The Econometrics Journal

Econometrics Journal (2020), volume **00**, pp. 1–13.
doi: 10.1093/ectj/utaa016

Identification of a class of index models: A topological approach

MOGENS FOSGERAU[*] AND DENNIS KRISTENSEN[†]

[*]*Department of Economics, University of Copenhagen, Øster Farimagsgade 5, 1353 København K, Denmark.*
Email: mogens.fosgerau@econ.ku.dk

[†]*Department of Economics, University College London, Gower Street, London, WC1E 6BT, UK.*
Email: d.kristensen@ucl.ac.uk

First version received: 30 December 2019; final version accepted: 17 April 2020.

Summary: We establish nonparametric identification in a class of so-called index models by using a novel approach that relies on general topological results. Our proof strategy requires substantially weaker conditions on the functions and distributions characterising the model than those required by existing strategies; in particular, it does not require any large-support conditions on the regressors of our model. We apply the general identification result to additive random utility and competing risk models.

Keywords: *Competing risks*, *discrete choice*, *index model*, *nonparametric identification*.

JEL codes:: *C14*, *C35*, *C36*, *C41*.

1. INTRODUCTION

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

$$\Pi(w, x, z) = \Lambda(a(w, x), z), \tag{1.1}$$

where

$$a(w, x) = g(w) + h(x) \tag{1.2}$$

is a vector of additively separable index functions, and $\Lambda : \mathbb{R}^J \times \mathbb{R}^{d_Z} \mapsto \mathbb{R}^J$, $g : \mathbb{R}^{d_W} \mapsto \mathbb{R}^J$, 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, whereas $z \in \mathbb{R}^{d_Z}$ corresponds to values of a set of control variables, Z. We take as a high-level assumption that we know (have observed from data) the function $\Pi(w, x, z)$, for (w, x, z) in the support of (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. We refer to this class of models as index models because 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. Most existing results on identification within this class of models require availability of a set of 'special' continuously distributed regressors; identification is then achieved by sending each of these special regressors

off to the boundary of their support. Estimators based on such a 'thin-set identification' argument were analyzed by Khan and Tamer (2010), who showed that they tend to be irregularly behaved with slow convergence rates. In contrast, we achieve identification as long as the random index $a(W, X)$ exhibits sufficient, but potentially bounded, variation. We expect this to translate into better-behaved estimators.

Second, we impose weak conditions on the functions of interest and distributions of the random variables (W, X, Z). We do not require continuity or differentiability of the functions entering the model in order to show identification, whereas most existing results at a minimum require these to be differentiable. Similarly, we require only $g(W)$ to have continuous support, whereas (X, Z) can be either discrete or continuous, or a mix of the two, as long as their supports satisfy certain conditions. Thus, our results cover models with thresholds and kinks in Λ, g, and h, which existing results cannot handle. In the case of discrete choice models, such features may occur if the decision-maker optimises subject to constraints; see, e.g., Cantillo and de Dios Ortúzar (2006). These models have traditionally been formulated in a parametric fashion; our theory demonstrates how they can be identified without parametric constraints. There is a growing literature on nonparametric estimation with unknown thresholds and kinks, which we conjecture can be used in our setting to translate our identification result into actual estimators; see, e.g., Chiou et al. (2018).

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. In comparison, most other papers that allow for control variables show identification at a fixed arbitrary value of Z, in which case variation in Z is unnecessary for 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 because they can be used for identification in other settings. The two key elements of our approach are the notions of relative identification and connected sets. Below, we state our formal definition of the former:

DEFINITION 1. *A function h is said to be relatively identified on a given set $\mathcal{X}$ if identification of $h(x^*)$ at some point $x^* \in \mathcal{X}$ implies that $h(x)$ is also identified at all other $x \in \mathcal{X}$.*

Next, recall the topological notion of *connnectedness*: A connected set cannot be contained in the union of two non-empty disjoint open sets while having non-empty intersection with both. In particular, it is not possible to split a connected open set into disjoint open subsets.

Our identification strategy then proceeds in three steps, where we here initially suppress the presence of Z for simplicity. First, we decompose the support of X into suitable subsets and achieve relative identification on each of these. This is done via two features of our model: (a) For a given x, we are able to identify the relative variation in $\Lambda(a)$, with $a = h(x) + g(w)$, through the observed variation in $\Pi(w, x)$ with respect to w through the known function $g(w)$. (b) By injectivity of Λ, we are then able to identify the relative value of a, which in turn yields the relative value of $h(x) = a - g(w)$ on suitably chosen subsets of the support of X. Second, we achieve global identification on the union of these subsets by using the second main ingredient of our proof strategy, connectedness. We will require the support of $a(W, X)$ to be connected, which is used to extend relative local identification to global identification. Third, reintroducing Z, we again rely on the supports of $X|Z = z$ to be suitably connected across different values of z in the support of Z to enlarge the identification region further.

