain features of the model.

There is also a nascent literature on nonparametric identification of so-called BLP models (Berry et al., 1995) as used in industrial organization; see, for example, Berry and Haile (2018) and Chiappori et al. (2018). The setting of the BLP model is somewhat different, though, since there the observed choice probabilities contain unobserved product characteristics that have to be controlled for. This leads to a different identification problem compared to ours.

Finally, there is also a literature on identification in competing risk models. The two most closely related papers in terms of modelling are Heckman and Honoré (1989) and Lee and Lewbel (2013). Heckman and Honoré (1989) achieves identification by letting the index of the duration variable go to zero and so their result falls in the "thin set identification" category. Lee and Lewbel (2013) provide a high-level assumption for identification involving a rank condition of an integral operator. Primitive conditions for this to hold are not known. We complement these two studies by showing identification under primitive conditions without relying on "thin set identification".

In the next section, we give two motivating examples in form of a random utility model and a competing risk model that both fall within the setting of eq. (1). We present our general framework in Section 3 and the assumptions we will work under, and provide our identification results in Section 4. Section 5 concludes.

2 Two Motivating Examples

The model (1) comprises a range of models that are met in economics. We here present two classes of models that fall within our framework.

2.1 Discrete choice models

We here first demonstrate that additive random utility models (ARUM) belong to the class of models (1). Using existing results in the literature, this implies that our results also apply to a broad class of rational inattention discrete choice models (Fosgerau et al., 2018) and an even wider class of perturbed utility models.

2.1.1 Additive random utility

Consider an agent choosing between $J+1$ alternatives, each of which being associated with an indirect utility of the form

$$U_j = g_j(W) + h_j(X) + \varepsilon_j, \quad j = 0, 1, ..., J,$$

where (W, X) is a set of observed covariates while $\varepsilon = (\varepsilon_0, \varepsilon_1, ..., \varepsilon_J)$ is unobserved. This model was initially proposed by McFadden (1974) and has since become one of the workhorses in applied microeconomics; see e.g. Ben-Akiva and Lerman (1985) and Maddala (1986). As is standard in the literature, we impose the following normalization on the "outside"option $j = 0$: $g_0(w) = h_0(x) = 0$. We collect the remaining functions in $g(W) = (g_1(W), ..., g_J(W))$ and $h(X) = (g_1(X), ..., g_J(X))$.

Some of the regressors (W, X) may potentially be dependent on ε. To handle this situation, we assume the availability of a set of control variables Z so the following conditional independence assumption is satisfied:

Assumption 1 $F_{\varepsilon|(W,X,Z)} = F_{\varepsilon|Z}$ *where* $F_{\varepsilon|Z}$ *has a conditional density with full support and finite first moments.*

In addition to (W, X, Z), the researcher also observes the utility maximizing choice, $D = \arg\max_{j\in\{0,1,...,J\}} U_j$. Thus, the conditional choice probabilities,

$$\Pi_j(w, x, z) := P(D = j|\,(W, X, Z) = (w, x, z)), \quad j = 0, 1, ..., J, \tag{2}$$

are identified in the population. We collect these in the vector-valued function $\Pi(w, x, z) = \{\Pi_j(w, x, z) : j = 1, ..., J\} \in \mathbb{R}^J$ where we leave out the choice probability of the outside option. Define the surplus function

$$G(a_0, ...a_J, z) = E\left[\max_{j=0,...,J}\{\varepsilon_j + a_j\}\,|Z = z\right],$$

for any given $(a_0, a_1, ..., a_J) \in \mathbb{R}^{J+1}$. Then by the conditional independence in Assumption 1 together with the Williams-Daly-Zacchary Theorem (McFadden, 1981), the gradient of the surplus function is the vector of choice probabilities. That is, for $j = 1, ..., J$,

$$\Pi_j(w, x, z) = \Lambda_j(g(w) + h(x), z), \quad \Lambda_j(a_1, ...a_J, z) = \left.\frac{\partial G(a_0, a_1...a_J, z)}{\partial a_j}\right|_{a_0=0},$$

and so (1) holds. Moreover, due to $F_{\varepsilon|Z}$ having a conditional density with full support, the conditional choice probability mapping $\Lambda(\cdot, z)$ is invertible for each z, c.f. Hofbauer and Sandholm (2002, Thm 2.1), and so injective which we will require to achieve identification.

