Structural modeling of simultaneous discrete choice

Andrew Chesher
Adam M. Rosen

Structural modeling of simultaneous discrete choice.[*]

Andrew Chesher and Adam M. Rosen
CeMMAP & UCL & Duke

February 19, 2020

Abstract

Models of simultaneous discrete choice may be incomplete, delivering multiple values of outcomes at certain values of the latent variables and covariates, and incoherent, delivering no values. Alternative approaches to accommodating incompleteness and incoherence are considered in a unifying framework afforded by the Generalized Instrumental Variable models introduced in Chesher and Rosen (2017). Sharp identification regions for parameters and functions of interest defined by systems of conditional moment equalities and inequalities are provided. Almost all empirical analysis of simultaneous discrete choice uses models that include parametric specifications of the distribution of unobserved variables. The paper provides characterizations of identified sets and outer regions for structural functions and parameters allowing for any distribution of unobservables independent of exogenous variables. The methods are applied to the models and data of Mazzeo (2002) and Kline and Tamer (2016) in order to study the sensitivity of empirical results to restrictions on equilibrium selection and the distribution of unobservable payoff shifters, respectively. Confidence intervals for individual parameter components are provided using a recently developed inference approach from Belloni, Bugni, and Chernozhukov (2018). The relaxation of equilibrium selection and distributional restrictions in these applications is found to greatly increase the width of resulting confidence intervals, but nonetheless the models continue to sign strategic interaction parameters.

KEYWORDS: Discrete endogenous variables, Discrete choice, Endogeneity, Incoherent models, Incomplete models, Instrumental variables, Set Identification, Structural econometrics.

JEL CODES: C10, C14, C50, C51.

1

1 Introduction

Simultaneous determination of discrete outcomes arises in many economic contexts. For example, couples make decisions about labour force participation, each partner choosing to work in paid employment part time, full time or not at all. Couples make other discrete lifestyle choices, for example about retirement, activities and pursuits, and work location. There are many joint discrete choices in which each person's actions may affect the welfare of others. Considering network formation people choose to make and break friendships and forge and sunder long-standing relationships and firms choose to trade or not with other firms and decide whether or not to enter markets and which products to sell. In all of these situations the discrete choices of one agent may affect the welfare or profit of others.

Many structural models of simultaneous discrete choice write the utility or profit of each agent as a function of the endogenous discrete choices of others and of exogenous variables, some observed and others unobserved. Any collection of choices by agents that deliver nonnegative utility or profit for all may eventuate. In any particular case there may be many such choices, or just one, or none at all. Which situation arises can depend on the values of observed and unobserved exogenous variables, on the particular functional forms admitted by the model and, in parametric models, on parameter values. In applications researchers must decide how to deal with these possibilities. Many strategies have been proposed.

When multiple outcomes are possible - models admitting this possibility are called *incomplete* - it is feasible to conduct inference using an incomplete model which does not specify which outcome will occur. An alternative is to augment the model with a selection mechanism that always delivers a single outcome. Some researchers augment a model with additional restrictions that ensure a single outcome.

When an empty set of outcomes can be delivered by a model - models admitting this possibility are called *incoherent* - it is possible to proceed to inference with no modification of the model if the null event is observed, and if it is not observed, by assuming unobserved variables are realized only on regions of their support in which the null outcome cannot arise. Some researchers augment the model with a mechanism that delivers a single value of the outcome when the null event occurs. Others impose restrictions that ensure the model they employ is coherent.

In this paper we present a unified treatment of this class of problems. We achieve this by embedding them in the Generalized IV (GIV) framework set out in Chesher and Rosen (2017), henceforth CR17. We characterize sharp identified sets of model unknowns as sets of functions or parameter values that satisfy systems of moment inequalities and equalities. Singleton sets can result, in which case there is point identification.

Partial identification analysis has long been regarded as one possible avenue for sensitivity

analysis.[1] The framework allows (i) the use of different restrictions on outcomes obtained under incoherence, (ii) the relaxation of equilibrium selection restrictions, and (iii) the removal of parametric restrictions on the distribution of unobserved variables. We show how, using the GIV framework, the sensitivity of inference to such restrictions can be assessed. Regarding point (iii) above, almost all empirical analysis of simultaneous discrete choice employs parametric restrictions on the distribution of unobserved variables, usually restricted to be Gaussian. Such restrictions have no provenance in economics. The relaxation of distributional assumptions in moment (in)equality models has been independently considered by Christensen and Connault (2019) for the sake of characterizing bounds on counterfactuals when the distribution of unobservables is restricted to a nonparametric neighborhood of a baseline specification. The approach taken here is different, focusing on characterizations of identified sets for structural functions or their parameters in the absence of any distributional restrictions beyond independence.

We illustrate the application of the GIV framework and its use for investigating sensitivity to relaxations of equilibrium selection and distributional restrictions using the models and data employed in Mazzeo (2002) and Kline and Tamer (2016). These examples are now introduced.

2 Examples

Example 1: KT16. Simultaneous firm entry.

This is the simultaneous binary outcome model considered in Kline and Tamer (2016), henceforth KT16. Binary Y_{LCC} and Y_{OA} indicate the presence of respectively low cost carriers (LCC) and other airlines (OA) operating on an air route in the USA. There are exogenous variables listed in vector $Z \in \mathcal{R}_Z$, structural equations

$$Y_{LCC} = 1\left[Z\beta_{LCC} + Y_{OA}\Delta_{LCC} + U_{LCC} > 0\right], \tag{2.1}$$

$$Y_{OA} = 1\left[Z\beta_{OA} + Y_{LCC}\Delta_{OA} + U_{OA} > 0\right], \tag{2.2}$$

where $Y \equiv (Y_{LCC}, Y_{OA}) \in \mathcal{R}_Y \equiv \{0,1\}^2$, $(Y, Z) \in \mathcal{R}_Y \times \mathcal{R}_Z$ are observable, and $U \equiv (U_{LCC}, U_{OA}) \in \mathbb{R}^2$ is an unobservable 2-vector. This type of model is studied in many papers including Heckman (1978), Bresnahan and Reiss (1990, 1991), and Tamer (2003). In KT16 and most other applications of this model U is restricted to be normally distributed independent of Z with mean zero (a normalization since one element of Z is constant) and unknown covariance matrix, with variances

---

normalized to one.

The indices $Z\beta_{LCC} + Y_{OA}\Delta_{LCC} + U_{LCC}$ and $Z\beta_{OA} + Y_{LCC}\Delta_{OA} + U_{OA}$ can be interpreted as the profit made by respectively LCC and OA carrier types in a market, an airline type operating on a route if profit is positive. In KT16 the data comprise realizations observed on 7882 routes in the US in the second quarter of 2010.

Example 2: M02. Product choice and oligopoly market structure:

We present a simplified version of the model considered in Mazzeo (2002), henceforth M02. There are two types of motel operator: one (L) operating low quality motels, and the other (H) operating high quality motels.

In a market with y_L low quality motels and y_H high quality motels the profit made by a $T \in \{L, H\}$ type firm on opening an additional motel is

$$\pi_T(z, y_L, y_H, u_T) = g_T(z, y_L, y_H) + u_T$$

and an additional type T motel is opened in a market with $y = \{y_L, y_H\}$ motels present if $\pi_T(z, y_L, y_H, u_T) > 0$. Here z is a vector of values of observed exogenous variables and $u = (u_L, u_H)$ are values of unobserved cost shifters.

In the simplified model used here the function $g_T(z, y_L, y_H)$ is the linear index:

$$g_T(z, y_L, y_H) = z\beta_T + \alpha_{TL} y_L + \alpha_{TH} y_H. \tag{2.3}$$

There are more complex specifications in M02, all involving linear indexes.

What inequalities restrict y_L and y_H if operators are making positive profits? In a market with $y = \{y_L, y_H\}$ motels present it cannot be profitable to open an additional type L motel or an additional type H motel, so the inequalities:

$$\pi_L(z, y_L, y_H, u_L) \leq 0 \Longleftrightarrow u_L \leq -g_L(z, y_L, y_H) \tag{2.4}$$

$$\pi_H(z, y_L, y_H, u_H) \leq 0 \Longleftrightarrow u_H \leq -g_H(z, y_L, y_H) \tag{2.5}$$

must both hold at locations with $y = \{y_L, y_H\}$. Additionally, where $y = \{y_L, y_H\}$ it must be more profitable for the market to be in the state $y = \{y_L, y_H\}$ than in neighboring states, $\{y_L - 1, y_H\}$, $\{y_L, y_H - 1\}$ so the following inequalities must hold.

$$\pi_L(z, y_L - 1, y_H, u_L) > 0 \Longleftrightarrow u_L > -g_L(z, y_L - 1, y_H) \tag{2.6}$$

$$\pi_H(z, y_L, y_H - 1, u_H) > 0 \Longleftrightarrow u_H > -g_H(z, y_L, y_H - 1) \tag{2.7}$$

In M02 these positive profit conditions, that is the system of inequalities (2.4) - (2.7) obtained as y varies across all pairs of nonnegative integers, are presented as conditions for a pure strategy Nash

equilibrium. In M02 unobserved $U \equiv (U_L, U_H)$ is restricted to be independent of Z and bivariate Gaussian.[2] Data comprise realizations of $\{Y_L, Y_H\}$ and exogenous Z observed in the mid 1990's at 492 small rural exits on 30 US Interstate Highways at which at least one motel was located. A consequence of this sampling scheme is that there is no data on markets with $Y = \{0, 0\}$ which requires attention in the econometric analysis. To simplify the econometric analysis in M02 the data are censored such that y_T recorded as 3 indicates 3 or more motels of type $T \in \{L, H\}$.

A selection mechanism is imposed in M02 resulting in a complete model which is estimated using maximum likelihood methods. The empirical analysis here is done using the incomplete model with no selection mechanism. Our characterizations of identified sets of model unknowns apply whether or not a selection mechanism is imposed.

3 The GIV framework

All the models considered in this paper require that endogenous discrete outcomes listed in vector Y, observed exogenous variables listed in vector Z, and unobserved variables listed in vector $U \in \mathcal{R}_U$ are generated by structures that, for some admissible function h satisfy

$$\mathbb{P}[h(Y, Z, U) = 0] = 1 \tag{3.1}$$

where (Y, Z, U) are defined on a probability space $(\Omega, \mathsf{L}, \mathbb{P})$, endowed with the Borel sets on Ω. These variables may have any finite dimension. A structure comprises a couple $(h, \mathcal{G}_{U|Z})$ where h is a structural function as in (3.1) and

$$\mathcal{G}_{U|Z} \equiv \{G_{U|Z}(\cdot|z) : z \in \mathcal{R}_Z\}$$

is a collection of conditional distributions of U given $Z = z$. $\mathcal{R}_Z$ is the support of Z and $G_{U|Z}(\mathcal{S}|z)$ is the probability mass on the set $\mathcal{S} \subseteq \mathcal{R}_U$ conditional on $Z = z$.

A GIV model comprises restrictions which determine which structures are admissible. Considering the two examples, in both cases unobserved U is restricted independent of Z over their joint support and U is restricted to be bivariate Gaussian. So the collection $\mathcal{G}_{U|Z}$ is restricted to be simply the singleton set $\{G_U(\cdot)\}$ comprising a Gaussian distribution with some unknown parameters. In Example 1 a suitable structural function is:

$$h((y_{LCC}, y_{OA}), z, (u_{LCC}, u_{OA})) =$$
$$\left(y_{LCC} \cdot |z\beta_{LCC} + y_{OA}\Delta_{LCC} + u_{LCC}|_-\right) + \left((1 - y_{LCC}) \cdot |z\beta_{LCC} + y_{OA}\Delta_{LCC} + u_{LCC}|_+\right)$$
$$+ \left(y_{LOA} \cdot |z\beta_{OA} + y_{LCC}\Delta_{OA} + u_{OA}|_-\right) + \left((1 - y_{OA}) \cdot |z\beta_{OA} + y_{LCC}\Delta_{OA} + u_{OA}|_+\right),$$

---

[2]The correlation is restricted to be zero in the empirical analysis reported in M02.

where $|\cdot|_-$ and $|\cdot|_+$ denote the negative and positive part of their arguments, respectively. In Example 2 the structural function

$$h((y_L, y_H), z, (u_L, u_H)) \equiv \max(0, g_L(z, y_L, y_H) + u_L) + \max(0, g_H(z, y_L, y_H) + u_H)$$
$$+ \max(0, -g_L(z, y_L - 1, y_H) - u_L) + \max(0, -g_H(z, y_L, y_H - 1) - u_H)$$

captures the restrictions.[3] In both examples the models impose linear index restrictions and there may be restrictions on the admissible values of parameters, for example requiring that variation in some elements of Z has no effect on one or other of the indexes.

The set of outcomes a structure can deliver at a value of (Z, U) is a zero level set of the structural function, as follows.

$$\mathcal{Y}(z, u; h) \equiv \{y : h(y, z, u) = 0\}$$

In a structure where there are values of (z, u) for which the outcome set $\mathcal{Y}(z, u; h)$ has more than one element, multiple outcomes are feasible, and the structure is termed *incomplete*. A structure in which there are values of (z, u) for which $\mathcal{Y}(z, u; h)$ is empty is termed *incoherent*. Models which admit incomplete structures respectively incoherent structures are incomplete respectively incoherent models and there are models with both attributes. The rather loaded term "proper" has been use to describe models that are both complete and coherent.

A level set on the support of U that is dual to the outcome set $\mathcal{Y}(z, u; h)$ is the residual set

$$\mathcal{U}(y, z; h) \equiv \{u : h(y, z, u) = 0\}$$

which will play an important role in the subsequent analysis. This set contains all values of unobserved U that deliver $Y = y$ when $Z = z$. In many econometric models this is a singleton set but in models of simultaneous discrete choice, when U is continuously distributed, it is not.[4] Complete structures have disjoint U level sets at every value of Z because where residual sets intersect one value of U can deliver more than one value of Y.[5] Structures which are both complete

---

[3] This specification of h allows either $y_T - 1$ or y_T when the marginal profit of y_T for type T is zero. Strictly speaking this differs from (2.4) - (2.7), but only on a set of values of U of Lebesgue measure zero, and is hence of no consequence. Likewise, the specification of h in Example 1 allows each y_T to be either zero or one when the profit from entering is exactly equal to zero, which is a zero probability event.

[4] There are also nonsingleton residual sets in models with continuous or discrete outcomes that admit unobservables with higher dimension than outcomes, for example panel data models with error components and random coefficient models.