Like us, Berry and Haile (2018) and Evdokimov (2010), among others, relied on connectness to achieve global identification, but in those papers, the restriction was 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 additive versions of multiple discrete choice 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, Matzkin (1993) is probably the most closely related to our setting, but the assumptions made and identification strategy pursued in that paper are very different from ours. Our and her sets of assumptions are not clearly ranked, with some of our assumptions being stronger and others weaker than hers. One key feature of her proof strategy was the introduction of assumptions that ensure the multinomial model can be converted into a binary choice problem followed by a thin-set identification argument. More recently, Allen and Rehbeck (2019) provided conditions under which one can identify how regressors alter the desirability of alternatives by using only average demands. Their conditions are weaker than ours, but on the other hand they were able to identify only certain features of the model, not the underlying data-generating structure.

There is also a nascent literature on nonparametric identification of so-called BLP models (Berry, Levinsohn, and Pakes, 1995) as used in industrial organisation; see, for example, Berry and Haile (2018) and Chiappori et al. (2018). The setting of the BLP model is somewhat different, though, because there the choice probabilities are treated as observed variables that depend on unobserved product characteristics, which have to be controlled for. This leads to a different identification problem from 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) achieved identification by assuming the index [in our notation $a(W, X)$] has support on $(0, \infty)^J$ and then achieved identification of a given component of the index by letting the other components go to zero, and so their result falls in the thin-set identification category. Abbring and van den Berg (2003) weakened this assumption substantially for the class of mixed proportional-hazard models, a subclass of competing risk models. Lee and Lewbel (2013) provided a high-level assumption for identification of the general model involving a rank condition of an integral operator. Primitive conditions for this to hold are not known. Honoré and Lleras-Muney (2006) derived bounds for the functions of interest when only discrete covariates are available. We complement these studies by showing identification in the general competing risk model under primitive conditions that allow for the presence of discrete covariates but at the same time impose more structure on the index; cf. Equation (2.2).

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 Equation (1.1). We present our general framework in Section 3, as well as the assumptions we will work under, and provide our identification results in Section 4. Section 5 applies our general result to the two examples, and Section 6 concludes.

2. TWO MOTIVATING EXAMPLES

The model (1.1) comprises a range of models that are met in economics. We here present two classes of models that fall within our framework. We will return to these two classes of models in Section 5, where we apply our general identification result to each of them.

2.1. Discrete choice models

We here first demonstrate that the class of additive random utility models (ARUM) can be mapped into (1.1). On the basis of existing results in the literature, this in turn implies that our results also apply to a broad class of rational inattention discrete choice models (Fosgerau et al., 2019) and an even wider class of perturbed utility models.

2.1.1. Additive random utility. Consider an agent choosing between $J+1$ alternatives, each carrying an associated indirect utility of the form

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

where (W, X) is a set of observed covariates and $\varepsilon = (\varepsilon_0, \varepsilon_1, ..., \varepsilon_J)$ is unobserved. This model was initially proposed by McFadden (1973) 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 normalisation on the 'outside' option $j = 0$: $a_0(w, x) = 0$.

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 that (W, X) are independent of ε conditional on Z. In addition to (W, X, Z), the researcher also observes the utility-maximising choice, $D = \arg\max_{j\in\{0,1,...,J\}} U_j$. Thus, the conditional choice probabilities (CCPs),

$$\Pi_j(w, x, z) := P(D = j | (W, X, Z) = (w, x, z)), \quad j = 0, 1, ..., J, \tag{2.2}$$

are identified in the population. We collect these in the vector-valued function $\Pi(w, x, z) = \left\{\Pi_j(w, x, z) : j = 1, ..., J\right\} \in \mathbb{R}^J$, where we leave out the CCP of the outside option. It now follows from standard results in the literature that $\Pi(w, x, z)$ can be written in the form of (1.1), with Λ being the gradient of the so-called surplus function; see Section 5 for further details.

Our identification result requires the researcher to group the observed covariates into two sets. The first set, denoted W, contains the 'special' regressors that enter the index a through a known function $g(W)$, as specified by the researcher; cf. Equation (2.2). The second set, denoted X, then enters a through $h(X)$, which is left unspecified.

The choices of W and $g(W)$ are application specific and should be guided by two considerations. First, $g(W)$ needs to exhibit sufficient continuous variation on $\mathbb{R}^J$ because this is a key requirement for our identification result to go through. Second, because $g_j(W)$ affects the utility of the jth alternative positively by definition, it should be specified accordingly.

As an example of this joint modelling and identification strategy, let us consider the problem of estimating willingness to pay for different goods, a common problem in various applied fields of economics (e.g., Fosgerau, 2006; Bontemps and Nauges, 2016). In this setting, choosing g to be $g_j(W_j) = -\ln W_j$, where W_j is the price of alternative $j, j = 1, ..., J$, transforms a positive price vector into a vector that can in principle attain values in all of $\mathbb{R}^J$. With this choice, $h_j(X) + \varepsilon_j$ captures the log willingness to pay for good j, where X contains characteristics of the agent and other characteristics of the different alternatives. Prices generally exhibit continuous variation and so satisfy the first of the two aforementioned requirements. This example assumes the availability of

alternative-specific regressors, W_1, ..., W_J. However, our identification result may still be applied if this is not true. In this case, the researcher needs to construct alternative-specific regressors $g_1(W)$, ..., $g_J(W)$ from a set of underlying covariates W.