Fosgerau et al. (2018) show that any ARUM satisfying the conditions above is observationally equivalent to a rational inattention discrete choice model in which the prior is held constant. This generalizes the result by Matějka and McKay (2015) that the multinomial logit model has a foundation as a rational inattention model. The Fosgerau et al. (2018) result implies that our identification result extends without effort to a broad class of rational inattention models.

2.1.2 Perturbed utility

The class of perturbed utility models (Fosgerau et al., 2012; Fudenberg et al., 2015) generalizes the class of additive random utility models. As shown by Hofbauer and Sandholm (2002), the conditional choice probabilities of an ARUM arise as the solution to a utility maximization problem where a consumer chooses the vector of choice probabilities to maximize a function that consists of a linear term and a concave term. Here we present a more general version that allows controls to affect the concave term, i.e.

$$\Lambda(a, z) = \arg\max_{q \in \Delta} \{a^\intercal q + \Omega(q|z)\},$$

where $a \in \mathbb{R}^{J+1}$ is a vector of utility indexes, $\Delta = \{q \in \mathbb{R}_+^{J+1} : \sum_{j=0}^J q_j = 1\}$ is the unit simplex and $\Omega(\cdot|z)$ is a concave function for each $z \in \mathcal{Z}$. We further specify the indexes as $a = a(w, x) = g(w) + h(x)$ in which case the implied conditional choice probabilities $\Pi(w, x, z)$ satisfy (1).

We will show that Λ is injective for each z as a function of a, which will include perturbed utility models among the models for which we establish identification. First, in order to rule out zero demands, we assume that the norm of the gradient $\nabla_q \Omega(q|z)$ approaches infinity as q approaches the boundary of the unit simplex. Second, we assume that $\Omega(q|z)$ is differentiable.[1] Third, we normalize the outside option so that $g_0(w) = h_0(x) = 0$. Now, for each value of the control z, the demand solves the first-order condition for an interior solution

$$a + \nabla_q \Omega(\Lambda(a, z)|z) = \lambda \iota,$$

where λ is a scalar constant and $\iota \in \mathbb{R}^J$ is a vector consisting of ones. To show that Λ is injective, consider this equation at a_1 and a_2 and assume that $\Lambda(a_1, z) = \Lambda(a_2, z)$. Define a matrix M such that $Mx = x - x_0\iota$ for all $x = (x_0, ..., x_J) \in \mathbb{R}^{J+1}$. Pre-multiply this matrix onto the first-order condition to obtain that

$$a_1 + M\nabla_q \Omega(\Lambda(a_1, z)|z) = a_2 + M\nabla_q \Omega(\Lambda(a_2, z)|z),$$

which implies that $a_1 = a_2$ as required.

2.2 Accelerated failure time models for competing risks

Consider a competing risk model as in Heckman and Honoré (1989) with J competing causes of failure. A latent failure time $T_j > 0$ is associated with each cause $j \in \{1, ..., J\}$. The econometrician observes the duration until the first failure, $Y = \min_{j\in\{1,...,J\}} T_j$, and the associated cause of failure, $D = \arg\min_{j\in\{1,...,J\}} T_j$, together with a set of covariates (X, W, Z). Assume that the jth failure time satisfies

$$\ln T_j = g_j(W) + h_j(X) - \varepsilon_j,$$

for some functions g_j and h_j, $j = 1, ..., J$. We collect the unobservables in $\varepsilon = (\varepsilon_1, ..., \varepsilon_J)$ which again is required to satisfy Assumption 1. The model may then be termed a multivariate generalized accelerated failure time model (Kalbfleisch

[1] Note we do not require a Hessian.

and Prentice, 1980; Fosgerau et al., 2013).