[5] The terminology used here follows Tamer (2003) and Lewbel (2007) in distinguishing between incoherent and incomplete models. Some earlier papers in the literature, e.g. Gourieroux, Laffont and Monfort (1980) and Blundell and Smith (1994) use incoherence to mean either incoherence or incompleteness as we have defined them. This alternative definition renders coherence equivalent to the existence of a unique reduced form, as described by Lewbel (2007), and is also equivalent to our use of the term *proper*. Many of the models employed in the early history of econometrics were proper, and in such models a unique conditional distribution of outcomes given observed exogenous variables can be obtained by transformation, via the structural function, of the conditional distribution of unobserved

Table 1: Values of competition parameters in graphs of residual sets for the M02 model

Figure	α_{LL}	α_{LH}	α_{HL}	α_{HH}
2	-1.5	-0.75	-2.0	-0.8
3	-1.3	-1.8	-1.4	-1.9
4	-1.8	-1.3	-1.4	-1.8

and coherent have residual sets that partition the support of U at each value of Z. Then with a parametric specification of the distribution of U a likelihood analysis is feasible.

Figure 1 shows the residual sets of the simultaneous firm entry model of Example 1 for four particular structures at a particular value of Z. The upper panes have $\Delta_{LCC}\Delta_{OA} > 0$, both negative in the left hand pane, both positive in the right hand pane. In each case there are residual sets with non-empty intersections indicating incomplete structures. Considering the left hand upper pane the overlapping residual sets are $\mathcal{U}((0,1), z; h)$ and $\mathcal{U}((1,0), z; h)$. The union of these residual sets has no intersection with any other residual set so the structure is complete for the outcome $Y^+ \equiv Y_{LCC} + Y_{OA}$ which is the number of carrier types operating on a route. With the Gaussian restriction the KT16 model is a complete parametric model for the outcome Y^+ and maximum likelihood estimation is feasible although this will not use all the information contained in the data. ML estimates are reported later. Similarly the structure with residual sets shown in the right hand pane of Figure 1 is complete for the outcome $Y^- \equiv Y_{LCC} - Y_{OA}$. The lower panes in Figure 1 show residual sets for structures with $\Delta_{LCC}\Delta_{OA} < 0$. For values of U in the central region labelled $\mathcal{U}(\phi, z; h)$ there are no values of Y that satisfy the restrictions of the model, that is, for all $u \in \mathcal{U}(\phi, z; h)$, $\mathcal{Y}(u, z; h) = \emptyset$, the empty set. The other residual sets have no intersection so these structures are complete and incoherent. Structures with $\Delta_{LCC}\Delta_{OA} = 0$ have residual sets which partition the support of U. Such structures are complete and coherent.

Figures 2, 3 and 4 show residual sets for the M02 model for three particular structures with different relative magnitudes of the competition parameters α_{LL}, α_{LH}, α_{HL} and α_{HH} as shown in Table 1. In all cases there are many intersecting residual sets indicating that these are incomplete structures and the nature of the intersections varies with the relative magnitudes of the competition parameters. In some cases (not shown in these graphs) a positive value for a competition parameter produces an incoherent structure. The structure generating Figure 4 is complete for the outcome $Y^+ \equiv Y_L + Y_H$ which is reminiscent of the KT16 model. Recall that high values of Y_L and Y_H are censored with values recorded as 3 indicating a value at least equal to 3. The M02 data records no locations with $Y = (0,0)$ so in the empirical analysis the distribution of U will be the truncated distribution which places zero probability mass on $\mathcal{U}((0,0), z; h)$.

variables given observed exogenous variables.

3.1 Plan of the remainder of the paper

Chesher and Rosen (2017) provides characterizations of identified sets of structures in GIV models under the conditions set out in Restrictions A1-A6 below. The result from CR17 employed here is given in Theorem 1. The following Section 4 applies the result to characterize identified sets in structural models of simultaneous discrete choice under various treatments of incompleteness and incoherence. In Section 5 the results are used in an analysis of the models and data of KT16 and M02. Section 6 provides characterization of a fast-to-compute outer region for the projection of an identified set of structures onto the space of structural functions, or parameters in case of a parametric specification. A method for determining whether a structural function in the outer region is in the projection is provided for the case in which all observed variables are discrete. The results are applied to the KT16 model and data. Section 7 concludes and outlines ongoing research.

3.2 Identified sets delivered by GIV models

Here are the restrictions employed in the GIV framework used here.[6]

Restriction A1: (Y, Z, U) are random vectors defined on a probability space $(\Omega, \mathsf{L}, \mathbb{P})$, endowed with the Borel sets on Ω. The support of (Y, Z, U) is a subset of a finite-dimensional Euclidean space. $\square$

Restriction A2: A collection of conditional distributions on $\mathcal{R}_Y$ denoted

$$\mathcal{F}_{Y|Z} \equiv \left\{ F_{Y|Z}\left(\cdot|z\right) : z \in \mathcal{R}_Z \right\}$$

is identified by the sampling process, where for all $\mathcal{T} \subseteq \mathcal{R}_{Y|z}$, $F_{Y|Z}\left(\mathcal{T}|z\right) \equiv \mathbb{P}\left[Y \in \mathcal{T}|z\right]$. $\square$

Restriction A3: There is an L-measurable function $h\left(\cdot, \cdot, \cdot\right) : \mathcal{R}_{YZU} \to \mathbb{R}$ such that

$$\mathbb{P}\left[h\left(Y, Z, U\right) = 0\right] = 1,$$

and there is a collection of conditional distributions on $\mathcal{R}_U$ denoted

$$\mathcal{G}_{U|Z} \equiv \left\{ G_{U|Z}\left(\cdot|z\right) : z \in \mathcal{R}_Z \right\},$$

where for all $\mathcal{S} \subseteq \mathcal{R}_{U|z}$, $G_{U|Z}\left(\mathcal{S}|z\right) \equiv \mathbb{P}\left[U \in \mathcal{S}|z\right]$. $\square$

Restriction A4: The pair $\left(h, \mathcal{G}_{U|Z}\right)$, termed a structure, belongs to a known set of admissible structures $\mathcal{M}$. $\square$

Restriction A5: $\mathcal{U}\left(Y, Z; h\right) \equiv \{u : h(Y, Z, u) = 0\}$ is closed almost surely $\mathbb{P}\left[\cdot|z\right]$, each $z \in \mathcal{R}_Z$. $\square$

Restriction A6: $\mathcal{Y}\left(Z, U; h\right) \equiv \{y : h(y, Z, U) = 0\}$ is closed almost surely $\mathbb{P}\left[\cdot|z\right]$, each $z \in \mathcal{R}_Z$. $\square$

---

[6]Throughout, the notation $\mathcal{R}_A$ is used to denote the support of a random vector A, and likewise $\mathcal{R}_{AB}$ and $\mathcal{R}_{A|b}$ are used to denote the joint support of (A, B) and the conditional support of A given $B = b$, respectively.

Restriction A1 defines the probability space on which (Y, Z, U) reside and restricts their support to Euclidean space. Restriction A2 requires that for each $z \in \mathcal{R}_Z$, $F_{Y|Z}(\cdot|z)$ is identified. It is convenient to have notation

$$f_{Y|Z}(y|z) \equiv F_{Y|Z}(\{y\}|z) = \mathbb{P}[Y = y|Z = z], \quad y \in \mathcal{R}_{Y|z}$$

for the simple probability mass function which gives the probability mass on a single point of support of Y given $Z = z$. Restriction A3 posits the existence of structural relation h, and provides notation for the collection of conditional distributions $\mathcal{G}_{U|Z}$ of U given Z.

Restriction A4 imposes model $\mathcal{M}$, the collection of admissible structures $(h, \mathcal{G}_{U|Z})$. Unlike the previous restrictions, it is refutable based on knowledge of $\mathcal{F}_{Y|Z}$ in that it is possible that there is no $(h, \mathcal{G}_{U|Z}) \in \mathcal{M}$ such that $\mathbb{P}[h(Y, Z, U) = 0] = 1$. In such cases the identified set of structures is empty, indicating model misspecification. Let $\mathcal{H} = \{h : \exists \mathcal{G}_{U|Z} \quad s.t. \ (h, \mathcal{G}_{U|Z}) \in \mathcal{M}\}$ denote the projection of $\mathcal{M}$ onto the space of structural functions.

Restrictions A5 and A6 restrict $\mathcal{U}(Y, Z; h)$ and $\mathcal{Y}(Z, U; h)$ to be random closed sets. These restrictions are satisfied for example if $\mathcal{M}$ specifies that all admissible h are continuous in their first and third arguments, respectively, but can also hold more generally. A given econometric model can generally be represented through a variety of different but substantively equivalent structural functions h, and judicious choice of this function can often be made to ensure these requirements are satisfied.

Here is a definition of the conditional containment functional of the random set $\mathcal{U}(Y, Z; h)$ for a set $\mathcal{S}$.[7] This gives the conditional probability that the random set $\mathcal{U}(Y, Z; h)$ is a subset of $\mathcal{S}$ for a structural function h. It plays a key role in what follows.

Definition 1 *The conditional containment functional of the random set $\mathcal{U}(Y, Z; h)$ for a set $\mathcal{S} \subseteq \mathcal{R}_{U|z}$ given $Z = z$ is:*

$$C_h(\mathcal{S}|z) \equiv \mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{S}|Z = z] = F_{Y|Z}(\mathcal{A}(\mathcal{S}, z; h)|z) = \sum_{y \in \mathcal{A}(\mathcal{S}, z; h)} f_{Y|Z}(y|z)$$

where

$$\mathcal{A}(\mathcal{S}, z; h) \equiv \{y : \mathcal{U}(y, z; h) \subseteq \mathcal{S}\} \tag{3.2}$$

is the set of values of Y that can only be realized when $U \in \mathcal{S}$.

Now we define a collection of subsets of $\mathcal{R}_{U|z}$ which comprises all possible unions of residual sets $\mathcal{U}(y, z; h)$.

---

[7] For the formal definition of a random closed set see e.g. Molchanov and Molinari (2018) or Molinari (2020).

Definition 2 *The conditional support of random set* $\mathcal{U}(Y, Z; h)$ *given* $Z = z$ *is* $\mathsf{U}(h, z)$.

$$\mathsf{U}(h, z) \equiv \left\{ \mathcal{U} \subseteq \mathcal{R}_U : \exists y \in \mathcal{R}_{Y|z} \text{ such that } \mathcal{U} = \mathcal{U}(y, z; h) \right\}.$$

$\mathsf{U}^*(h, z)$ *is the collection of all sets that are unions of the sets in the collection* $\mathsf{U}(h, z)$:

$$\mathsf{U}^*(h, z) \equiv \left\{ \mathcal{U} \subseteq \mathcal{R}_U : \exists \mathcal{Y} \subseteq \mathcal{R}_{Y|z} \text{ such that } \mathcal{U} = \bigcup_{y \in \mathcal{Y}} \mathcal{U}(y, z; h) \right\}.$$

There is the following Theorem which follows directly from Corollary 1, Lemma 1 and Theorem 3 of CR17.

Theorem 1 *Under Restrictions A1-A6, the identified set of structures is*

$$\mathcal{M}^* \equiv \left\{ (h, \mathcal{G}_{U|Z}) \in \mathcal{M} : \forall \mathcal{S} \in \mathsf{Q}(h, z), \ C_h(\mathcal{S}|z) \leq G_{U|Z}(\mathcal{S}|z), \ a.e. \ z \in \mathcal{R}_Z \right\}, \quad (3.3)$$

where $\mathsf{Q}(h, z)$, *defined in Theorem 3 of CR17, is a selection of sets from the collection* $\mathsf{U}^*(h, z)$.

The main practical import of the refinement of $\mathsf{U}^*(h, z)$ delivered by Theorem 3 of CR17 to produce $\mathsf{Q}(h, z)$ is that, when all residual sets are connected, it allows exclusion from consideration of those sets in $\mathsf{U}^*(h, z)$ that are disjoint. Corollary 2 of CR17 gives conditions under which inequalities in the characterization of $\mathcal{M}^*$ become equalities. This leads to the conclusion that the identified set of structures is entirely characterized by *equality* restrictions either when residual sets are singleton with probability 1 or when the model under consideration is complete, or when both conditions are satisfied.[8]

These results are now applied to models of simultaneous discrete choice. In all the cases considered identified sets of structures are characterized by systems of inequalities as in (3.3), the residual sets under consideration and the collections of distributions $\mathcal{G}_{U|Z}$ varying from case to case.

4 Identification

In this section we consider four approaches to the analysis of incoherent and incomplete models. The first three deal with incompleteness in the same way as the recent literature on set identification in incomplete models, the observed outcome being permitted to be any one of the set

---

[8] For instance, the multinomial logit model studied in McFadden (1974) is complete. In their analysis of instrumental variable models of multiple discrete choice, Chesher, Rosen and Smolinski (2011) in Section 3.2 show how moment inequalities that characterize the identified set reduce to moment equalities when there are no endogenous variables. In that case the model is complete, and if the alternative-specific utility functions are restricted to be linear with additive unobservable utility shifters distributed i.i.d. Gumbel, the moment inequalities correspond to the likelihood contributions of the model studied in McFadden (1974).

Table 2: A summary of the determination of outcome Y for the approaches of section 4 when a model delivers no or multiple solutions.

| | Restrictions for the determination of Y when: | |
	$\mathcal{Y}(U, Z; h) = \phi$: (no value)	$\mathrm{card}(\mathcal{Y}(U, Z; h)) > 1$ (multiple values)
Approach IC1	$Y = \phi$	$Y \in \mathcal{Y}(U, Z; h)$
Approach IC2	not observed	$Y \in \mathcal{Y}(U, Z; h)$
Approach IC3	$Y \in R_{Y|z}$	$Y \in \mathcal{Y}(U, Z; h)$
Approach IC4	not observed	not observed

of outcomes delivered by an *incomplete* structure. The approaches differ in the way the event $\{\mathcal{Y}(Z, U; h) = \emptyset\}$ is handled in incoherent structures. The fourth approach, proposed by Dagenais (1997) and Hajivassiliou (2008), treats incoherence and incompleteness identically in a way that enables construction of a unique likelihood function when there is a parametric specification of the distribution of unobserved U. The four approaches are summarized in Table 2.

Maximum likelihood estimation of the KT16 model using data on the number of airline types on each route (which we show is feasible in section 5.1) delivers results reported in Section 5.1 suggesting that the KT16 data is delivered by an incoherent structure. So attention to incoherency is necessary.