Our assumption of $g(W)$'s being known has antecedents in the literature on identification in discrete choice models. For example, in the context of binary choice ($J = 1$), Lewbel et al. (2000) also assumed the presence of a 'special' regressor, in our notation W, that entered the utility of alternative 1 in a known fashion. However, that paper furthermore restricted $h(x)$ to be linear, $h(x) = \beta x$, and, importantly, identification of β was achieved through variation of $g(W)$ on the boundary of its support. Our identification result does not rely on any such argument.

Our framework also includes the so-called rational inattention discrete choice model. Fosgerau et al. (2019) showed 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 generalises the finding of Matějka and McKay (2015), who showed that the multinomial logit model has a foundation as a rational inattention model. Thus, 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 and McFadden, 2012; Fudenberg et al., 2015; Allen and Rehbeck, 2019) is another generalisation of the class of ARUM. As shown by Hofbauer and Sandholm (2002), the CCPs of an ARUM can be represented as the solution to a maximisation problem in which an agent chooses the vector of CCPs to maximise a function that consists of a linear term and a concave term. Here, we present an extended version that includes controls affecting the concave term, i.e.,

$$\Lambda(a, z) = \arg\max_{q \in \Delta} \left\{ a^{\intercal} q + \Omega(q|z) \right\}, \tag{2.3}$$

where $a \in \mathbb{R}^{J+1}$ is a vector of utility indices, $\Delta = \{q \in \mathbb{R}_{+}^{J+1} : \sum_{j=0}^{J} q_j = 1\}$ is the unit simplex, and $\Omega(\cdot\,|z)$ is a concave function for each $z \in \mathcal{Z}$. The perturbed utility model includes ARUM as a special case, while allowing an individual to have strict preference for randomisation rather than to choose a vertex of the probability simplex. As noted by Allen and Rehbeck (2019), observing only realisations of lotteries across choice options is sufficient for identification, which requires only the vector of CCPs, $\Pi(w, x, z)$. We show in Section 5 that the implied CCPs satisfy (1.1).

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 observed covariates (X, W, Z). Assume that the jth failure time satisfies

$$\ln T_j = a_j(W, X) - \varepsilon_j, \tag{2.4}$$

for some function $a_j(w, x)$, $j = 1, ..., J$. The model may then be termed a multivariate generalised accelerated failure time model (Kalbfleisch and Prentice, 1980; Fosgerau et al., 2013). The econometrician has knowledge of

$$\Pi_j(w, x, z) := E[\ln Y | W = w, X = x, Z = z] \cdot P(D = j | W = w, X = x, Z = z), \tag{2.5}$$

for $j = 1, ..., J$, where Z is used to control for potential dependence between (W, X) and ε. We collect the unobservables in $\varepsilon = (\varepsilon_1, ..., \varepsilon_J)$ and again require them to be conditionally independent of (X, W), in which case, as shown in Section 5, Π defined above again satisfies Equation (1.1).

Typical applications of the above model are in the modelling of (un)employment spells in which an exit from the unemployment register can be the result of finding a full- or part-time job in a different sector or another change of status. Thus, in this setting, $j = 1, ...J$ indexes the different exits (types of non-unemployment), and (W, X) contain both variables characterising the types of employment (such as salary in a given type/sector of employment) and individual-specific controls (such as age and marital status). Similar to discrete choice models, we would then need to construct $g(W)$ to capture risk-specific characteristics with continuous variation and then include all other covariates in X. Most empirical applications assume a parametric structure for the index, e.g., $a(W, X) = \alpha W + \beta X$. In this setting, requiring g to be known effectively assumes fixing $\alpha \in \mathbb{R}^{J \times d_Y}$. At the same time, we impose very weak restrictions on the distributional features of the regressors X and how they enter the index $a(W, X)$.

3. GENERAL FRAMEWORK

We now return to the general model given in Equations (1.1) and (2.2), where $g : \mathbb{R}^{d_Y} \to \mathbb{R}^J$ is assumed to be a known function, whereas $h : \mathbb{R}^{d_X} \to \mathbb{R}^J$ and $\Lambda : \mathbb{R}^J \times \mathbb{R}^{d_Z} \to \mathbb{R}^J$ are unknown functions. In the following, let int$\mathcal{A}$ denote the interior of a given set $\mathcal{A}$, and let supp(Y) denote the support of a given random variable Y. We then 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; cf. the examples in the previous section.