The econometrician has knowledge of

$$\Pi_j(w,x|z) = E(\ln Y|(W,X,Z) = (w,x,z)) \cdot P(D = j|(W,X,Z) = (w,x,z)), : j = 1,...,J,$$

which satisfies eq. (1) with

$$\Lambda_j(a,z) = G(a,z) \cdot \frac{\partial G(a,z)}{\partial a_j}, : j = 1,...,J,$$

where as before $a = (a_1,...,a_J)$ while $G(a,z)$ is now the expected log failure time,

$$G(a,z) = E[\ln Y|g(W) + h(X) = a, Z = z] = -E\left[\max_{j=1,\ldots,J}\{-a_j + \varepsilon_j\}|Z = z\right].$$

Injectivity of $\Lambda(a,z) = (\Lambda_1(a,z),...,\Lambda_J(a,z))$ for each value of z follows from Assumption 1 by recycling the arguments of the previous section where now normalization of one of the causes of failure is not required since the level $G(a,z)$ is included.

3 General framework

We now return to the general model given in eq. (1) where $g : \mathbb{R}^J \to \mathbb{R}^J$ is a known function while $h : \mathbb{R}^{d_X} \to \mathbb{R}^J$ and $\Lambda : \mathbb{R}^J \times \mathbb{R}^{d_Z} \to \mathbb{R}^J$ are unknown functions. We take $\Pi(w,x,z)$ as given and known to us for all $(w,x,z) \in \text{supp}(W,X,Z) \subseteq \mathbb{R}^J \times \mathbb{R}^{d_X} \times \mathbb{R}^{d_Z}$ where (W,X,Z) denote the random variables that we have observed, c.f. the examples in the previous section.

For notational convenience, define

$$a(w,x) := g(w) + h(x). \tag{3}$$

Given that g is known to us, identification of a is equivalent to identification of h. We then wish to identify the functions $a(w,x)$ and $\Lambda(a,z)$ for $(w,x) \in \mathcal{M}_0$, $a \in \mathcal{A}_0$ and $z \in \mathcal{Z}_0$ where the sets $\mathcal{M}_0$, $\mathcal{A}_0$ and $\mathcal{Z}_0$ are defined below. Specifically, these sets will be constructed according to certain features of the underlying covariates and the functions of interest as explained in the following.

The covariates W play a special role in our approach in that we need sufficient continuous variation in these to achieve identification. First, we will throughout require that $\dim(W) = J$ in order to vary each of the J components of a in $\Lambda(a, z)$ independently of each other. We will now strengthen this and require that W exhibit continuous variation: For given values of $z \in \operatorname{supp}(Z)$ and $x \in \operatorname{supp}(X|Z = z)$, define

$$\mathcal{M}(z) = \cup_{x \in \operatorname{supp}(X|Z=z)} \operatorname{int} \operatorname{supp}(W|X = x, Z = z) \times \{x\} \subseteq \mathbb{R}^J \times \mathbb{R}^{d_X}, \qquad (4)$$

where $\operatorname{int}\mathcal{M}$ denotes the interior of a given set $\mathcal{M}$, and let

$$\mathcal{A}(z) = a(\mathcal{M}(z)) = \{a(w, x) \,|\, (w, x) \in \mathcal{M}(z)\} \subseteq \mathbb{R}^J \qquad (5)$$

denote the image of $\mathcal{M}(z)$ under a. We then wish to show that a and Λ are identified on

$$\mathcal{M}_0 = \cup_{z \in \mathcal{Z}_0} \mathcal{M}(z), \text{ and } \mathcal{A}_0 = \cup_{z \in \mathcal{Z}_0} \mathcal{A}(z), \qquad (6)$$

respectively, where $\mathcal{Z}_0 \subseteq \operatorname{supp}(Z)$ will be specified below.