In all of the models considered now U is specified to be continuously distributed and U and Z are restricted to be stochastically independent on their joint support.

Restriction A7: Unobservable U is absolutely continuously distributed with respect to Lebesgue measure on its support $\mathcal{R}_U = \mathbb{R}^{k_u}$ and U and Z are stochastically independent on their joint support $\mathcal{R}_U \times \mathcal{R}_Z$, which is a subset of a finite dimensional Euclidian space. $\square$

4.1 Approach IC1: Observed Null Outcomes

This approach can be employed when the occurrence of the event $\{\mathcal{Y}(Z, U; h) = \emptyset\}$ can be observed. Data obtained in the field rarely contain such information but it could be present in data generated by experimental studies.

We use ϕ as a place-holder for the null realization that Y takes when the event $\{\mathcal{Y}(Z, U; h) = \emptyset\}$ occurs and we define $\mathcal{U}(\phi, z; h)$ as the set of values of U such that no value $y \in \mathcal{R}_Y$ satisfies $h(y, z, u) = 0$. Thus we have the equivalence of events

$$\{Y = \phi\} \Leftrightarrow \{\mathcal{Y}(Z, U; h) = \emptyset\} \Leftrightarrow \{U \in \mathcal{U}(\phi, z; h)\}. \tag{4.1}$$

and for all $z \in \mathcal{R}_Z$, $\mathbb{P}[\mathcal{Y}(Z, U; h) = \emptyset | Z = z] = \mathbb{P}[Y = \phi | Z = z]$. The probability mass allocated by $\mathbb{P}[\cdot | Z = z]$ over the extended support $\mathcal{R}_Y^* \equiv \mathcal{R}_Y \cup \phi$ is one. With the event $\{Y = \phi\}$ observable, $\mathbb{P}[Y \in \mathcal{T} | Z = z] = F_{Y|Z}(\mathcal{T}|z)$ is identified for all sets $\mathcal{T} \in \mathcal{Y}^*$ and almost every $z \in \mathcal{R}_Z$. We

extend the definition of the structural function $h\left(\cdot,\cdot,\cdot\right)$ to cover the extended support $\mathcal{R}_Y^*$ with the following restriction.

Restriction B1 (Observed Null Outcomes): The outcome variable Y has support $\mathcal{R}_Y^* = \mathcal{R}_Y \cup \phi$, and there is the equivalence of events $\{U \in \mathcal{U}(\phi, Z; h)\} \Leftrightarrow \{Y = \phi\}$. Further, for any (u, z), $h\left(\phi, z, u\right) = 0$ if for all $y \in \mathcal{R}_Y$, $h(y, z, u) \neq 0$, and $h\left(\phi, z, u\right) \neq 0$ if $h(y, z, u) = 0$ for some $y \in \mathcal{R}_Y$. $\square$

Under Restrictions A1-A7 and B1 Theorem 1 characterizes the identified set of structures with (i) the structural function h interpreted as the extended structural function of Restriction B1, (ii) the residual set $\mathcal{U}(\phi, z; h)$ contributing in the construction of the unions of residual sets that comprise the collection of sets $\mathsf{Q}(h, z)$, (iii) $\mathcal{G}_{U|Z}$ replaced by G_U and $G_{U|Z}(\cdot|z)$ replaced by $G_U(\cdot)$ by virtue of Restriction A7.

4.2 Approach IC2: Truncated null outcomes

This approach can be taken when all the observable realizations of (Y, Z) come from realizations of (U, Z) such that $\mathcal{Y}(u, z; h)$ is non-empty. If the structural function h is incoherent there could be realizations of (U, Z) such that $\mathcal{Y}(u, z; h)$ is empty, but in the scenario considered here the occurrence of the event $\{\mathcal{Y}(U, Z; h) = \emptyset\}$ is not observed. This will commonly be the case when data are gathered in the field.

In this situation the observable realizations of (Y, Z) are associated with realizations of (U, Z) from the *truncated* support: $\{(u, z) \in \mathcal{R}_{UZ} : \mathcal{Y}(u, z; h) \neq \emptyset\}$. Accordingly there is Restriction B2.

Restriction B2 (Truncated Null Outcomes): All realizations of (Y, Z) are such that $\mathcal{Y}(u, z; h)$ is non-empty. For each $z \in \mathcal{R}_Z$ the distribution of Y conditional on $Z = z$, $F_{Y|Z}(\cdot|z)$, identified by the sampling process is the distribution of Y given $Z = z$ and $\mathcal{Y}(U, Z; h) \neq \emptyset$. $\square$

The probability of the event $\{\mathcal{Y}(U, Z; h) = \emptyset\}$ is

$$\mathbb{P}[\mathcal{Y}(U, Z; h) = \emptyset | Z = z] = G_U(\mathcal{U}(\phi, z; h))$$

where the restriction A7 that U and Z are stochastically independent on their full joint support is imposed.

Under Restrictions A1-A7 and B2, Theorem 1 characterizes the identified set of structures with

$$G_{U|Z}(S|z) = \frac{G_U(S)}{1 - G_U(\mathcal{U}(\phi, z; h))} \tag{4.2}$$

which embodies both Restriction A7 and the effect of truncation of the distribution of U.

4.3 Approach IC3: Indeterminate allocation of null outcomes

In this approach when the event $\{\mathcal{Y}(U, Z; h) = \emptyset\}$ occurs any value of Y on its support can be realized and the model places no restriction on the selection of this value. This reflects the idea that the model conveys no information about the way in which Y is generated when no value of Y satisfies the condition $h(y, z, u) = 0$. This approach, captured in Restriction B3, was considered by Beresteanu, Molchanov and Molinari (2011, online supplement p.53).

Restriction B3 (Indeterminate allocation of null outcomes): If (U, Z) are such that $\mathcal{Y}(U, Z; h) = \emptyset$, equivalently if $U \in \mathcal{U}(\phi, Z; h)$ then any value of Y in $\mathcal{R}_Y$ is feasible. $\square$

In this case the residual sets are

$$\mathcal{U}(y, z; h) = \{u : h(y, z, u) = 0\} \cup \{u : \mathcal{Y}(u, z; h) = \emptyset\} \tag{4.3}$$

equivalently

$$\mathcal{U}(y, z; h) = \{u : \tilde{h}(y, z, u) = 0\}$$

where $\tilde{h}$ is a modified structural function.

$$\tilde{h}(y, z, u) = 1[\mathcal{Y}(u, z; h) \neq \emptyset] \times h(y, z, u)$$

Under Restrictions A1-A7 and B3 Theorem 1 characterizes the identified set of structures with the structural function $\tilde{h}$ in place of the structural function h and with $G_U(\mathcal{S})$ in place of $G_{U|Z}(\mathcal{S}|z)$ which results from Restriction A7.

In Example 1, in the lower panes of Figure 1 which shows residual sets for incoherent structures, the residual sets under Approach 3 are each the union of a residual set with the central rectangular region in which $\mathcal{Y}(u, z; h) = \emptyset$. Since the resulting residual sets have non-null intersections the model obtained under Approach IC3 is incomplete. In our analysis of the KT16 data we do not constrain the signs of the competition parameters, allowing the possibility of incoherence and employ Approach IC2 and Approach IC3 in complementary analyses.

4.4 Approach IC4: Truncation of null and non-unique outcomes

In this approach, employed by Dagenais (1997) and Hajivassiliou (2008) it is assumed that observed data are regarded as realizations of (Y, Z) obtained from realizations of (U, Z) such that $\mathcal{Y}(U, Z; h)$ is a non-null singleton set.

Restriction B4 (Truncated null and nonunique outcomes): All realizations of (Y, Z) are such that $\mathcal{Y}(u, z; h)$ is a non-empty singleton set. For each $z \in \mathcal{R}_Z$ the distribution of Y conditional on $Z = z$, $F_{Y|Z}(\cdot|z)$, identified by the sampling process is the distribution of Y given $Z = z$ and card $(\mathcal{Y}(U, Z; h)) = 1$. $\square$

For this approach define the residual sets as follows

$$\mathcal{U}(y, z; h) \equiv \{u : h(y, z, u) = 0 \text{ and } \forall y' \neq y \quad h(y', z, u) \neq 0)\} \tag{4.4}$$

and define the set

$$\mathcal{B}(z; h) \equiv \{u : \text{card}\left(\mathcal{Y}(u, z; h)\right) \neq 1\}$$

Under Restrictions A1-A7 and B2, Theorem 1 characterizes the identified set of structures with the residual sets as defined in (4.4) and with

$$G_{U|Z}(S|z) = \frac{G_U(S)}{1 - G_U(\mathcal{B}(z; h))}$$

which embodies both Restriction A7 and the effect of truncation of the distribution of U.

The model obtained under Approach IC4 is complete and coherent by construction and the characterization delivered by Theorem 1 comprises a system of moment equalities. Estimation by maximum likelihood is feasible with a parametric specification of G_U.

4.5   Complete and coherent models via restrictions on structures

In some applications incompleteness and incoherence are ruled out by imposing restrictions on structural functions that guarantee outcome sets $\mathcal{Y}(Z, U; h)$ are singleton with probability 1 for all admissible h.

Amemiya (1974) imposes such conditions in a class of simultaneous Tobit models as does Heckman (1978) in models with simultaneous equations in binary outcomes. Gourieroux, Laffont and Monfort (1980) and Blundell and Smith (1994) impose coherency conditions in their analysis of simultaneous equations regime-switching models and simultaneous equations models of qualitative or censored outcomes, respectively. For a more thorough review of this approach in a variety of models see e.g. Schmidt (1981) and Maddala (1983).

Alternative approaches to completing incomplete models include redefining the outcome variable or specifying a selection mechanism that chooses one from multiple potential outcomes. In simultaneous binary response models of firm entry, Bresnahan and Reiss (1990, 1991) circumvent problems posed by incompleteness by exploiting the observation that their model uniquely determines the number of entrants. Bjorn and Vuong (1984) and Kooreman (1994) augment models which are incomplete by specifying that whenever $\mathcal{Y}(Z, U; h)$ contains more than one outcome, one of these is selected by some stochastic process characterized by additional parameters. Bajari, Hong and Ryan (2004) illustrate how to incorporate equilibrium selection mechanisms into general discrete games of complete information.

All these approaches yield specifications of the structural function that result in a model that is complete and coherent. The methods that we employ apply to such models with the simplification

that identified sets in such complete models are characterized by a system of moment *equalities* as shown in CR17.

5 Applications

In this Section we employ the data used in KT16 and M02 to estimate the identified sets delivered by their parametric models.[9] Where parameter values lead to an incomplete structure the observed outcome is permitted to be any member of the set of outcomes delivered by the structure, the model being silent about the manner in which a value emerges. Where parameter values lead to incoherent structures we adopt approaches IC2 and IC3 in turn. The focus is first and mainly on the KT16 data and model.

5.1 KT16 model and data

The KT16 data come from the second quarter of the 2010 Airline Origin and Destination Survey (DB1B).[10] The data contain 7882 markets, defined as trips between two airports irrespective of intermediate stops. For each market there are binary indicators recording the presence of each airline type, Y_{LCC} for Low Cost Carriers and Y_{OA} for Other Airlines and values of binary explanatory variables: Z^{size} which is a market specific variable with common values across airline types measuring market size and Z^{pres}_{LCC} and Z^{pres}_{OA} which are market- and airline- type-specific variables measuring market presence.

In KT16 (page 356) the calculation of the market presence and size variables is described as follows.

> The analysis considers two explanatory variables: market presence and market size. The first explanatory variable is market presence, which is a market- and airline-specific variable: for each airline and for each airport, compute the number of markets that airline serves from that airport and divide by the total number of markets served from that airport by any airline. The market presence variable for a given market and airline is the average of these ratios (excluding the one market under consideration) at the two endpoints of the trip, providing some proxy for an airline's presence in the airports associated with that market. See also Berry (1992)....[T]he market presence for the LCC firm (resp. OA firm) is the maximum among the actual airlines in the LCC category (resp. OA category). The second explanatory variable is market size, which is a market-specific variable (but shared by all airlines in that market), which is defined as the population at the endpoints of the trip. The market size and market presence

---

[9]In the case of M02 we use a much simplified version of the model.

[10]The data has also been subsequently used in applications studied in Chen, Christensen, and Tamer (2018) and Kaido, Molinari, and Stoye (2019).

variables actually used in the empirical application are discretized binary variables based on the continuous variables just described. They take the value of 1 if the variable is higher than its median value and 0 otherwise.

The profit made by airline type $t \in \{LCC, OA\}$ operating in market m is

$$\beta_t^{cons} + \beta_t^{size} Z_m^{size} + \beta_t^{pres} Z_{t,m}^{pres} + \Delta_t Y_{s,m} + U_{t,m}$$

where $s \neq t$ with $s \in \{LCC, OA\}$.[11] For further details see KT16. Define the parameter vector θ.

$$\theta \equiv (\beta_{LCC}^{cons}, \beta_{OA}^{cons}, \beta_{LCC}^{size}, \beta_{OA}^{size}, \beta_{LCC}^{pres}, \beta_{OA}^{pres}, \Delta_{LCC}, \Delta_{OA}, \rho).$$

First consider a measure of the distance of a value of θ from the identified set defined by the inequalities given in (3.3). The distance measure used here is $D(\theta)$ defined as

$$D(\theta) \equiv \sum_{z \in \mathcal{R}_Z} \sum_{\mathcal{S} \in \mathsf{Q}(\theta, z)} \max \left(0, C_\theta \left(\mathcal{S} | z \right) - G_{U|Z} \left(\mathcal{S} | z; \theta \right) \right) \tag{5.1}$$

where

$$z = \left(z^{size}, z_{LCC}^{pres}, z_{OA}^{pres} \right)$$

The notation makes explicit that with the KT16 parametric specification the 9-element parameter vector θ characterizes a structure. Note that θ appears as an argument of the distribution since in the Gaussian specification there is the correlation parameter ρ which is an element of θ. The distance measure $D(\theta)$ is zero for values of θ in the identified set and positive for values outside it.

For values of θ that deliver an incoherent structure we define $D_2(\theta)$ for the case in which we adopt approach IC2 and $D_3(\theta)$ for the case in which we adopt approach IC3.[12] In calculating $D_2(\theta)$, $G_{U|Z} \left(\cdot | z; \theta \right)$ is the truncated Gaussian distribution with zero probability mass on $\mathcal{U}(\phi, z; \theta)$, defined in (4.2). In calculating $D_3(\theta)$, $G_{U|Z} \left(\cdot | z; \theta \right)$ is bivariate Gaussian *independent* of z while the residual sets used in calculating $C_\theta \left(\mathcal{S} | z \right)$ are the sets defined in (4.3).