The covariates contained in $g(W)$ play a special role in our approach in that we need sufficient continuous variation in these to achieve identification. First note that $\dim g\,(W) = J$. Thus, sufficient continuous variation of $g(W)$, which is known to us, permits us to identify the relative variation of $\Lambda(a, z)$ with respect to a. Formally, for any given pair $(x, z) \in \text{supp}(X, Z)$, define

$$\mathcal{G}\,(x, z) = \text{int supp}\,(g\,(W)\,|X = x, Z = z)\,, \quad \mathcal{X}\,(z) = \text{supp}\,(X|Z = z)\,. \tag{3.1}$$

We will then throughout implicitly require that some of the open sets $\mathcal{G}\,(x, z)$, $(x, z) \in \text{supp}(X, Z)$ are non-empty, and then we will achieve identification at the values of x for which this is true. A sufficient condition for a given $\mathcal{G}(x, z)$ to be non-empty is that the distribution of $W|X = x$, $Z = z$ is continuous and that g maps open sets into open sets; however, this is not required, and $g(W)$ may contain discrete components as long as they fall within the support of the continuous component. Our identification result still applies if any values of a discrete component fall outside the continuous support, but it excludes these values. This also rules out that some components of W are included in Z because in this case $\mathcal{G}\,(x, z) = \emptyset$. At the same time, however, (X, W) can depend on Z; we just need sufficient variation in (X, W) conditional on Z. Moreover, no continuity restrictions are imposed on the distribution of (X, Z), which may be completely discrete. Finally, we would like to stress that we do not impose any large-support restrictions on $g(W)$, which is in contrast to most existing results in the literature, as discussed in the introduction. If, for example, $\mathcal{G}\,(x, z) = \mathbb{R}^J$ for all x, then our result demonstrates that $h(x)$ is identified on all of supp(X), but it is not necessary; identification on all of supp(X) can be achieved without such a full-support condition.

Next, let

$$\mathcal{M}(x,z) = \mathcal{G}(x,z) \times \{x\}, \;\; \mathcal{A}(x,z) = a(\mathcal{M}(x,z)) = \mathcal{G}(x,z) + \{h(x)\},$$

denote the support of $(W, X)|(X, Z) = (x, z)$ and $a(W, X)|(X, Z) = (x, z)$, respectively, and let

$$\mathcal{M}(z) = \cup_{x \in \mathcal{X}(z)} \mathcal{M}(x,z), \;\; \mathcal{A}(z) = \cup_{x \in \mathcal{X}(z)} \mathcal{A}(x,z) = a(\mathcal{M}(z)), \tag{3.2}$$

denote the supports of the same random variables but now conditioning only on $Z = z$. Finally, for some set $\mathcal{Z}_0 \subseteq \text{supp}(Z)$ chosen according to certain assumptions stated below, let

$$\mathcal{A}_0 = \cup_{z \in \mathcal{Z}_0} \mathcal{A}(z), \;\; \mathcal{X}_0 = \cup_{z \in \mathcal{Z}_0} \mathcal{X}(z)$$

be the supports of $a(W, X)$ and X conditional on $Z \in \mathcal{Z}_0$, respectively. We will then show identification of $h(x)$ and $\Lambda(a, z)$ for $x \in \mathcal{X}_0$, $a \in \mathcal{A}_0$, and $z \in \mathcal{Z}_0$. Specifically, $\mathcal{Z}_0$ will be constructed according to certain properties of the underlying covariates and the functions of interest. Observe the dependence of $\mathcal{M}_0$ and $\mathcal{A}_0$ on the set $\mathcal{Z}_0$. To achieve 'maximal' identification, we would ideally like to choose $\mathcal{Z}_0 = \text{supp}(Z)$. However, we potentially have to restrict $\mathcal{Z}_0$. First, we require $a \mapsto \Lambda(a, z)$ to satisfy the following condition for all $z \in \mathcal{Z}_0$:

ASSUMPTION 3.1. *For any $z \in \mathcal{Z}_0$, $a \mapsto \Lambda(a, z)$ is injective on $\mathcal{A}(z)$ as defined in (3.2).*

By asking for $\Lambda(a, z)$ to be injective, we can identify the relative variation in $a(w, x)$ through the observed variation in $\Pi(w, x, z)$. In a given application, Assumption 3.1 may not hold for all $z \in \text{supp}(Z)$, in which case we need to 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 may achieve identification on the whole support.