At a first glance, the construction of $\mathcal{M}_0$ in terms of $\mathcal{M}(z)$ may look somewhat odd and one could perhaps be tempted to instead attempt to achieve identification on $\tilde{\mathcal{M}}_0 = \cup_{z \in \mathcal{Z}_0} \tilde{\mathcal{M}}(z)$ where $\tilde{\mathcal{M}}(z) = \operatorname{int} \operatorname{supp}(W, X|Z = z)$. However, the alternative version $\tilde{\mathcal{M}}(z)$ will be empty if, for example, X is discrete; in contrast, $\mathcal{M}(z)$ will be non-empty as long as $\operatorname{supp}(W|X = x, Z = z)$ has non-empty interior for some values of (x, z), regardlessly of X having discrete components. A sufficient condition for $\mathcal{M}(z)$ to be non-empty is that the distribution of $W|X = x, Z = z$ has a continuous component but it allows for this to be combined with discrete components. However, if the discrete support points are not contained within the support of the continuous component, we will not be able to show identification at these values. This also rules out that some components of W are included in Z since in this case $\operatorname{int} \operatorname{supp}(W|X = x, Z = z) = \emptyset$. At the same time, however, (X, W) are allowed to be dependent on Z; we just need sufficient variation in (X, W) conditional on Z. Finally, we would like to stress that we do not impose any large-support restrictions on W, which is in contrast to most existing results in the literature; see discussion in the Introduction.

Observe the dependence of $\mathcal{M}_0$ and $\mathcal{A}_0$ on the set $\mathcal{Z}_0 \subseteq \text{supp}(Z)$. To achieve "maximal" identification, we would ideally like to choose $\mathcal{Z}_0 = \text{supp}(Z)$. However, we potentially have to restrict $\mathcal{Z}_0$ so that the functions $a \mapsto \Lambda(a, z)$ and $w \mapsto g(w)$ satisfy certain conditions as we vary (w, x) over $\mathcal{M}_0$. Specifically, we implicitly restrict $\mathcal{Z}_0$ so that the following two assumptions are satisfied on the resulting set:

Assumption 2 *For any $z \in \mathcal{Z}_0$, $a \mapsto \Lambda(a, z)$ is injective on $\mathcal{A}(z)$ as defined in (5).*

Assumption 3 *$w \mapsto g(w)$ is injective and, for any $z \in \mathcal{Z}_0$, takes open sets to open sets on $\left\{w \in \mathbb{R}^J : \exists x \in \mathbb{R}^{d_X} : (w, x) \in \mathcal{M}(z)\right\}$.*

In a given application, Assumptions 2-3 may not hold for all $z \in \text{supp}(Z)$ in which case we will remove such values from $\mathcal{Z}_0$. In the worst case scenario, this leaves us with $\mathcal{Z}_0$ being empty and our identification result becomes void. At the other extreme, $\mathcal{Z}_0 = \text{supp}(Z)$ and we achieve identification on the whole support. Regarding Assumption 2, asking that Λ is injective is weaker than the assumption that $\varepsilon | Z = z$ has a continuous distribution, which we used in our two examples to guarantee injectivity. Regarding Assumption 3, first note that functions that take open sets into open sets are referred to as open maps in topology. A sufficient condition for Assumption 3 to hold is that g is one-to-one with its inverse being continuous. However, an open map does not necessarily have to be continuous and so Assumptions 2-3 potentially allow for discontinuities in both Λ and g.

Due to the structure of $a(w, x)$, as given in eq. (3), it follows from the definition of $\mathcal{M}(z)$ together with Assumption 3 that $\mathcal{A}(z)$ and thereby also $\mathcal{A}_0$ are open sets. We add to this by also requiring it to be connected. Recall that a set $\mathcal{A}$ is said to be connected if for any non-empty open sets $\mathcal{O}_1$ and $\mathcal{O}_2$ the following holds: $\mathcal{A} \subseteq \mathcal{O}_1 \cup \mathcal{O}_2$, $\mathcal{A} \cap \mathcal{O}_1 \neq \varnothing$, $\mathcal{A} \cap \mathcal{O}_2 \neq \varnothing \Rightarrow \mathcal{O}_1 \cap \mathcal{O}_2 \neq \varnothing$. Thus a connected set cannot be contained in two non-empty disjoint open sets. We then impose:

Assumption 4 *$\mathcal{A}_0$ defined in eq. (6) is connected.*

The latter assumption allows us to extend identification from each $\mathcal{A}(z)$ to all of $\mathcal{A}_0$, but, importantly, without requiring that each $\mathcal{A}(z)$ is connected. It still requires, however, that we can connect all points in $\mathcal{A}_0$ as we vary $z \in \mathcal{Z}_0$. Note the potential tension between this last assumption and Assumptions 2-3: It may very well be that the set of values for z at which Assumptions 2-3 hold is so small that the corresponding set $\mathcal{A}_0$ becomes disconnected. This situation can be handled under additional constraints: Suppose, for example, $\mathcal{Z}_0 = \mathcal{Z}_{0,1} \cup \mathcal{Z}_{0,2}$ where $\mathcal{A}_{0,k} = \cup_{z \in \mathcal{Z}_{0,k}} \mathcal{A}(z)$, $k = 1, 2$, are two connected sets but $\mathcal{A}_1 \cup \mathcal{A}_2$ is not connected. Then we can apply the results presented in the next section to $\mathcal{Z}_{0,1}$ and $\mathcal{Z}_{0,2}$ separately to achieve identification on $\mathcal{A}_{0,1} \cup \mathcal{A}_{0,2}$. This does, however, require two normalizations – one for each of the two separate sets – since we need to impose the following normalization on the function h:

Assumption 5 *There exists known $z_0 \in \mathcal{Z}_0$ and $(w_0, x_0) \in \mathcal{M}(z_0)$ so that $h(x_0) = 0$.*

This is needed to identify the level of h since, for any given pair of (Λ, h), we have $\Lambda(g(w) + h(x), z) = \tilde{\Lambda}\left(g(w) + \tilde{h}(x), z\right)$ where $\tilde{\Lambda}(a, z) = \Lambda(a + c, z)$ and $\tilde{h}(x) = h(x) - c$ for some given value of $c \in \mathbb{R}^J$.

4 Results

As explained earlier, we shall make use of the following notion of relative identification in our proof of identification:

Definition 1 *A function $a(w, x)$ is relatively identified on a given set $\mathcal{M}$ if there exists x such that for all $(w', x') \in \mathcal{M}$ there exists w with $(w, x) \in \mathcal{M}$ and $a(w, x) = a(w', x')$.*

In particular, if a is relatively identified on $\mathcal{M}$ and if a is identified at a point in $\mathcal{M}$, then a is also identified on all of $\mathcal{M}$. Note that $\mathcal{M}$ may, for example, take the form of $\mathcal{M} = \mathcal{W} \times \{x\}$ in which case we are only able to identify $a(w, x)$ at the single value x. We then have the following first result:

Lemma 1 *Under Assumptions 2-3 the following hold: For any $z \in \mathcal{Z}_0$ and $(w, x) \in \mathcal{M}(z)$, there exists an open neighbourhood $\mathcal{O} \subseteq \mathcal{A}(z)$ of $a(w, x)$ such that $a(\cdot, \cdot)$ is relatively identified on $a^{-1}(\mathcal{O}) \cap \mathcal{M}(z)$.*

Proof. Let $(w, x) \in \mathcal{M}(z)$ be given. By definition of $\mathcal{M}(z)$, there exists an open neighbourhood $\mathcal{O}_w$ of w such that $(w', x) \in \mathcal{M}(z)$ for all $w' \in \mathcal{O}_w$, which in turn implies that $\mathcal{O} := a(\mathcal{O}_w, x) = g(\mathcal{O}_w) + h(x) \subseteq \mathcal{A}(z)$ is open by Assumption 3. Now, consider an arbitrary point $(w', x') \in a^{-1}(\mathcal{O}) \cap \mathcal{M}(z)$. By construction of $\mathcal{O}$, there exists $w'' \in \mathcal{O}_w$ such that $a(w'', x) = a(w', x')$ Then also $\Pi(w', x', z) = \Pi(w'', x, z)$. By definition this corresponds to solving $\Lambda(g(w') + h(x'), z) = \Lambda(g(w'') + h(x), z)$ w.r.t. w'', which will have a unique solution since $a \mapsto \Lambda(a, z)$ is injective on $\mathcal{A}(z)$ by Assumption 2 and g is invertible by Assumption 3. Thus, w'' is well-defined and unique, and is identified since Π is known/identified. Hence, $a(\cdot, \cdot)$ is relatively identified on $a^{-1}(\mathcal{O})$. ■