We study *estimated* distance measures, $\hat{D}_s(\theta)$, $s \in \{2, 3\}$, in which containment probabilities are estimated using the KT16 data. The KT16 data are used to estimate 32 conditional probabilities, $F_{Y|Z}(\{y\}|z) = f_{Y|Z}(y|z)$, of each of 4 values of Y, $y \in \mathcal{R}_Y = \{(0,0), (1,0), (0,1), (1,1)\}$ and 8 values of the three binary exogenous variables. The estimates, simple relative cell frequencies calculated from the 7882 observations are denoted $\hat{f}_{Y|Z}(y|z)$ and the formula for the estimated containment

---

[11] In KT16 the term payoff is used where here we use profit.
[12] The data carry no record of a route for which $Y = \phi$, so approach IC1 is not available and we choose not to consider approach IC4.

probabilities is a linear function of these probabilities

$$\hat{C}_\theta \left(\mathcal{S} | z \right) = \sum_{y \in \mathcal{R}_Y} 1[y \in \mathcal{A}(\mathcal{S}, z; \theta)] \times \hat{f}_{Y|Z}(y|z)$$

where $\mathcal{A}(\mathcal{S}, z; \theta)$ is as defined in (3.2) with θ replacing h in view of the parametric specification. There are up to 112 inequalities and equalities contributing to the values of the estimated distance measures.

The minimum value of the estimated distance measures using either of the approaches IC2 and IC3 is positive and small, equal to 0.43. The values of the parameters that minimize the distance measures are shown in column 2 of Table 3. The distance-measure-minimizing values of Δ_{LCC} and Δ_{OA} are both negative and structures with these parameter values are coherent so it is to be expected that approaches IC2 and IC3 give the same results.[13]

Because the minimum value of the distance measures is positive the analog estimate of the identified set of parameter values is *empty*. This could be an indication of model misspecification but a likely major factor is inward bias in the analog set estimate arising because of sampling variation in the estimated containment probabilities. The identified set comprises values of θ such that $D(\theta) \leq 0$, equivalently such that

$$\max_{z \in \mathcal{R}_Z} \max_{\mathcal{S} \in \mathsf{Q}(\theta, z)} \left(C_\theta \left(\mathcal{S} | z \right) - G_{U|Z} \left(\mathcal{S} | z; \theta \right) \right) \leq 0.$$

Applying the nonlinear max operator to noisy estimates of the containment probability $C_\theta \left(\mathcal{S} | z \right)$ will tend to result in an overestimate of $D(\theta)$ leading to a set estimate which is inward biased.

Inspecting the contributions to the estimated distance measure by the 112 restrictions we find that there are 48 equality restrictions, 6 at each value of the exogenous variables. This is as predicted by Corollary 2 of CR17, arising because at and near the parameter values at which the distance measure is minimized there are residual sets and unions of residual sets in each collection $\mathsf{Q}(z, \theta)$ which have no intersection with other residual sets. Indeed as noted in Bresnahan and Reiss (1990, 1991) the KT16 model is complete at these parameter values for the outcome $Y^+ = Y_1 + Y_2$ which is the number of airline types in the market, and identified sets in complete models are characterized by moment equalities as noted earlier. All the equality restrictions and a few of the inequality restrictions are violated at the distance-measure-minimizing parameter value.

Exact probabilities were calculated for an incomplete and coherent KT16 structure with the parameter values shown in Table 4.[14] The value of the distance measure at that parameter value

---

[13] It was necessary to consider both approaches in case a distance-measure-minimizing parameter value delivering an incoherent structure was found using either approach.

[14] A selection mechanism was imposed with random selection of one from multiple potential outcomes such that each potential outcome was equally likely to be chosen.

is zero to within machine accuracy.[15] Evaluating the distance measure in a hypercube with side of length 0.002 around the parameter value that generates the probabilities delivers positive values between 0.001 and 0.004 possibly suggestive of local point identification.

In the presence of this large number of equality restrictions it is possible that the KT16 model is point identifying at parameter values around those that minimize the distance measure and inward bias in estimating a singleton set leads to an estimated set that is empty. Even if there is not point identification the presence of equality restrictions can lead to the identified set being a manifold with lower dimension than θ. In this case inward bias due to sampling variation in estimated probabilities will cause analog estimators of the identified set to be empty sets.

It is clearly essential to take the effect of sampling variation in estimated probabilities into account and to this end we employ the procedure introduced in Belloni, Bugni and Chernozhukov (2018) (BBC18) to calculate 95% confidence regions for projections of the 9-element θ onto each of its axes in turn. The procedure is designed to perform well for inference on low dimensional functions of parameters in high dimensional parametric settings when identified sets are defined by a large number of moment inequalities, some of which may be equalities. It accommodates point and partially identifying models.

Let the functions of moments required to be nonpositive at a parameter value θ in an identified set be denoted $m_j(\theta)$, $j \in \{1, \ldots, J\}$. Let the core determining sets in $\mathsf{Q}(z, \theta)$ be denoted as follows

$$\mathsf{Q}(z, \theta) = \left\{ (\mathcal{S}_{k(j)}, z_{l(j)}) : j \in \{1, \ldots, J\} \right\}$$

where $\{k(j) : j \in \{1, \ldots, J\}\}$ is a list of indexes identifying unions of residual sets, $\{l(j) : j \in \{1, \ldots, J\}\}$ is a list of indexes identifying values of $Z \in \mathcal{R}_Z$ and each set $\mathcal{S}_{k(j)}$ is a union of residual sets determined by the value of $z_{l(j)}$ and θ.

In the KT16 application there are $J = 112$ inequalities, the moment conditions are

$$\sum_{y \in \mathcal{R}_Y} 1[y \in \mathcal{A}(\mathcal{S}_{k(j)}, z_{l(j)}; \theta)] \times f_{Y|Z}(y|z_{l(j)}) - G_{U|Z}\left(\mathcal{S}_{k(j)}|z_{l(j)}; \theta\right) \leq 0, \quad j \in \{1, \ldots, J\}$$

and the moment functions are defined as

$$m_j(\theta) = \sum_{y \in \mathcal{R}_Y} \left(1[y \in \mathcal{A}(\mathcal{S}_{k(j)}, z_{l(j)}; \theta)] - G_{U|Z}\left(\mathcal{S}_{k(j)}|z_{l(j)}; \theta\right) \right) \times f_{YZ}(y, z_{l(j)}) \leq 0$$

where $f_{YZ}(y, z)$ denotes the *joint* probability that $Y = y$ and $Z = z$.

With data $\{(Y_i, Z_i) : i \in \{1, \ldots, N\}\}$ define

$$m_{ij}(\theta) \equiv \left(1[y \in \mathcal{A}(\mathcal{S}_{k(j)}, z_{l(j)}; \theta)] - G_{U|Z}\left(\mathcal{S}_{k(j)}|z_{l(j)}; \theta\right) \right) \times 1[Y_i = y \wedge Z_i = z_{l(j)}]$$

---

[15]Calculated as 10^{-16}. The calculations were done in R, (R Core Team (2018)). Bivariate Gaussian probabilities were calculated using the `pmvnorm` function in the `mvtnorm` package, Genz and Bretz (2009), Genz et al (2018).

and let $\hat{m}_j(\theta)$ denote the estimator

$$
\begin{aligned}
\hat{m}_j(\theta) &= N^{-1} \sum_{i=1}^{N} m_{ij}(\theta) \\
&= \sum_{y \in \mathcal{R}_Y} \left(1[y \in \mathcal{A}(\mathcal{S}_{k(j)}, z_{l(j)}; \theta)] - G_{U|Z}\left(\mathcal{S}_{k(j)}|z_{l(j)}; \theta\right) \right) \hat{f}_{YZ}(y, z_{l(j)})
\end{aligned}
$$

where

$$
\hat{f}_{YZ}(y, z) = N^{-1} \sum_{i=1}^{N} 1[Y_i = y \wedge Z_i = z]
$$

is a simple relative frequency estimator of the probability that $Y = y$ and $Z = z$. Let $\hat{\sigma}_j^2(\theta) = N^{-1} \sum_{i=1}^{N} (m_{ij}(\theta) - \hat{m}_j(\theta))^2$ denote an estimator of the variance of $N^{1/2}\hat{m}_j(\theta)$.

The $100(1 - \alpha)\%$ confidence region for the projection of the identified set onto the space of an element θ_k of θ is

$$
CI(\theta_k, \alpha) = \left\{ r : \inf_{\{\theta : \theta_k = r\}} \max_{j \in \{1, \dots J\}} \left(N^{1/2} \frac{\hat{m}_j(\theta)}{\hat{\sigma}_j(\theta)} \right) \leq c_N(J, \alpha) \right\}
$$

where $c_N(J, \alpha)$ is the critical value

$$
c_N(J, \alpha) = \frac{\Phi^{-1}(1 - \alpha/J)}{\sqrt{1 - \Phi^{-1}(1 - \alpha/J)^2/N}}.
$$

with Φ denoting the standard Gaussian cumulative distribution function.

The 95% regions are shown in columns 3 and 4 of Table 3.[16] Both approaches IC2 and IC3 were employed with identical results.[17] The data are quite informative about the values of the parameters in the context of the KT16 model.[18] The effect of the presence of each airline type on the profits of the other is negative and significantly different from zero. Higher market presence is associated with higher profits with positive estimates significantly different from zero. The 95% regions accord quite well with the graphs of marginal posterior distributions of the parameters shown in KT16.

Corollary 2 of CR17 shows which of the inequalities defining the identified set of parameter

---

[16]Values of ρ very close to 1 arise. We have taken care to use functions to calculate Gaussian probabilities that are robust to singularity of the covariance matrix, specifically the `pmvnorm` function in the package `mvtnorm` in R. We obtain similar results when in the optimization with respect to elements of θ in the calculation of the 95% regions optimization is done with respect to $\tanh^{-1}(\rho)$ which lies in $(-\infty, \infty)$.

[17]Even though the 95% regions for Δ_{LCC} and Δ_{OA} contain only negative values it is necessary to allow for the possibility of positive values when calculating the regions because in minimizing over values of θ in the calculation of the region positive values of these competition parameters may be found.

[18]In just one case the simple distance-measure-minimizing value of a parameter (β_{OA}^{pres}) does not lie in the 95% region. But this can happen because the estimated moment functions are not weighted in calculating the simple distance measure whereas they are weighted by estimates of the accuracy of their estimation in computing the confidence regions.

values are equalities at any parameter value. In the KT16 model the result is that unless $\Delta_{LCC} > 0$ and $\Delta_{OA} > 0$ the model is complete for the outcome $Y^+ \equiv Y_{LCC} + Y_{OA}$, the number of airline types on a route. This accords with the analysis in Bresnahan and Reiss (1990, 1991) for the coherent case. With the Gaussian restriction on the distribution of the unobservables a maximum likelihood (ML) estimate can be calculated using (Y^+, Z) data as long as the parameter vector θ is point identified using the information contained in the distribution of Y^+ and Z, which it appears to be.[19]

ML estimates are shown in column 5 of Table 3. Strikingly the ML estimate of Δ_{LCC} is *positive* indicating that an *incoherent* structure generates the data.[20] Calculating a Wald test of $H_0 : \Delta_{LCC} = 0$ delivers a p-value smaller than 0.001. A constrained ML estimate holding $\Delta_{LCC} = 0$ is reported in column 6 of Table 3 and it can be seen that the maximized log likelihood function, $\max_\theta LL(\theta)$, is significantly higher absent the constraint. The constrained ML estimate of ρ is positive while the unconstrained estimate is negative.

These ML results using information on only the number of airline types working a route are at odds with the results obtained using all the information in the data which reveals the identity of the airline types on a route in addition to the number of airline types. The 95% confidence regions on the competition parameters using Y data rather then Y^+ data clearly exclude positive values for these parameters. This suggests that there is much to be gained here in using all the information in the data.

These results raise the possibility that the KT16 model is misspecified in some respect. There should not be grossly different results obtained using data aggregated at different levels. To investigate, Information Matrix (IM) tests (White (1982)) were calculated, focussing on each parameter in turn delivering p-values for unconstrained and constrained ML estimates as shown in columns 7 and 8 of Table 3.[21] Three of the IM tests lead to rejection of the specification with the nonnegativity constraint imposed on Δ_{LCC}; only one leads to rejection when the constraint is lifted.

This result and the differences found when using Y^+ information rather than all the information on outcomes lead us to consider estimation under a relaxation of the restrictions of the KT16

---

[19]Using the exact probabilities at the parameter value shown in Table 4 we find that the Information Matrix associated with the likelihood function for the Y^+ outcome is nonsingular at that parameter value indicating the possibility of local point identification, see Rothenberg (1971).

[20]The calculation of the likelihood function takes account of the signs of the competition parameters. When one of these is negative and the other is positive the model is incoherent but remains complete for Y^+ as long as we use approach IC2 which we do in the ML calculation.

[21]IM tests associated with the constant terms are identically zero. Information matrix tests compare elements of two estimates of the Information Matrix associated with a likelihood function. One estimate is the negative of the average of contributions to the matrix of second derivatives of the log likelihood function; the other is the variance-covariance matrix of contributions to the score vector. In the application here the calculation is done as proposed in Chesher (1983) and we use numerical approximations to first and second derivatives of log likelihood contributions calculated using the `grad` and `hessian` functions in R's `numDeriv` package. The Chesher (1983) representation follows from the interpretation in Chesher (1984) of the IM test as a score test of the hypothesis that there is no parameter heterogeneity.