Because of the structure of $a(w, x)$, it follows from the definition of $\mathcal{G}(x, z)$ that $\mathcal{A}(x, z)$ and thereby also $\mathcal{A}(z)$ and $\mathcal{A}_0$ are open sets. We add to this by also requiring $\mathcal{A}(z)$ to be connected for all $z \in \mathcal{Z}_0$. An open set $\mathcal{A}$ is connected if $\mathcal{A} = \mathcal{O}_1 \cup \mathcal{O}_2$ implies that $\mathcal{O}_1 \cap \mathcal{O}_2 \neq \varnothing$ whenever $\mathcal{O}_1$ and $\mathcal{O}_2$ are non-empty open sets. Thus, an open connected set cannot be separated into two non-empty disjoint open sets. We then impose:

ASSUMPTION 3.2. *$\mathcal{A}(z)$ is connected for all $z \in \mathcal{Z}_0$.*

Assumption 3.2 allows us to go from local identification at a given point $x \in \mathcal{X}(z)$ to relative identification on all of $\mathcal{X}(z)$, $z \in \mathcal{Z}_0$ via the image of $a(x, w)$. The assumption imposes restrictions on the support of the random variable $a(X, W)$ instead of (X, W) themselves. This is done to impose minimal restrictions on the distribution of X and the smoothness of h. Recall that W is assumed to contain a continuous component. Thus, Assumption 3.2 includes, for example, the case of X being unbounded and discrete, or the case of X being continuous while $h(X)$ is discontinuous everywhere. Assumption 3.2 is not verifiable from data, but the same holds for smoothness conditions that are regularly imposed in existing identification results. If we are willing to entertain certain smoothness conditions, such as the inverse of $\Lambda(a, z)$ being continuous with respect to a, then the assumption is implied by connectedness of $\Pi(\mathcal{M}(z)|z) = \Lambda(\mathcal{A}(z), z)$—this latter property being verifiable. Similarly, if we restrict X and h to both be continuous, it will be implied by connectedness of $\mathcal{M}(z)$.

Once we have achieved relative identification on each $\mathcal{X}(z)$, $z \in \mathcal{Z}_0$, global identification is then reached through the following assumption:

ASSUMPTION 3.3. *If $\mathcal{Z}_1 \cup \mathcal{Z}_2 = \mathcal{Z}_0$, $\mathcal{Z}_1, \mathcal{Z}_2 \neq \emptyset$, then $(\cup_{z \in \mathcal{Z}_1} \mathcal{M}(z) \cap (\cup_{z \in \mathcal{Z}_2} \mathcal{M}(z) \neq \emptyset$.*

This is used to paste together the relatively identified sets $\mathcal{X}(z)$ across z. Again, this assumption does not require X and/or h to be continuous, only that the sets $\text{supp}(W, X|Z = z)$, $z \in \mathcal{Z}_0$ overlap. Finally, the following normalisation on the function h gives us identification on $\mathcal{X}(z_0)$:

ASSUMPTION 3.4. *There exist $z_0 \in \mathcal{Z}_0$ and $(w_0, x_0) \in \mathcal{M}(z_0)$ so that $h(x_0) = 0$.*

Such a normalisation is needed to identify the level of h because 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. MAIN RESULT

As explained earlier, we shall make use of the notion of relative identification in our proof of identification. As a first step, we show relative identification on any two overlapping images of a; this is achieved through injectivity of $\Lambda(a, z)$, which allows us to map the overlapping images into overlapping images of Π.

LEMMA 4.1. *Suppose that Assumption 3.1 holds and that $h(x^*)$ is identified at $x^* \in \mathcal{X}(z)$ for some $z \in \mathcal{Z}_0$. Then the set $\mathcal{X}^*(z) := \{x \in \mathcal{X}(z) | \mathcal{A}(x^*, z) \cap \mathcal{A}(x, z) \neq \emptyset\}$ is identified, and $h(x)$ is identified on $\mathcal{X}^*(z)$.*

Proof. By definition, $\mathcal{A}(x^*, z) \cap \mathcal{A}(x, z) \neq \emptyset$ if and only if there exist w^* and w so that $g(w^*) \in \mathcal{G}(x^*, z)$, $g(w) \in \mathcal{G}(x, z)$, and $a(w^*, x^*) = a(w, x)$. Noting that $\Lambda(a|z)$ is injective by Assumption 3.1, we see that the last equality is equivalent to $\Lambda(a(w^*, x^*), z) = \Lambda(a(w, x), z)$, which we recognise as

$$\Pi\left(w^*, x^*, z\right) = \Pi(w, x, z), \tag{4.1}$$

where Π is known to us. Thus, $\mathcal{X}^*(x, z)$ is identified as the set of solutions x to (4.1) as we vary (w^*, w). Next, for any given $x \in \mathcal{X}^*(x, z)$, let w^* and w be the corresponding values for which (4.1) holds. Because these are known, the value $a(w^*, x^*) = g(w^*) + h(x^*)$ is also known to us. This in turn implies that $h(x) = a(w, x) - g(w) = a(w^*, x^*) - g(w)$ is identified. □

We then use this lemma in conjunction with the connectedness of $\mathcal{A}(z)$ to show relative identification on each of the sets $\mathcal{X}(z)$:

LEMMA 4.2. *Suppose that Assumptions 3.1 and 3.2 hold. Then, for all $z \in \mathcal{Z}_0$, $h(x)$ is relatively identified on $\mathcal{X}(z)$ as defined in Equation (3.1).*