Assumption 4 now allows us to take the step from relative identification on a cover of $\mathcal{A}_0$ by open sets to identification everywhere on $\mathcal{M}_0$:

Theorem 1 *Under Assumptions 2-4, $a(\cdot, \cdot)$ is identified on $\mathcal{M}_0$.*

Proof. For any given $z \in \mathcal{Z}_0$, recall that $\mathcal{A}(z)$ is open. Using Lemma 1, find an open cover of $\mathcal{A}(z)$, $\cup_{i \in \mathcal{I}_z} \mathcal{O}_i(z)$, where $\mathcal{O}_i(z) \subseteq \mathcal{A}(z)$ and a is relatively identified on each $a^{-1}(\mathcal{O}_i(z)) \cap \mathcal{M}(z)$, $i \in \mathcal{I}_z$. Then

$$\cup_{z \in \mathcal{Z}_0} \cup_{i \in \mathcal{I}_z} a^{-1}(\mathcal{O}_i(z)) \cap \mathcal{M}(z) \tag{7}$$

is an open cover of $\mathcal{M}_0$ where a is relatively identified on each $a^{-1}(\mathcal{O}_i(z)) \cap \mathcal{M}(z)$.

Define $\mathcal{S} \subseteq \mathcal{M}_0$ as the set of points at which $a(\cdot, \cdot)$ is identified. $\mathcal{S}$ is nonempty, since $(w_0, x_0) \in \mathcal{S}$.

Consider any $\mathcal{O}_i(z)$ for $i \in \mathcal{I}_z$ and $z \in \mathcal{Z}_0$, so that $a(\mathcal{S}) \cap \mathcal{O}_i(z) \neq \emptyset$. By construction, $\mathcal{O}_i(z) \subseteq \mathcal{A}(z)$ and hence $\mathcal{S} \cap a^{-1}(\mathcal{O}_i(z)) \cap \mathcal{M}(z) \neq \emptyset$. Hence $a^{-1}(\mathcal{O}_i(z)) \cap \mathcal{M}(z) \subseteq \mathcal{S}$ due to Lemma 1. Moreover, $\mathcal{O}_i(z) \subseteq a(\mathcal{S})$ for all such $\mathcal{O}_i(z)$ and we conclude that $a(\mathcal{S})$ is open.

Now suppose, to obtain a contradiction, that identification does not extend to all of $\mathcal{M}_0$, i.e. $\mathcal{M}_0 \backslash \mathcal{S} \neq \emptyset$. By assumption, there exists $\mathcal{O}_i(z)$ with $\mathcal{O}_i(z) \cap a(\mathcal{M}_0 \backslash \mathcal{S}) \neq \emptyset$, and we have $\mathcal{O}_i(z) \cap a(\mathcal{S}) = \emptyset$ since otherwise $a^{-1}(\mathcal{O}_i(z)) \cap \mathcal{M}(z) \subseteq \mathcal{S}$. Then the union of such such neighbourhoods $\mathcal{O}_i(z)$ is open and disjoint from the open set $a(\mathcal{S})$. But this is a contradiction since $\mathcal{A}_0$ is connected. ■

Importantly, the above argument shows that it is possible to achieve identification across different values of z without continuity due to connectedness of the image $\mathcal{A}$. Moreover, this result holds without having identified Λ. Once we have identified a we can also identify Λ:

Theorem 2 *Under Assumptions 2-4, Λ is identified on $\mathcal{A}_0$.*

Proof. Let $z \in \mathcal{Z}_0$ and $a \in \mathcal{A}(z)$ be given. By definition of $\mathcal{A}(z)$, there exists some pair $(w, x) \in \mathcal{M}(z)$ such that $a = a(w, x)$. Since $a(\cdot, \cdot)$ is identified, the pair (w, x) is also identified. But then we also know $\Pi(w, x, z)$ and so $\Lambda(a, z) = \Pi(w, x, z)$ is uniquely identified. ■

5 Conclusion

We have established an general identification result for a wide class of index models, whereby identification relies solely on general topological properties. Smoothness is not required. No large support condition is imposed on the regressors. Controls variables may contribute to achieving identification.

References

Allen, R. and Rehbeck, J. (2019) Identification with Additively Separable Heterogeneity *Econometrica* .

Ben-Akiva, M. and Lerman, S. R. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand* Vol. 6 MIT Press Cambridge, MA.

Berry, S. T. and Haile, P. A. (2018) Identification of Nonparametric Simultaneous Equations Models With a Residual Index Structure *Econometrica* **86**(1), 289–315.

Berry, S. T., Levinsohn, J. and Pakes, A. (1995) Automobile Prices in Market Equilibrium *Econometrica* **63**(4), 841–890.

Chiappori, P.-A., Komunjer, I. and Kristensen, D. (2018) Nonparametric identification and estimation of discrete choice models.

Evdokimov, K. (2010) Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity.

Fosgerau, M., McFadden, D. and Bierlaire, M. (2013) Choice probability generating functions *Journal of Choice Modelling* **8**.

Fosgerau, M., McFadden, D. L., Fosgerau, M. and McFadden, D. L. (2012) A theory of the perturbed consumer with general budgets *NBER Working Paper* pp. 1–27.

Fosgerau, M., Melo, E., de Palma, A. and Shum, M. (2018) Discrete Choice and Rational Inattention: A General Equivalence Result *SSRN Electronic Journal* .

Fudenberg, D., Iijima, R. and Strzalecki, T. (2015) Stochastic Choice and Revealed Perturbed Utility *Econometrica* **83**(6), 2371–2409.

Heckman, J. J. and Honoré, B. E. (1989) The identifiability of the competing risks model *Biometrika* **76**(2), 325–330.

Hofbauer, J. and Sandholm, W. H. (2002) On the global convergence of stochastic fictitious play *Econometrica* **70**(6), 2265–2294.

Kalbfleisch, J. D. and Prentice, R. L. (1980) *The statistical analysis of failure time data* Vol. 2nd. of *Wiley Series in probability and statistics* Wiley Hoboken, New Jersey.

Khan, S. and Tamer, E. (2010) Irregular Identification, Support Conditions, and Inverse Weight Estimation *Econometrica* **78**(6), 2021–2042.

Lee, S. and Lewbel, A. (2013) Nonparametric identification of accelerated failure time competing risks models *Econometric Theory* **29**(05), 905–919.

Lewbel, A., Shen, W. and Zhang, H. M. (2000) Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables *Journal of Econometrics* **97**(530), 145–177.

Maddala, G. S. (1986) *Limited-dependent and qualitative variables in econometrics.* Cambridge University Press Cambridge.

Manski, C. F. (1975) Maximum score estimation of the stochastic utility model of choice *Journal of Econometrics* **3**(3), 205–228.

Matějka, F. and McKay, A. (2015) Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model *American Economic Review* **105**(1), 272–298.

Matzkin, R. L. (1993) Nonparametric identification and estimation of polychotomous choice models *Journal of Econometrics* **58**(1-2), 137–168.

McFadden, D. (1981) Econometric Models of Probabilistic Choice *in* C. Manski and D. McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications* MIT Press Cambridge, MA, USA pp. 198–272.

McFadden, D. L. (1974) Conditional logit analysis of qualitative choice behavior *Frontiers in Econometrics* Academic Press New York pp. 105–142.