Table 3: Results obtained using the KT16 data and model: column 2 gives results on minimizing a measure of distance from an identified set; columns 3-4 give lower and upper bounds of confidence regions on projections; columns 5 and 7 give the unconstrained MLE and associated IM test p-values; columns 6 and 8 give a constrained MLE and associated IM test p-values

| | | 95% regions | | | | IM tests | |
	$\arg\min_{\theta} \hat{D}(\theta)$	lower bound	upper bound	MLE (std err)	Constrained MLE (std err)	Column 5 p value	Column 6 p value
1	2	3	4	5	6	7	8
β_{LCC}^{cons}	-0.80	-1.01	-0.67	-2.363 (0.0856)	-2.015 (0.0791)	-	-
β_{LCC}^{size}	0.39	0.19	0.50	-0.173 (0.1313)	0.130 (0.0837)	0.178	0.195
β_{OA}^{cons}	0.51	0.39	0.59	0.553 (0.0476)	0.501 (0.0334)	-	-
β_{OA}^{size}	0.52	0.44	0.54	0.466 (0.0419)	0.463 (0.0429)	0.198	0.010
β_{LCC}^{pres}	1.64	1.57	1.95	1.923 (0.2488)	2.216 (0.1484)	0.667	0.007
β_{OA}^{pres}	0.38	0.44	0.64	0.374 (0.0419)	0.437 (0.0363)	0.285	0.336
Δ_{LCC}	-1.44	-1.65	-1.20	1.262 (0.3927)	0.000 (constr)	0.029	-
Δ_{OA}	-1.37	-1.54	-1.07	-1.385 (0.1591)	-1.273 (0.1202)	0.664	0.480
ρ	0.95	0.84	0.99	-0.326 (0.2352)	0.432 (0.0925)	0.144	0.080
$\min_{\theta} D(\theta)$	0.43						
$\max_{\theta} LL(\theta)$				-5709.1	-5712.6		

model. Many relaxations could be considered. Here we ask what information about the parameters is provided by the data when the bivariate Gaussian restriction is dropped while maintaining the independence of U and Z and the linear index restrictions. That is the subject of Section 6.

First we give results obtained using the M02 model and data.

5.2 M02 model and data

The M02 data comprise realizations of Y_L and Y_H, the number of respectively low and high quality motels and values of exogenous variables in the mid 1990's at 492 small rural exits (markets) on 30 US Interstate Highways at which at least one motel was located. In the empirical analysis reported here the values of Y_L and Y_H are censored with a value of 2 indicating 2 or more motels.

We use a much simplified version of the model employed in M02 in which the profit made by a motel operator of type $T \in \{H, L\}$ on opening an additional motel of type T at exit m when there

Table 4: Value of the parameter vector used in simulations and exact probability calculations for the KT16 model

Parameter	Parameter value
β_{LCC}^{cons}	-1.0
β_{LCC}^{size}	0.5
β_{OA}^{cons}	1.0
β_{OA}^{size}	0.5
β_{LCC}^{pres}	1.5
β_{OA}^{pres}	0.5
Δ_{LCC}	-1.5
Δ_{OA}	-1.3
ρ	0.7

are $Y_{m,H}$ high and $Y_{m,L}$ low quality motels already present is:

$$\beta_{T0} + \beta_{T1} Z_{m,1} + \beta_{T2} Z_{m,2} + \alpha_{TL} Y_{m,L} + \alpha_{TH} Y_{m,H} + U_{m,T}$$

and $Z_{m,1}$ and $Z_{m,2}$ are values of binary explanatory variables, Z_1 and Z_2. The variable $Z_{m,1}$ is equal to 1 (0) if at exit m the variable $SPACING$ defined in M02 is above (below) its median value taken across all exits. The variable $SPACING$ is defined in M02 as "The distance in miles from the market exit to the closest exits along the highway with motels (the sum of the distance to the closest competitors on either side)".[22] The variable $Z_{m,2}$ is equal to 1 (0) if at exit m the variable $PLACEPOP$ defined in M02 is above (below) its median value. The variable $PLACEPOP$ is defined in M02 as the population of the town nearest the market exit.

The M02 data contains no information about intersections with no motels. We accommodate this by assuming that sampling is from the truncated distribution of (Y, Z) which has support $((\{0, 1, 2, ...\} \times \{0, 1, 2, ...\}) \backslash (0,0)) \times \mathcal{R}_Z$ so that observable realizations of (Y, Z) are delivered by realizations of U from the truncated distribution

$$G_{U|Z}(\cdot|z) = \frac{G_U(\cdot)}{1 - G_U(\mathcal{U}((0,0), z; \theta))}, \quad U \in \mathcal{R}_U \backslash \mathcal{U}((0,0), z; \theta).$$

The same truncation approach is taken in M02.[23] The distribution of $U = (U_L, U_H)$ is restricted to be mean zero Gaussian independent of $Z = (Z_1, Z_2)$ on their joint pre-truncation support. Truncation induces dependence between U and Z. The variance matrix is parameterized with variances σ_{LL} and σ_{HH} and correlation ρ.

A selection mechanism is specified in M02 which renders the model complete and ML estimation

---

[22] Page 232 of M02.

[23] In M02 footnote 14 it is noted that because there are no $(0,0)$ outcomes recorded in the data, the probabilities computed are conditional on unobservables such that firms of at least one type would find it profitable to enter.

is conducted. Here no selection mechanism is specified. The data carry no record of a highway intersection for which $Y = \phi$, so in dealing with incoherent structures approach IC1 is not available and we choose to consider only approach IC2 which specifies sampling from a truncated distribution of the unobservables excluding support regions on the support of U on which $\mathcal{Y}(u, z; \theta) = \emptyset$. There are thus potentially two sources of truncation.

As in the analysis of the KT16 data we start by considering the simple distance measure (5.1) and seek parameter values that minimize this measure. Results are shown in Table 5. The distance measure is minimized at a parameter value in which the correlation between U_L and U_H is $\rho = 1$. The minimized distance measure is equal to 1.13. The distance-measure-minimizing values of the competition parameters $\alpha_{TT'}$ are all negative - see column 3 of Table 5.

In M02 ρ is constrained to be zero. When we impose this constraint the minimized distance measure is equal to 15.72, much larger than the value obtained when ρ is allowed to vary freely. With ρ constrained to be zero the distance-measure-minimizing parameter value for α_{LH} is *positive*, indicating an incoherent structure; see column 4 of Table 5. The MLEs of the cross-type competition parameters reported in M02 with ρ constrained to be zero are all close to zero compared with the reported estimated standard errors.

The final three columns of Table 5 show distance-measure-minimizing parameter values with the correlation ρ allowed to vary freely and with α_{LH} and α_{HL} constrained to be zero.[24] Here very different values of the parameters lead to very similar values of the minimized distance measure. In all the cases reported there is a similar value of ρ around 0.4. The minimized distance measure takes values close to 18.34, somewhat larger than is obtained when α_{LH} and α_{HL} are not constrained and considerably larger than is obtained when ρ is also unconstrained.

We can find no parameter values that deliver a *zero* distance measure and we conclude that the analog estimate of the identified set is *empty*. This set is characterized by a large number of inequality and equality restrictions (916 in total) and the equality restrictions cannot be satisfied using the estimates of probabilities of Y given Z delivered by finite samples. It is possible that the M02 model, despite being incomplete and potentially incoherent, is point identifying with a singleton identified set. If this is indeed the case, then the inequality and equality restrictions would be satisfied using exact probabilities generated by the M02 model. It may simply be that the analog set estimate is empty due to the inward bias of this estimate when employing estimated probabilities.

It is clear that allowance must be made for the effect of sampling variation and to this end we calculate 95% confidence regions for each parameter in turn using the BBC18 procedure described in Section 5.1. Only the 95% regions for the own type competition parameters α_{LL} and α_{HH} are

---

[24]Optimization was performed using R package `nloptr`, Ypma (2018), Johnson (2007–2019).

informative and these are unbounded below but bounded above by negative values, thus.

$$CI(\alpha_{LL}, 0.05) = (-\infty, -0.16) \qquad CI(\alpha_{HH}, 0.05) = (-\infty, -0.18)$$

So the competition parameters can be signed with some confidence but the data tells us little about the magnitudes.

Thus when a selection mechanism is not imposed the M02 data and model produce an empty analog set estimate. When sampling variation is taken into account the data and model appear quite uninformative about parameter magnitudes, although the confidence sets nonetheless sign the competition parameters. There are several potential contributing factors to the increase in the span of the confidence intervals relative to the empty analog estimates. Part of the problem here may be the relatively small sample size. Another issue to be considered is that in conducting this analysis, in order to reduce the number of inequalities to be considered, we have introduced more censoring than was done in M02.[25] In addition, we have used a much simplified specification of the profit functions of the operator types and a coarse binary encoding of the exogenous variables. For inference we used the self-normalized critical value specified in BBC18, which is easy to compute, but can be conservative. A potentially significant contributory factor is that the M02 model specifies no exclusion restrictions on the exogenous variables that affect the profits of the two operator types. In the KT16 model there are such restrictions because there are firm-type-specific variables measuring market presence, providing exogenous variation that can shift one type's profit independently of the other. There are no such firm-type-specific variables in the M02 model or data.[26]

6 What can be known of the structural functions absent a parametric specification of the distribution of unobservables?

It is commonplace to find parametric distributional restrictions imposed in structural models of simultaneous discrete choice. In the vast majority of cases, as in K16 and M02, unobserved variables are restricted to have multivariate Gaussian distributions independent of observed exogenous variables. There is rarely a good reason to expect the Gaussian restriction to hold, certainly not exactly, so it is good to have the methods developed here for assessing the sensitivity of inference to particular parametric specifications. The methods are first set out and then applied to the KT16 model and data.[27] We find that absent a parametric distributional restriction the KT16 data has

---

[25] In our analysis $y_T = 2$ indicates 2 or more motels of type T whereas in M02 $y_T = 3$ indicates 3 or more motels of type T.

[26] In exploratory simulations (not reported here) of an M02-type model in which there are effective exclusion restrictions we find that informative confidence regions for all parameters can be found, but only with samples of the order of 10 times larger than that used in M02, closer to the sample size in the KT16 data.

[27] The M02 data is not used in this sensitivity analysis because, for reasons discussed at the end of the last section, the data were already found to produce wide confidence intervals with parametric distributional restrictions in the

Table 5: Results obtained using the M02 data and a version of the model allowing incompleteness: values of parameters that minimize a measure of distance from an identified set - column 3 has no constraints imposed, columns 5-7 have constraints imposed as shown

	Parameter	$\arg\min_\theta D(\theta)$	$\arg\min_\theta D(\theta)$ $\rho = 0$	$\arg\min_\theta D(\theta)$ $\alpha_{LH} = 0, \quad \alpha_{HL} = 0$		
1	2	3	4	5	6	7
1	β_{L0}	0.45	-1.41	-0.33	-0.31	-2.86
2	β_{L1}	2.97	-0.39	0.85	0.81	7.68
3	β_{L2}	0.79	0.03	0.35	0.33	3.28
4	β_{H0}	0.09	-0.34	-0.66	-1.58	-0.92
5	β_{H1}	2.90	2.29	2.41	5.82	3.52
6	β_{H2}	0.81	0.83	0.67	1.62	0.98
7	α_{LL}	-1.88	-1.14	-1.01	-0.97	-9.41
8	α_{LH}	-1.62	1.64	0	0	0
9	α_{HL}	-1.24	-0.73	0	0	0
10	α_{HH}	-1.24	-0.92	-1.18	-2.86	-1.76
11	σ_{LL}	2.22	0.94	0.77	0.69	5.25
12	ρ	1.00	0	0.39	0.39	0.38
13	σ_{HH}	1.64	0.97	0.95	2.72	1.73
	$\min_\theta D(\theta)$	1.13	15.72	18.34	18.34	18.36

little to say about the determinants of competition on airline routes.

First a characterization of an outer region for the structural function h of structures (h, G_U) is given when the restriction that unobserved U and exogenous Z are independently restricted is maintained. This is an outer region in the sense that the projection of the identified set for (h, G_U) onto the space of structural functions is guaranteed to be a subset of the outer region which may contain structural functions which do not lie in the identified set. Then a procedure for determining if a structural function in the outer region is in the identified set is presented for the case in which all observed variables are discrete. The results are applied to assess the sensitivity of our analysis of the KT16 model and data to relaxation of parametric distributional restrictions.

Define $\mathcal{H}^*$, the projection of a set of identified structures onto the space of structural functions.

Definition 3 *The projection of an identified set of structures $\mathcal{M}^*$ onto the space of structural functions is*

$$\mathcal{H}^* \equiv \{h \in \mathcal{H} : \exists \mathcal{G}_U \quad s.t. (h, \mathcal{G}_U) \in \mathcal{M}^*\}$$

In a parametric model with a structural function characterized by a parameter vector θ, the structural function h can be replaced by θ. Only the case in which Restriction A7 holds, that is $U \perp\!\!\!\perp Z$ on their full support, is considered.

---

absence of equilibrium selection restrictions.

6.1 Outer regions for projections

Theorem 2, below, defines an outer region $\bar{\mathcal{H}}^* \supseteq \mathcal{H}^*$ for the projection $\mathcal{H}^*$. The Theorem is a restatement of Corollary 3 of CR17 which showed that for all closed sets $\mathcal{S} \subseteq \mathcal{R}_U$ when $U \perp\!\!\!\perp Z$ the inequalities

$$C_h\left(\mathcal{S}|z\right) \leq 1 - C_h\left(\mathcal{S}^c|z'\right) \tag{6.1}$$

hold for all z and z' in $\mathcal{R}_Z$ for all structural functions $h \in \mathcal{H}^*$ where $\mathcal{S}^c$ denotes the closure of the complement of $\mathcal{S}$. This follows because, when $U \perp\!\!\!\perp Z$, for all sets $\mathcal{S}$, closed or open, and for all $z \in \mathcal{R}_Z$

$$C_h\left(\mathcal{S}|z\right) \leq G_U(\mathcal{S}) \tag{6.2}$$

for all structures $(h, G_U) \in \mathcal{M}^*$ so substituting $\mathcal{S}^c$ for $\mathcal{S}$

$$C_h\left(\mathcal{S}^c|z\right) \leq G_U(\mathcal{S}^c) = 1 - G_U(S)$$

which with (6.2) delivers the inequality (6.1). The system of inequalities obtained from (6.1) using all closed sets on $\mathcal{R}_U$ and all z and z' in $\mathcal{R}_Z$, depending as it does on only the structural function h, provides an outer region for the structural function. The inequalities define an *outer region* for the projection because there may exist h in the region for which there does not exist a *proper* distribution G_U such that for all $\mathcal{S}$, z and z'

$$C_h\left(\mathcal{S}|z\right) \leq G_U(\mathcal{S}) \leq 1 - C_h\left(\mathcal{S}^c|z'\right).$$

Theorem 2 *Let* $\mathsf{K}(\mathcal{R}_U)$ *comprise the collection of all closed sets on* $\mathcal{R}_U$. *If Restrictions A1-A7 hold then the set*

$$\bar{\mathcal{H}}^* = \left\{ h : \forall \mathcal{K} \in \mathsf{K}(\mathcal{R}_U), \quad \sup_{z \in \mathcal{R}_Z} C_h\left(\mathcal{K}|z\right) \leq \inf_{z \in \mathcal{R}_Z} \left(1 - C_h\left(\mathcal{K}^c|z\right)\right) \right\} \tag{6.3}$$

comprises an outer region for the structural function h *in the sense that* $\mathcal{H}^* \subseteq \bar{\mathcal{H}}^*$.