Proof. Let $x^* \in \mathcal{X}(z)$ be given, and suppose we know the value of $h(x^*)$. Let $\mathcal{X}^*(z) \subseteq \mathcal{X}(z)$ be the set on which $h(x)$ is identified, and let $\mathcal{A}^*(z) = \cup_{x \in \mathcal{X}^*(z)} \mathcal{A}(x, z)$ be the corresponding values of $a(w, x)$. By assumption, $x^* \in \mathcal{X}^*(z)$, and so the identified set is non-empty. This in turn implies that $\mathcal{A}^*(z)$ is non-empty and open. Now, seeking a contradiction, suppose that $\mathcal{X}^{**}(z) := \mathcal{X}(z) \setminus \mathcal{X}^*(z) \neq \varnothing$. Then define $\mathcal{A}^{**}(z) = \cup_{x \in \mathcal{X}^{**}(z)} \mathcal{A}(x, z)$, which is also open and non-empty. Because $\mathcal{A}^*(z) \cup \mathcal{A}^{**}(z) = \mathcal{A}(z)$, which is connected according to Assumption 3.2, there must exist $x \in \mathcal{X}^*(z)$ and $x' \in \mathcal{X}^{**}(z)$ so that $\mathcal{A}(x, z) \cap \mathcal{A}\left(x', z\right) \neq \varnothing$. Lemma 4.1 then implies that x' and $h(x')$ are also identified, which is a contradiction. □

Finally, the 'connectedness' of $\cup_{z \in \mathcal{Z}_0} \mathcal{M}(z)$ as stated in Assumption 3.3, together with the normalisation in Assumption 3.4, give us global identification:

THEOREM 4.1. *Under Assumptions 3.1 through 3.4, h(x) is identified on* $\mathcal{X}_0 = \cup_{z\in\mathcal{Z}_0}\mathcal{X}(z)$.

Proof. Let $\mathcal{X}^*$ be the identified set. By Lemma 4.2, $\mathcal{X}^* = \cup_{z\in\mathcal{Z}^*}\mathcal{X}(z)$ for some $\mathcal{Z}^* \subseteq \mathcal{Z}_0$. By Assumption 4, $z_0 \in \mathcal{Z}^*$, and so the set is non-empty. Seeking a contradiction, suppose that $\mathcal{Z}^{**} := \mathcal{Z}_0 \backslash \mathcal{Z}^* \neq \varnothing$. By definition $\mathcal{Z}^* \cup \mathcal{Z}^{**} = \mathcal{Z}_0$, and so $\{\cup_{z\in\mathcal{Z}^*}\mathcal{M}(z)\} \cap \{\cup_{z\in\mathcal{Z}^{**}}\mathcal{M}(z)\} \neq \emptyset$ by Assumption 3.3. This implies that there exist $z^* \in \mathcal{Z}^*$ and $z^{**} \in \mathcal{Z}^{**}$ so that $\mathcal{M}(z^*) \cap \mathcal{M}(z^{**}) \neq \emptyset$, which in turn implies that there exists $x^* \in \mathcal{X}(z^*) \cap \mathcal{X}(z^{**})$ for which $h(x^*)$ is identified. Then Lemma 4.2 implies that $h(x)$ is identified on all of $\mathcal{X}(z^{**})$, which is a contradiction. □

Once we have identified h, we can also identify Λ:

THEOREM 4.2. *Under Assumptions 3.1 through 3.4,* $\Lambda(a, z)$ *is identified on* $\{(a, z)\,|a \in \mathcal{A}(z), z \in \mathcal{Z}_0\}$.

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 $a = a(w, x)$. Because $h(\cdot)$ and thereby also $a(\cdot, \cdot)$ are identified, the pair (w, x) is known. Then we also know $\Pi(w, x, z)$, and so $\Lambda(a, z) = \Pi(w, x, z)$ is uniquely identified. □

5. APPLICATIONS

This section applies the general result to the two main examples of Section 2, the ARUM and the competing risk model, and compares our identification results for these two models with existing ones found in the literature. In both examples, we impose the following conditional independence restriction on the error term:

ASSUMPTION 5.1. *(i)* ε *is conditionally independent of (X, W),* $F_{\varepsilon|(W,X,Z)}(\cdot\,|\,\cdot,\cdot,z) = F_{\varepsilon|Z}(\cdot\,|z)$ *for all* $z \in \mathcal{Z}_0$ *for some* $\mathcal{Z}_0 \subseteq \text{supp}(Z)$*; (ii)* $F_{\varepsilon|Z}(\cdot\,|z)$ *has a conditional density with full support for all* $z \in \mathcal{Z}_0$.

We demonstrate in the next two subsections that part (i) implies that Π, as defined in Equations (2.1) and 2.4, respectively, can be written on the forms (1.1) and (2.2) for all $z \in \mathcal{Z}_0$, whereas part (ii) ensures that the model-specific $\Lambda(a, z)$ is injective with respect to a for all $z \in \mathcal{Z}_0$.