Proof. *The result follows directly from Corollary 3 of CR17.* $\blacksquare$

The set $\bar{\mathcal{H}}^*$ characterized by Theorem 2 is defined by an infinite number of inequalities. Theorem 3 below provides a smaller collection of inequalities which delivers the same outer region as the system of inequalities (6.3) that employs all closed sets. Recall the definition of $\mathsf{U}^*(h, z)$, the collection of sets that can be expressed as unions of the residual sets of structural function h when

$Z = z$.

$$\mathsf{U}^*(h, z) \equiv \left\{ \mathcal{U} \subseteq \mathcal{R}_U : \exists \mathcal{Y} \subseteq \mathcal{R}_{Y|z} \text{ such that } \mathcal{U} = \bigcup_{y \in \mathcal{Y}} \mathcal{U}(y, z; h) \right\}$$

Theorem 3 *Define the set*

$$\mathcal{H}^o = \{h : \forall \mathcal{S} \in \mathsf{U}^*(h, z), \forall (z, z') \in \mathcal{R}_Z \times \mathcal{R}_Z, \quad C_h(\mathcal{S}|z) \leq 1 - C_h(\mathcal{S}^c|z')\}. \tag{6.4}$$

If Restrictions A1-A7 hold then $\mathcal{H}^o = \bar{\mathcal{H}}^*$.

Proof. *Rearranging the inequality in (6.4) and using the definition of $C_h(\cdot|z)$ the sets $\bar{\mathcal{H}}^*$ and $\mathcal{H}^o$ can be expressed as follows.*

$$\bar{\mathcal{H}}^* = \left\{ h : \forall \mathcal{K} \in \mathsf{K}(\mathcal{R}_U), \forall (z, z') \in \mathcal{R}_Z \times \mathcal{R}_Z, \quad \mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{K}|z] + \mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{K}^c|z'] \leq 1 \right\}$$

$$\mathcal{H}^o = \left\{ h : \forall \mathcal{S} \in \mathsf{U}^*(h, z), \forall (z, z') \in \mathcal{R}_Z \times \mathcal{R}_Z, \quad \mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{S}|z] + \mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{S}^c|z'] \leq 1 \right\}$$

Define $\mathcal{U}_{\mathcal{S}}(z, h)$, the union of all residual sets of structural function h contained in a set $\mathcal{S}$ when $Z = z$.

$$\mathcal{U}_{\mathcal{S}}(z, h) \equiv \bigcup_{y \in \{y : \mathcal{U}(y, z; h) \subseteq \mathcal{S}\}} \mathcal{U}(y, z; h)$$

There is, for all $\mathcal{K}$ and z

$$\mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{K}|z] = \mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{U}_{\mathcal{K}}(z, h)|z]. \tag{6.5}$$

Let $h \in \mathcal{H}^o$. Consider any set $\mathcal{K} \in \mathsf{K}(\mathcal{R}_U)$. For any z and z'

$$\begin{aligned}
\mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{K}|z] + \mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{K}^c|z'] &= \mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{U}_{\mathcal{K}}(z, h)|z] + \mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{U}_{\mathcal{K}^c}(z', h)|z'] \\
&\leq \mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{U}_{\mathcal{K}}(z, h)|z] + \mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{U}_{\mathcal{K}}^c(z, h)|z'] \\
&\leq 1.
\end{aligned}$$

Here following CR17 Corollary 3 $\mathcal{U}_{\mathcal{K}}^c(z, h)$ is the closure of the complement of $\mathcal{U}_{\mathcal{K}}(z, h)$. The equality in the first line comes on using (6.5). Because $\mathcal{U}_{\mathcal{K}}(z, h) \subseteq \mathcal{K}$ there is $\mathcal{U}_{\mathcal{K}}^c(z, h) \supseteq \mathcal{K}^c$ and by definition $\mathcal{U}_{\mathcal{K}^c}(z', h) \subseteq \mathcal{K}^c$ so it follows that

$$\mathcal{U}_{\mathcal{K}}^c(z, h) \supseteq \mathcal{K}^c \supseteq \mathcal{U}_{\mathcal{K}^c}(z', h)$$

so

$$\mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{U}_{\mathcal{K}^c}(z', h)|z'] \leq \mathbb{P}[\mathcal{U}(Y, Z; h) \subseteq \mathcal{U}_{\mathcal{K}}^c(z, h)|z']$$

27

which leads to the first inequality above. The second inequality arises because $h \in \mathcal{H}^o$ and $\mathcal{U}_{\mathcal{K}}(z,h) \in \mathsf{U}^*(h,z)$.

It has been shown that for every structural function h in the set $\mathcal{H}^o$ the inequality (6.3) in the definition of the set $\bar{\mathcal{H}}^*$ holds for all sets $\mathcal{K} \in \mathsf{K}(R_U)$ which delivers the result of the Theorem. $\square$

In consequence of this Theorem, in determining the outer region given by Theorem 2 it is sufficient to consider only the inequalities arising from sets $\mathcal{S}$ which are unions of residual sets. Further, for any pair (z, z'), every set $\mathcal{S}$ to be considered is a union of residual sets obtained with $Z = z$, S^c is the complement of the set $\mathcal{S}$, the conditional containment probability on the left of the inequality (6.4) is conditioned on $Z = z$ while the conditional containment probability on the right hand side of that inequality is conditioned on $Z = z'$. Inequalities with $z = z'$ can be ignored as they are always satisfied.

6.2 Calculating a projection when observable variables are discrete

When all observable variables are discrete it is possible to determine if a structural function found to be in the outer region is in fact in the projection $\mathcal{H}^*$. The result is contained in Theorem 4. Some preparatory development is required.

The support of U, $\mathcal{R}_U$, is partitioned into a collection of what are termed *elemental sets*. These sets have the property that every intersection of residual sets taken across all $z \in \mathcal{R}_Z$ can be expressed as a union of elemental sets.

Definition 4 $\mathsf{E}(h) = \{\mathcal{E}_\ell : \ell \in \{1, \dots, L(h)\}\}$ *is a collection of elemental sets when (a)* $\mathsf{E}(h)$ *is a partition of* $\mathcal{R}_U$ *and (b) for every full-dimensional[28] intersection of residual sets,* $\mathcal{U}_{Int}(\mathcal{Y}, \mathcal{Z}; h)$, *where*

$$\mathcal{U}_{Int}(\mathcal{Y}, \mathcal{Z}; h) \equiv \bigcap_{y \in \mathcal{Y}, z \in \mathcal{Z}} \mathcal{U}(y, z; h), \quad \mathcal{Y} \subseteq \mathcal{R}_{Y|z}, \quad \mathcal{Z} \subseteq \mathcal{R}_Z$$

there exists a set of indexes $\mathcal{D}(\mathcal{Y}, \mathcal{Z}; h) \subseteq \{1, \dots, L(h)\}$ *such that*

$$\bigcup_{\ell \in \mathcal{D}(\mathcal{Y}, \mathcal{Z}; h)} \mathcal{E}_\ell = \mathcal{U}_{Int}(\mathcal{Y}, \mathcal{Z}; h).$$

Let $p(h) = \{p_\ell(h) : \ell \in \{1, \dots, L(h)\}\}$ be a collection of constants with $p_\ell(h) \geq 0$, $\ell \in \{1, \dots, L(h)\}$ and $\sum_{\ell=1}^{L(h)} p_\ell(h) = 1$. Each $p_\ell(h)$ will be interpreted as the probability that U takes a value in the elemental set $\mathcal{E}_\ell$.

---

[28] The residual sets are closed sets so they have boundaries which can intersect. These intersections are low dimensional manifolds embeded in the support of U which is required to be continuously distributed so they carry zero probability mass and need not be included in the collection of elemental sets.

Let $\{Q(z, h) : z \in \mathcal{R}_Z\}$ be a core determining collection of sets. These collections could be as defined in Theorem 3 of CR17 or, with some redundancy, they could be the collections of *all* unions of residual sets:

$$Q(z, h) = \left\{ \mathcal{S}(\mathcal{Y}, z, h) : \mathcal{Y} \subseteq \mathcal{R}_{Y|z} \right\}$$

where

$$\mathcal{S}(\mathcal{Y}, z, h) = \bigcup_{y \in \mathcal{Y}} \mathcal{U}(y, z; h) \tag{6.6}$$

denotes a union of residual sets. For each such set define the set of indexes of the collection of elemental sets that are contained in $\mathcal{S}(\mathcal{Y}, z, h)$

$$\mathcal{B}(\mathcal{Y}, z, h) \equiv \left\{ \ell \in \{1, \ldots, L(h)\} : \mathcal{E}_\ell \subseteq \mathcal{S}(\mathcal{Y}, z, h) \right\}.$$

The structural function h is in the projection $\mathcal{H}^*$ if there exist probabilities $p(h)$ such that for all $z \in \mathcal{R}_Z$

$$p_\ell(h) \geq 0, \quad \ell \in \{1, \ldots, L(h)\}$$

$$\sum_{\ell=1}^{L(h)} p_\ell(h) = 1$$

and

$$\sum_{\ell \in \mathcal{B}(\mathcal{Y}, z, h)} p_\ell(h) \geq C_h(\mathcal{S}(\mathcal{Y}, z, h)|z) \tag{6.7}$$

for all sets $\mathcal{S}(\mathcal{Y}, z, \theta)$ in the core determining collection $Q(z, h)$ under consideration.

On the left hand side of (6.7) is the probability that U takes a value in the set $\mathcal{S}(\mathcal{Y}, z, h)$ calculated as the sum of the probabilities on the elemental sets that are contained in $\mathcal{S}(\mathcal{Y}, z, h)$. On the right hand side is the containment probability, that is, the conditional probability that the random residual set delivered by the structural function h is a subset of the set $\mathcal{S}(\mathcal{Y}, z, h)$. For structural functions h in the projection $\mathcal{H}^*$ there may be many probability distributions $p(h)$ that satisfy (6.7). For structural functions $h \notin \mathcal{H}^*$ there will be no probability distribution $p(h)$ that satisfies (6.7).

The probabilities $p(h)$ are subject to linear equalities and inequalities so determining whether a structural function h is in the projection $\mathcal{H}^*$ comes down to determining whether the feasible region of a linear program is non-empty which can be done using Farkas's Alternative as in Theorem 4.

In preparation for the statement of the Theorem recall that identified sets of structures are characterized by inequalities delivered by a collection of core determining sets. Each set in such a collection, $\mathcal{S}(\mathcal{Y}, z, h)$, defined in (6.6), is a union of residual sets determined by a subset, $\mathcal{Y}$, of the support of Y, a value, z, of exogenous Z and the structural function under consideration, h. Let d index the collection of all subsets of the support of Y and let e index the values on the support of

Z.

Theorem 4 *A structural function h is in the sharp projection $\mathcal{H}^*$ if and only if there is a nonnegative solution for v in the following linear program:*

$$\min_{s,t,v} v$$

subject to:

$$
\begin{aligned}
sA + tB &\geq 0 \\
t &\geq 0 \\
s + t \cdot c &\leq v
\end{aligned}
$$

where $v \in \mathbb{R}^1$, $s \in \mathbb{R}^1$, $t \in \mathbb{R}^K$,

$$K \equiv \sum_{z \in \mathcal{R}_Z} \mathrm{card}\left(\mathsf{Q}(h, z)\right)$$

$$A \equiv \iota_{L(h)} = [\,\underbrace{1, \ldots, 1}_{L(h)\ times}\,]$$

$$B \equiv [[B_{ki}]], \qquad k \in \{1, \ldots, K\}, \quad i \in \{1, \ldots, L(h)\}$$

$$B_{ki} \equiv -1[i \in \mathcal{B}(\mathcal{Y}_{d(k)}, z_{e(k)}, h)]$$

$$c \equiv [-C_h(\mathcal{S}(\mathcal{Y}_{d(1)}, z_{e(1)}, h)|z_{e(1)}), \ldots, -C_h(\mathcal{S}(\mathcal{Y}_{d(K)}, z_{e(K)}, h)|z_{e(K)})]$$

where $\{d(1), \ldots, d(K)\}$ and $\{e(1), \ldots, e(K)\}$ are collections of integers such that in the kth of the K inequalities $\mathcal{S}(\mathcal{Y}_{d(k)}, z_{e(k)}, h)$ is the union of residual sets and $z_{e(k)}$ is the value of Z on which there is conditioning.

Proof. *In Border (2019, paragraph 12, Section 1.4) there is the following version of Farkas's Alternative.*[29]

Farkas's Alternative. *Let A be a $m \times n$ real matrix, let B be a $K \times n$ matrix, let $b \in \mathbb{R}^m$ and let $c \in \mathbb{R}^K$. Exactly one of the following alternatives hold. Either there exists $x \in \mathbb{R}^n$ satisfying*

$$
\begin{aligned}
Ax &= b \\
Bx &\leq c \\
x &\geq 0
\end{aligned}
\tag{6.8}
$$

---

[29] In the rendition here K, s and t have been substituted for respectively ℓ, p and q. Kim Border describes the result as "buried" in Farkas (1902).

or there exists $s \in \mathbb{R}^m$ and $t \in \mathbb{R}^K$ satisfying

$$
\begin{align}
sA + tB &\geq 0 \tag{6.9} \\
t &\geq 0 \\
s \cdot b + t \cdot c &< 0.
\end{align}
$$

In the application here $x = p(h)$ which contains the probabilities on the $L(h)$ elemental sets, $n = L(h)$, $m = 1$, $b = 1$, and K, A, B and c are as defined in the Theorem. The equality restriction in (6.8) is the requirement that probabilities on elemental sets sum to 1. The inequality restriction $Bx \leq c$ embodies the restrictions in (6.7). The nonnegativity condition in (6.8) requires probabilities to be nonnegative.

The structural function h is in the projection H^* if and only if probabilities $x = p(h)$ can be found that satisfy (6.8). This can be done if and only if there is no solution to (6.9). There is no solution to (6.9) if and only if the solution to the program of the Theorem has $v \geq 0$ for if there was a solution to (6.9) a negative value could be achieved in the program of the Theorem. ∎

In applications we place a negative lower bound on v, e.g. $v \geq -0.1$, to avoid the unbounded program that otherwise arises when h is not in the projection. Additional linear inequality restrictions on the probabilities can be accommodated by including additional rows in B and additional elements in c. For example if the elemental sets are axis aligned hyperrectangles projecting to form a lattice on each margin[30] then marginal probabilities are easily obtained as sums of elements in $p(h)$ and marginal moments are linear functions of $p(h)$ with precise values depending on how one chooses to measure the locations of elemental sets. One could then, for example, require some odd order marginal moments to be zero in an attempt to impose a degree of symmetry on the distribution of U. Such a strategy could be usefully accompanied by linear restrictions on probabilities to bring about a degree of smoothness, for example by requiring the second differences of marginal probabilities to be smaller in absolute value than some pre-chosen constant.

With some extension Theorem 4 can accommodate truncation of the distribution of U such as occurs in the IC2 and IC4 approaches to incoherence and also in cases like M02 in which there is no data recording certain outcomes by virtue of the sampling process. Let $\mathcal{T}(z, h)$ denote the set of values of U that is truncated when $Z = z$. With

$$
\mathcal{C}(\mathcal{Y}, z, h) \equiv \{\ell \in \{1, \ldots, L(h)\} : \mathcal{E}_\ell \subseteq \mathcal{T}(z, h)\}
$$

the inequality (6.7) becomes

$$
\frac{\sum_{\ell \in \mathcal{B}(\mathcal{Y}, z, h)} p_\ell(h)}{1 - \sum_{\ell \in \mathcal{C}(\mathcal{Y}, z, h)} p_\ell(h)} \geq C_h(\mathcal{S}(\mathcal{Y}, z, h)|z)
$$

---

[30] This is the case in the KT16 and M02 examples.

delivering the linear restriction on $p(h)$

$$\sum_{\ell \in \mathcal{B}(\mathcal{Y}, z, h)} p_\ell(h) \geq C_h(\mathcal{S}(\mathcal{Y}, z, h)|z)\left(1 - \sum_{\ell \in \mathcal{CY}, z, h)} p_\ell(h)\right)$$

which can be accommodated by making changes in the matrix B in Theorem 4. Care must be taken to ensure that there is positive probability mass on $\mathcal{T}^c(z, h)$ which can bring additional linear inequalities on board but these are easily accommodated.

6.3 Application to the KT16 model and data

With the distribution of U unrestricted other than requiring U and Z to be independently distributed, location and scale normalizations must be brought on board. To this end the intercepts, β_{LCC}^{cons} and β_{OA}^{cons} are set equal to zero and the coefficients on β_{LCC}^{size} and β_{OA}^{size} are set equal to one. With the Gaussian restriction removed the parameter ρ is irrelevant. There remain 4 parameters that are free to vary: β_{LCC}^{pres}, β_{OA}^{pres}, Δ_{LCC} and Δ_{OA}.

Calculating an analog estimate of the outer region for the structural function, $\mathcal{H}^o$ of Theorem 3, using the KT16 data delivers an *empty set*. This is undoubtedly the consequence of inward bias due to sampling variation in the estimated probabilities of Y given Z which are employed to calculate estimates of the containment probabilities that appear in the 784 inequalities that determine the outer region.[31]

In order to improve our understanding we simulated data from a structure admitted by the KT16 model using a parameter value at which the structure is incomplete and coherent.[32] The parameter value is shown in Table 4. Even with very large sample sizes, exceeding 5 million, analog estimates of the outer region $\mathcal{H}^o$ are empty sets.

We computed *exact* probabilities of Y given Z at the same parameter value used in the simulations as explained in section 5.1. Using the exact probabilities we can calculate the difference between the expressions on the left and right hand sides of the inequality in (6.4). Of course all these differences are nonnegative. The smallest value is very close to zero, 6.2×10^{-5}, and 5% of the 784 differences are smaller than 0.008. Given these tiny magnitudes it is not surprising to find empty analog set estimates even in very large simulated data sets and empty set estimates are to be expected using a sample of the size found in KT16.

Using the exact probabilities obtained using the parameter values shown in Table 4 we can

---

[31] In research underway we are investigating whether a modified version of the BBC18 procedure can be applied to obtain confidence regions for parameter values lying in the outer region. The issue here is that after expressing the inequalities in terms of joint rather than conditional probabilities the estimated moment functions are quadratic functions of estimates of the probabilities $\mathbb{P}[Y = y \wedge Z = z]$ not linear functions to which the BBC18 procedure can be applied directly.

[32] A selection mechanism was imposed in the simulations with random selection of one from multiple potential outcomes such that each potential outcome was equally likely to be chosen.

calculate the exact outer region $\mathcal{H}^o$ and do the calculation proposed in Theorem 4 to determine the projection $\mathcal{H}^*$. We find that the outer region $\mathcal{H}^o$ and the projection $\mathcal{H}^*$ *coincide* and comprise the orthant of the 4 dimensional parameter space in which $\beta_{LCC}^{pres} > 1$, $\beta_{OA}^{pres} > 0$, $\Delta_{LCC} < 0$ and $\Delta_{OA} < -1$. Dropping the parametric distribution restriction we are able to sign the four coefficients but the data carry almost no other information about the magnitudes of the parameters.

One reason for this disappointing result is the very coarse grouping of the data on the three exogenous variables. In KT16 the exogenous variables are each binary indicating whether or not the value under consideration is above or below the median value in the data and these binary representations of the exogenous variables have been used here. We have calculated the projection onto the space of structural function parameters using exact probabilities for a case in which the exogenous variables have richer support.

We generated these exact probabilities using the parameter value shown in Table 4 and for exogenous variables with 5 points of support each located at the sextiles of the values recorded in the KT16 data. For each variable the sextiles were scaled to take values around 1 by dividing each sextile value by the mean of the sextile values for that variable. The trivariate marginal distribution of the 3 exogenous variables located on 125 points of support was specified as uniform in this exploratory exercise. We find that in the exact projection onto the space of structural function parameters the coefficients on the exogenous variables β_{LCC}^{pres} and β_{OA}^{pres} are restricted to finite values while the competition parameters are bounded above by negative values and unbounded below.[33]

Projections onto the space of each of the 4 parameters in turn are shown in Table 6. Note that the parameter values shown here are those in Table 4 divided by 0.5, the common value of the coefficients on β_{LCC}^{size} and β_{OA}^{size} which are normalized to equal 1 in this exercise. The 6 panes of Figure 5 show projections onto the space of each pair of parameters.[34] The parameter values that generated the exact probabilities are plotted as well. By the nature of the projection the same probabilities could have been generated by any parameter value in the projection coupled with one or more proper probability distributions for unobserved U independent of exogenous Z. Clearly richer support for the exogenous variables increases the information about the structural parameter absent a parametric restriction on the distribution of the unobservables, but the projection onto the space of the parameters Δ_{LCC} and Δ_{OA} remains unbounded below.

Direct calculation shows that with richer support for the exogenous variables the sharp projection is a proper subset of the outer region. In the case studied here the outer region is characterized by $217,000$ inequalities. To determine whether a parameter value lies in the sharp projection using

---

[33] The linear program of Theorem 4 is solved using the R package `lpSolveAPI` (Konis and Schwendinger (2020)) which provides an interface to `Lp_solve v 5.5` (Berkelaar and others (2015)).

[34] These are convex hulls of points found to lie in the projections. There appear to be some very slight nonconvexities but allowing for them would make little difference to the impression this gives of the information content of the model concerning structural function parameters absent a parametric distributional restriction.

Table 6: Parameter values and lower and upper bounds of projections of the identified set onto the space of each parameter in turn absent a parametric specification of the distribution of unobservables. Exact probabilities are used in the calculations. Each exogenous variable has 5 points of support. Note that parameter values here are adjusted for normalization of the coefficients on one of the exogenouos variables.

Parameter	Parameter value	Lower bound	Upper bound
β_{LCC}^{pres}	3.0	2.83	3.48
β_{OA}^{pres}	1.0	0.96	1.25
Δ_{LCC}	-3.0	$-\infty$	-0.65
Δ_{OA}	-2.6	$-\infty$	-1.21

the method set out in Theorem 4 requires solving a linear program with 1752 decision variables and 2602 constraints which takes about 7 seconds on a typical desktop computer.

7 Concluding remarks

This paper has shown how identification analysis of structural models of simultaneous discrete choice can be done under alternative approaches to model incompleteness and incoherence using the unifying Generalized Instrumental Variable framework introduced in Chesher and Rosen (2017). To enable assessment of the sensitivity of inference to parametric distributional restrictions outer regions for projections of the identified set of structures onto the space of structural functions or parameters have been characterized. A simple method for determining which elements of an outer region lie in a projection has been presented for the case in which observed variables are discrete.

Applying the results to two models and accompanying data sets delivers mixed results.

Analysis of the data of Mazzeo (2002) using a cut down version of the model produces an empty analog set estimate but little information about the magnitudes of structural parameters when sampling variation is taken into account. A likely contributory factor is the lack of firm-type specific variables in the data. The approach in Mazzeo (2002) involves imposing a selection mechanism leading to a complete model whose parameters are estimated by maximum likelihood. This delivers some quite accurate parameter estimates, but our results using the incomplete partially identifying model suggest they may be sensitive to the equilibrium selection restriction as well as to the functional form and distributional restrictions. Nonetheless, the confidence intervals obtained using the incomplete model continue to deliver negative signs for strategic interaction parameters.

Analysis of the data of Kline and Tamer (2016) produces more nuanced results. The Kline and

Tamer (2016) model appears to be point identifying with the data to hand and maximum likelihood estimators of parameters can be calculated. They indicate that the data are generated by a complete and incoherent structure. Using all the information in the data employing the methods we propose and calculating confidence regions on projections of the identified set onto each parameter axis in turn gives a strong indication that the data generating structure is coherent and incomplete. These conflicting messages together with specification tests of the likelihood based model suggest a degree of misspecification. We investigate what can be known once the Gaussian restriction is dropped.

Dropping the Gaussian restriction on the distribution of the unobservables and calculating an analog estimator of the outer region for the parameters of the profit functions delivers an empty set. This is very likely due to inward bias caused by sampling variation in probability estimates that appear in over 700 inequalities defining the outer region.

Using exact probability distributions of observable variables calculated for the Kline and Tamer (2016) model at parameter values of the order suggested by the parametric distribution based empirical analysis we calculate the exact outer region and the exact projection for structural parameters obtained with the Gaussian restriction dropped. Absent the parametric distributional restriction very wide ranges of parameter values can deliver the probability distributions. Key structural parameters can be signed but rather little can be said about their magnitudes and in particular competition parameter values are unbounded below. Using exact probabilities in a case in which exogenous variables have richer support than is found in the Kline and Tamer (2016) application gives an identified set which contains only finite values for the coefficients on exogenous variables and with competition parameters negative but unbounded below.

In research underway we seek to determine whether introducing smoothness and shape restrictions on the distribution of unobserved variables, stopping short of parametric restrictions can deliver more encouraging results. We are also exploring another way of relaxing the model's restrictions taking a single equation approach. In the context of the KT16 model this involves estimating a binary outcome model for each airline type's entry separately using variables affecting the profits of the competitor type as instrumental variables for the binary endogenous explanatory variable recording the presence of the competitor type, similar to the instrumental variable binary outcome models applied in Section 8.2 of Chesher and Rosen (2020) and the single equation IV ordered outcome model employed in Chesher, Rosen, and Siddique (2019).

REFERENCES

AMEMIYA, T. (1974): "Multivariate Regression and Simultaneous Equation Models when the Dependent Variables are Truncated Normal," *Econometrica*, 42(6), 999-1012.

BAJARI, P., H. HONG, AND S. P. RYAN (2010): "Identification and Estimation of Discrete Games of Complete Information," *Econometrica*, 78(5), 1529-1568.

BELLONI, A., F. BUGNI, AND V. CHERNOZHUKOV (2018): "Subvector Inference in PI Models with Many Moment Inequalities," arXiv:1806.11466.

BERESTEANU, A., I. MOLCHANOV, AND F. MOLINARI (2011): "Sharp Identification Regions in Models with Convex Moment Predictions," *Econometrica*, 79(6), 1785-1821.

BERKELAAR, M., AND OTHERS (2015). lpSolve: Interface to 'Lp_solve' v. 5.5 to Solve Linear/Integer Programs. R package version 5.6.13. `https://CRAN.R-project.org/package=lpSolve`

BERRY, S.T., (1992): "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 60(4), 889-917.

BJORN, P., AND Q. VUONG (1984): "Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation," CIT working paper, SSWP 537.

BLUNDELL, R., AND R. J. SMITH (1994): "Coherency and Estimation in Simultaneous Models with Censored or Qualitative Dependent Variables," *Journal of Econometrics*, 64, 355-373.

BORDER, K.C., (2019): "Alternative Linear Inequalities," Caltech, version 2019.10.23::09.57, manuscript. `http://www.hss.caltech.edu/~kcb/Notes/Alternative.pdf`.

BRESNAHAN, T.F., AND P.J. REISS (1990): "Entry in Monopoly Markets," *Review of Economic Studies*, 57, 531-553.

BRESNAHAN, T.F., AND P.J. REISS (1991): "Empirical Models of Discrete Games," *Journal of Econometrics*, 48(1-2), 57-81.

CHEN, X., CHRISTENSEN, T., AND TAMER, E., (2018): "Monte Carlo Confidence Set for Identified Sets," *Econometrica*, 86(6), 1965-2018.

CHRISTENSEN, T. AND B. CONNAULT (2019): "Counterfactual Sensitivity and Robustness," arXiv working paper 1904.00989v2.

CHESHER, A., (1983): "The Information Matrix Test: Simplified Calculation via a Score Test Interpretation," *Economics Letters*, 13(1), 45-48.

CHESHER, A., (1984): "Testing for Neglected Heterogeneity," *Econometrica*, 52, 865-872.

CHESHER, A., AND A.M. ROSEN (2012): "Simultaneous equations for discrete outcomes: coherence, completeness, and identification," CeMMAP Working Paper, CWP21/12.

CHESHER, A., AND A.M. ROSEN (2017): "Generalized Instrumental Variable Models," *Econometrica*, 85, 959-989.

CHESHER, A., AND A.M. ROSEN (2020): "Generalized Instrumental Variable Models, Methods and Applications," in *The Handbook of Econometrics, Volume 7a*, ed., Durlauf, S., L.P. Hansen, J.J. Heckman and R. Matzkin, Elsevier.