5.1. ARUM

Define the surplus function,

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

for any given $(a_0, a_1, ..., a_J) \in \mathbb{R}^{J+1}$, where the second equality uses Equation (2.1.1) and Assumption 5.1(i). The Williams-Daly-Zachary theorem (McFadden, 1981) then implies that the CCPs, as defined in (2.1), can be written on the forms (1.1) and (2.2) with Λ defined as the gradient of the surplus function,

$$\Lambda(a, z) := \left.\frac{\partial G(a, z)}{\partial a}\right|_{a_0=0}.$$

We conclude:

COROLLARY 5.1. *Any ARUM on the form (2.1.1) that satisfies Assumptions 3.1 through 3.4 and 5.1(i) is identified.*

Next, we discuss each of Assumptions 3.1 through 3.4 in the context of ARUM and how these compare with existing ones found in the literature on identification of ARUM.

First, Assumption 3.1, injectivity of $\Lambda(\cdot, z)$ for each z, is implied by Assumption 5.1(ii); cf. Hofbauer and Sandholm (2002, Theorem 2.1). However, Assumption 5.1(ii) is not necessary for injectivity to hold. A simple example is the binomial model, where the probability for alternative zero is the cumulative distribution of ε_1. If the distribution includes point masses, then ties can occur, but this does not destroy injectivity. This is true for any tie-breaking rule. More generally, if the subdifferential of the surplus function is strictly cyclically monotone (Rockafellar, 1970), which does not require the existence of a density, then the utility-maximising choice probabilities under any tie-breaking rule are injective (Sørensen and Fosgerau, 2020).

Assumptions 3.2 and 3.3 impose restrictions on the joint variation of $(g(W), X)$. For Assumption 3.2 to hold, we need to identify J regressors, $g(W)$, that exhibit enough joint continuous variation so their joint support, conditional on (X, Z), has a non-empty interior on $\mathbb{R}^J$. One instance in which this can be achieved is if we have observed alternative-specific characteristics. In the case of demand modelling, one such choice would be a (transformation) of the (relative) prices of the different alternatives, while X contains all remaining regressors, possibly including other alternative-specific covariates. In this case, to control for potential endogeneity of prices, we could then include cost shifters in Z. Prices tend to exhibit continuous variation, and Assumption 3.2 would be likely to hold. Assumption 3.3 requires other observed product characteristics and the agent's observed characteristics to exhibit sufficient variation conditional on the controls in Z so that these have overlapping support across different values of Z.

As already mentioned in the introduction, there are few fully nonparametric identification results for ARUM. To our knowledge, the only results comparable to ours are found in Matzkin (1993). Her results also require the presence of alternative-specific regressors but impose stronger conditions on these and other covariates. Moreover, her setup did not include any control variables. On the other hand, she did not necessarily require that $a(W, X)$ be additive, which we assume throughout. Theorem 1 of Matzkin (1993) does allow for dependence between (W, X) and ε, but in this case, she required the observed component of the utilities to be identical across alternatives and strictly increasing in one of the arguments. In our notation, this requires $a_j(W, X)$, $j = 1, ..., J$ to all be identical. We do not impose any such constraints. Her Theorem 2 requires full independence between (W, X) and ε but, on the other hand, imposes fewer restrictions on $a(W, X)$ than we do. In both cases, she identified Λ by letting different components of W diverge to $+\infty$, which is an example of thin-set identification, discussed earlier.

5.2. *Perturbed discrete choice*

We here demonstrate that the CCPs for the perturbed discrete choice model again can be expressed on the forms (1.1) and (2.2), with Λ defined in (2.2) as being injective. This is done under the following restrictions: First, to rule out zero demands, the norm of the gradient $\nabla_q \Omega(q|z)$ has to approach infinity as q approaches the boundary of the unit simplex. Second, $\Omega(q|z)$ is differentiable.[1] Third, we normalise the outside option so that $g_0(w) = h_0(x) = 0$. Under these three restrictions, for each value of the control z, the demand solves the first-order condition for an interior solution,

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

[1] Note we do not require a Hessian.

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

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

which implies that $a_1 = a_2$ as required.

5.3. *Competing risk*

Define

$$\Lambda\left(a, z\right) := G\left(a, z\right) \cdot \frac{\partial G\left(a, z\right)}{\partial a}, \tag{5.1}$$

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

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

where the second equality uses Equation (2.3) and Assumption 5.1(i). Williams-Daly-Zacchary theorem (McFadden, 1981) then implies that Π, now defined by (2.4), can be written on the forms (1.1) and (2.2). Injectivity of $\Lambda(a, z)$, as given in Equation (5.1), is obtained by recycling the arguments of the previous subsection, except that no normalisation of one of the causes of failure is required because the level $G(a, z)$ is included.

COROLLARY 5.2. *Any competing risk model on the form (2.3) that satisfies Assumptions 3.1 and 3.4 and 5.1(i) is identified.*

Given that the competing risk model and the ARUM share a similar structure, the discussion of the remaining assumptions carries over to the present setting with obvious modifications.

Compared with existing results (Heckman and Honoré, 1989; Lee and Lewbel, 2013), we impose stronger conditions on the index $a(W, X)$ because we require it to be additive and with $g(W)$ known. On the other hand, Heckman and Honoré (1989) required $a(W, X)$ to go to zero as W diverges and thus relied on a thin-set identification argument, while Lee and Lewbel (2013) relied on a high-level functional rank condition. It is unclear which primitive conditions suffice for this rank condition to hold. Finally, Honoré and Lleras-Muney (2006) restricted themselves to the case of purely discrete regressors and were only able to derive bounds for objects of interest. We achieve point identification as long as there is some continuous variation in W, while X can be completely discrete.

6. CONCLUSION

We have established an identification result for a wide class of index models on the basis of general topological arguments. Three key features of our argument are that smoothness of the model is not required; no large-support condition is imposed on the regressors; and control variables may contribute to achieving identification. We leave the development of nonparametric estimators of the identified components for future research.

ACKNOWLEDGEMENT

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).

REFERENCES

Abbring, J. H. and G. J. van den Berg (2003). The identifiability of the mixed proportional hazards competing risks model. *Journal of the Royal Statistical Society 65*, 701–10.

Allen, R. and J. Rehbeck (2019). Identification with additively separable heterogeneity. *Econometrica 87*, 1021–54.

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

Berry, S. T. and P. A. Haile (2018). Identification of nonparametric simultaneous equations models with a residual index structure. *Econometrica 86*, 289–315.

Berry, S. T., J. Levinsohn and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica 63*, 841–90.

Bontemps, C. and C. Nauges (2016). The impact of perceptions in averting-decision models: an application of the special regressor method to drinking water choices. *American Journal of Agricultural Economics 98*, 297–313.

Cantillo, V. and J. de Dios Ortúzar (2006). Implications of thresholds in discrete choice modelling. *Transport Reviews 26*, 667–91.

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

Chiou, Y. Y., M. Y. Chen and J. e. Chen (2018). Nonparametric regression with multiple thresholds: Estimation and inference. *Journal of Econometrics 206*, 472–514.

Evdokimov, K. (2010). Identification and estimation of a nonparametric panel data model with unobserved heterogeneity.

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

Fosgerau, M., D. McFadden and M. Bierlaire (2013). Choice probability generating functions. *Journal of Choice Modelling 8*, 1–8.

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

Fosgerau, M., E. Melo, A. de Palma and M. Shum (2019). Discrete choice and rational inattention: A general equivalence result. *SSRN Electronic Journal*, doi.org/10.1111/iere.12469.

Fudenberg, D., R. Iijima and T. Strzalecki (2015). Stochastic choice and revealed perturbed utility. *Econometrica 83*, 2371–2409.

Heckman, J. J. and B. E. Honoré (1989). The identifiability of the competing risks model. *Biometrika 76*, 325–30.

Hofbauer, J. and W. H. Sandholm (2002). On the global convergence of stochastic fictitious play. *Econometrica 70*, 2265–94.

Honoré, B. E. and A. Lleras-Muney (2006). Bounds in competing risks models and the war on cancer. *Econometrica 74*, 1675–98.

Kalbfleisch, J. D. and R. L. Prentice (1980). *The Statistical Analysis of Failure Time Data, Vol. 2nd. of Wiley Series in Probability and Statistics*, Chapter 8, Competing Risks and Multistate Models. Hoboken, NJ: Wiley.

Khan, S. and E. Tamer (2010). Irregular identification, support conditions, and inverse weight estimation. *Econometrica 78*, 2021–42.

Lee, S. and A. Lewbel (2013). Nonparametric identification of accelerated failure time competing risks models. *Econometric Theory 29*, 905–19.

Lewbel, A., W. Shen and H. M. Zhang (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics 97*(530), 145–77.

Maddala, G. S. (1986). *Limited-Dependent and Qualitative Variables in Econometrics*, Chapter 5, Multivariate qualitative variables, Cambridge, UK: Cambridge University Press.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics 3*, 205–28.

Matějka, F. and A. McKay (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review 105*(1), 272–98.

Matzkin, R. L. (1993). Nonparametric identification and estimation of polychotomous choice models. *Journal of Econometrics 58*(1–2), 137–68.

McFadden, D. (1981). Econometric models of probabilistic choice, In Manski, C. and D. McFadden (Eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp. 198–272. Cambridge, MA, USA: MIT Press.

McFadden, D. L. (1973). Conditional logit analysis of qualitative choice behavior, In *Frontiers in Econometrics*, pp. 105–42. New York, NY: Academic Press.

Rockafellar, R. T. (1970). *Convex Analysis*, Chapter 12, Conjugates of Convex Functions, Princeton, NJ: Princeton University Press.

Sørensen, J. R.-V. and M. Fosgerau (2020). How McFadden met Rockafellar and learnt to do more with less, Discussion Papers 20-01, University of Copenhagen, Department of Economics.