CHESHER, A., A.M. ROSEN AND Z. SIDDIQUE (2019): "Estimating Endogenous Effects on Ordinal Outcomes," CeMMAP Working Paper, CWP66/19.

CHESHER, A., A.M. ROSEN AND K. SMOLINSKI (2013): "An Instrumental Variable Model of Multiple Discrete Choice," *Quantitative Economics*, 4(2), 157-196.

DAGENAIS, M.C. (1997): "A Simultaneous Probit Model," working paper, University of Montreal.

FARKAS, J., (1902): "Über die Theorie der einfachen Ungleichungen," *Journal für die Reine und Angewandte Mathematik* 124:1-27.

GENZ,A., AND F. BRETZ (2009): *Computation of Multivariate Normal and t Probabilities. Lecture Notes in Statistics, Vol. 195,* Springer-Verlag, Heidelberg. ISBN 978-3-642-01688-2.

GENZ,A., F. BRETZ, T. MIWA, X. MI, F LEISCH, F. SCHEIPL, AND T. HOTHORN (2018): mvtnorm: Multivariate Normal and t Distributions. R package version 1.0-8. `http://CRAN.R-project.org/package=mvtnorm`.

GILBERT, P., AND R. VARADHAN (2019): numDeriv: Accurate Numerical Derivatives. R package version 2016.8-1.1. `https://CRAN.R-project.org/package=numDeriv`.

GOURIEROUX, C., J. J. LAFFONT, AND A. MONFORT (1980): "Coherency Conditions in Simultaneous Linear Equation Models with Endogenous Switching Regimes," *Econometrica*, 48(3), 675-695.

HAJIVASSILIOU, V. (2008): "Novel Approaches to Coherency Conditions in LDV Models," working paper, LSE.

HECKMAN, J. J., (1978): "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 46, 931-959.

JOHNSON, S. G. (2007–2019): "The NLopt Nonlinear Optimization Package," `http://github.com/stevengj/nlopt`.

KAIDO, H., MOLINARI, F., AND STOYE, J., (2019): "Confidence Intervals for Projections of Partially Identified Parameters," *Econometrica*, 87(4), 1397-1432.

KLINE, B., AND TAMER, E., (2016): "Bayesian Inference in a Class of Partially Identified Models," *Quantitative Economics*, 7(2), 329-366.

KOOREMAN, P. (1994): "Estimation of Econometric Models of Some Discrete Games," *Journal of Applied Econometrics*, 9, 255-268.

LEWBEL, A. (2007): "Coherency and Completeness of Structural Models Containing a Dummy Endogenous Variable," *International Economic Review*, 48(4), 1379-1392.

KONIS, K., AND SCHWENDINGER, F. (2020). lpSolveAPI: R Interface to 'lp_solve' Version 5.5.2.0. R package version 5.5.2.0-17.6 `https://CRAN.R-project.org/package=lpSolveAPI`.

MADDALA, G.S., (1983): *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge University Press, Cambridge, U.K.

MANSKI, C.F.,(1990): "Nonparametric Bounds on Treatment Effects," *American Economic Review: Papers and Proceedings*, 80(2), 319-323.

MANSKI, C.F., (2003): *Partial Identification of Probability Distributions.* Spring-Verlag, New York.

MAZZEO, M., (2002): "Product Choice and Oligopoly Market Structure," *The RAND Journal of Economics*, 33(2), 221-242.

McFadden, D., (1974): "Conditional Logit Analysis of Qualitative Choice Behaviour," in *Frontiers in Econometrics* ed. by P. Zarembka, Academic Press, New York.

Molchanov, I., and F. Molinari (2018): *Random Sets in Econometrics*, Cambridge University Press, Cambridge, U.K.

Molinari, F., (2020): "Econometrics with Partial Identification," in *The Handbook of Econometrics, Volume 7a*, ed., Durlauf, S., L.P. Hansen, J.J. Heckman and R. Matzkin, Elsevier. Working paper version available at `https://www.ifs.org.uk/uploads/CW2519_Econometrics_with_Partial_Identification.pdf`.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. `https://www.R-project.org/`.

Rothenberg, T.J., (1971): "Identification in Parametric Models," *Econometrica*, 39(3), 577-591.

Schmidt, P. (1981): "Constraints on Parameters in Simultaneous Tobit and Probit Models," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. F. Manski, and D. L. McFadden, pp. 422-434. MIT Press.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Models with Multiple Equilibria," *Review of Economic Studies*, 70(1), 147-167.

Tamer, E. (2010): "Partial Identification in Econometrics," *Annual Review of Economics*, 2, 167-195.

White, H., (1982): "Maximum Likelihood Estimation of Misspecified Models, " *Econometrica*, 50(1), 1-25.

Ypma, J., (2018): "Introduction to nloptr: an R interface to NLopt, " `https://cran.r-project.org/web/packages/nloptr/vignettes/nloptr.pdf`.

Figure 1: Residual sets for the KT16 model of competition on US airline routes for 4 combinations of the interaction parameters Δ_{LCC} and Δ_{OA}. Labels $\mathcal{U}((y_{LCC}, y_{OA}))$ are short hand for $\mathcal{U}((y_{LCC}, y_{OA}), Z; \theta)$, $\{y_{LCC}, y_{OA}\} \in \{0,1\}^2$. Dark shaded areas are regions in which residual sets intersect, $\mathcal{U}((0,1), Z; \theta)$ and $\mathcal{U}((1,0), Z; \theta)$ in the upper left pane, $\mathcal{U}((0,0), Z; \theta)$ and $\mathcal{U}((1,1), Z; \theta)$ in the upper right pane. The unshaded areas in the lower panes are regions in the space of U in which there is no y such that $h(y, z, u) = 0$. Each residual set for a non-null value of Y is unbounded on two sides.
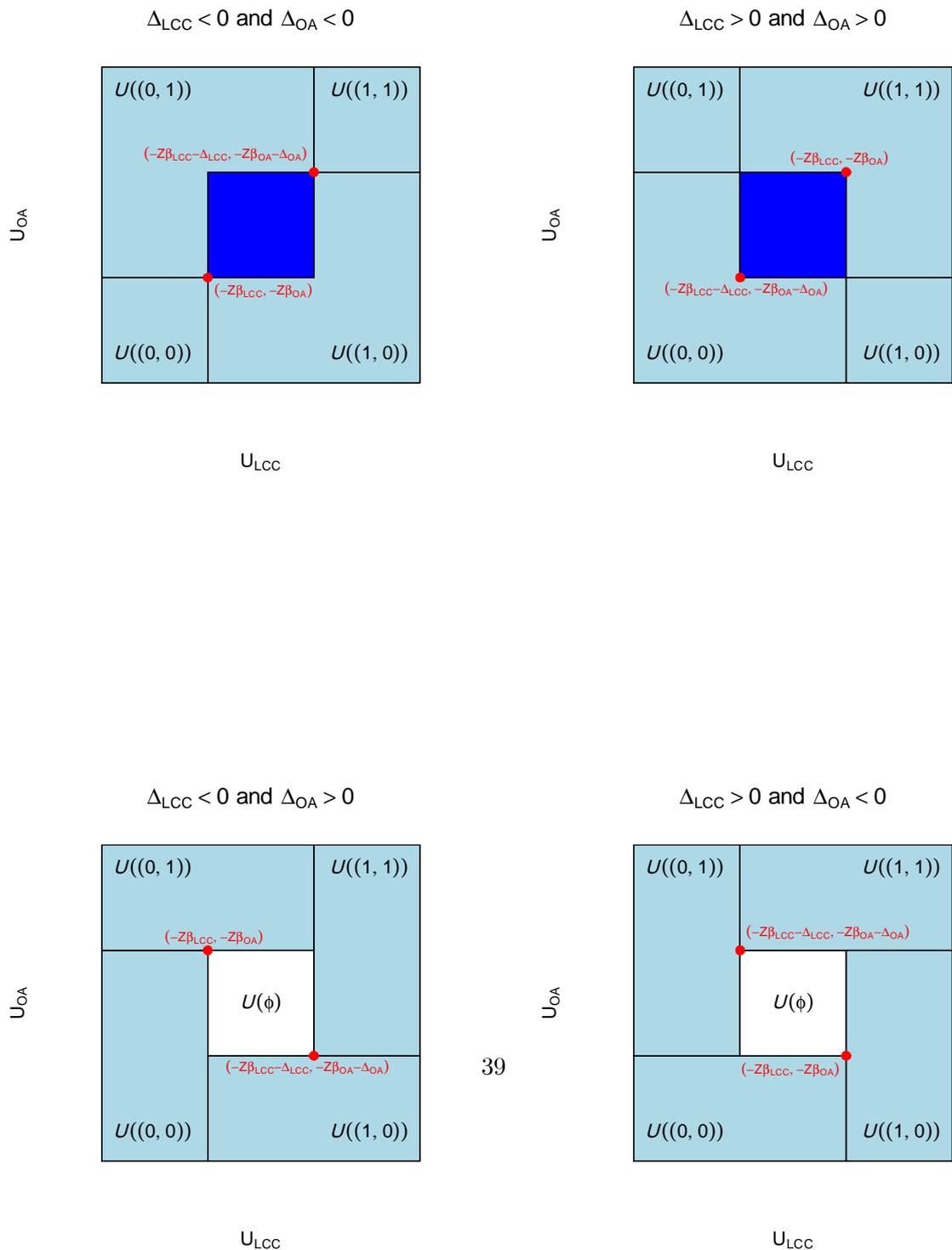

39

Figure 2: Residual sets for the simplied version of the M02 model used here with censoring at $Y_T = 3$, $T \in \{H, L\}$ and $\alpha_{LL} = -1.5$ $\alpha_{LL} = -0.75$ $\alpha_{HL} = -2.0$ $\alpha_{HH} = -0.8$. Labels (y_L, y_H) signify $\mathcal{U}((y_L, y_H), z; h)$ and are plotted at the vertical centre of a residual set. This allows one to tell, for example that $\mathcal{U}((1, 2), z; h)$ is a subset of $\mathcal{U}((2, 0), z; h)$. Residual set boundaries are drawn in various widths so that common boundaries are more easily identified.
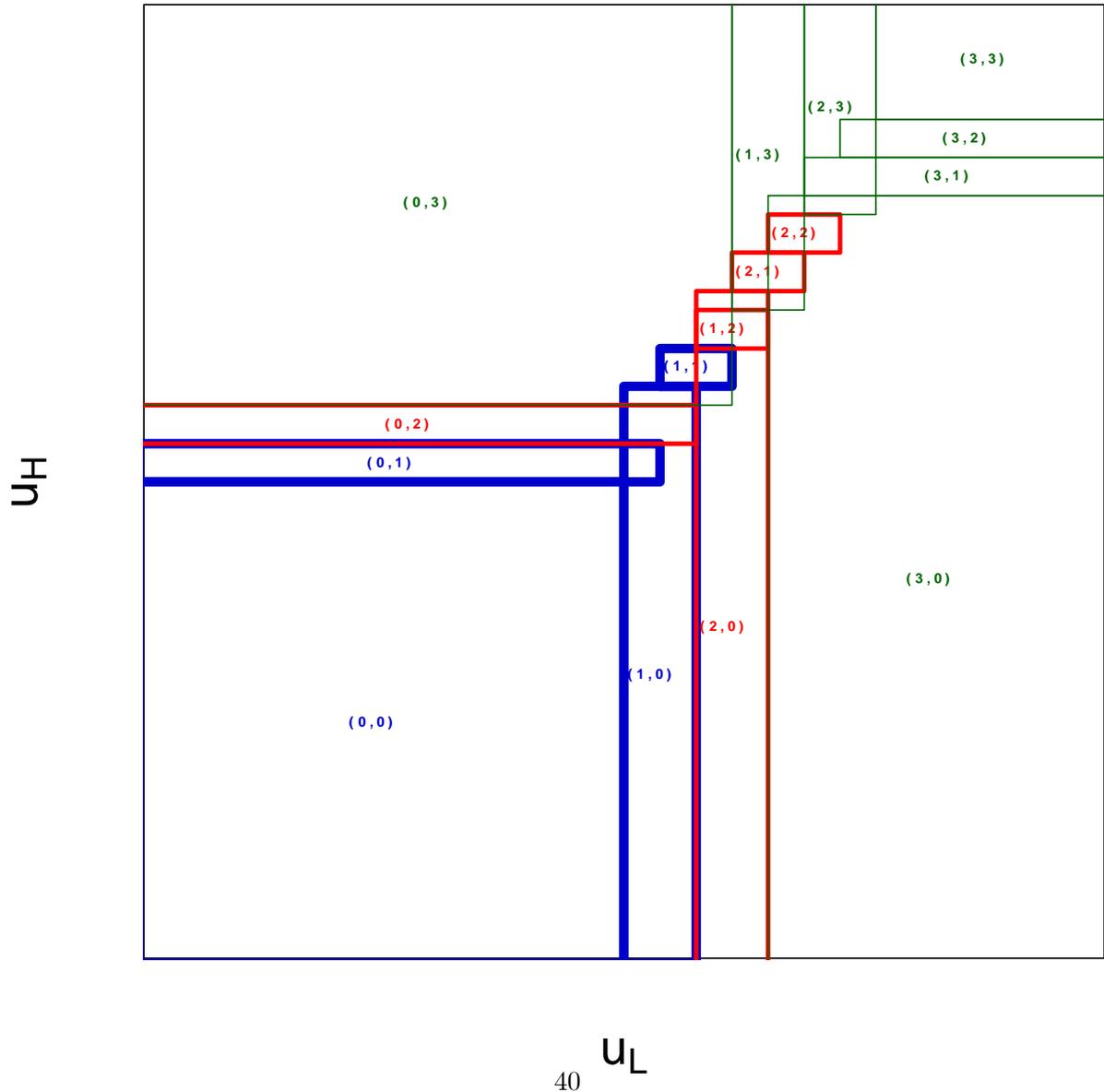
Figure 3: Residual sets for the simplied version of the M02 model used here with censoring at $Y_T = 3$, $T \in \{H, L\}$ and $\alpha_{LL} = -1.3$ $\alpha_{LL} = -1.8$ $\alpha_{HL} = -1.4$ $\alpha_{HH} = -1.9$. Labels (y_L, y_H) signify $\mathcal{U}((y_L, y_H), z; h)$ and are plotted at the vertical centre of a residual set. Residual set boundaries are drawn in various widths so that common boundaries are more easily identified.
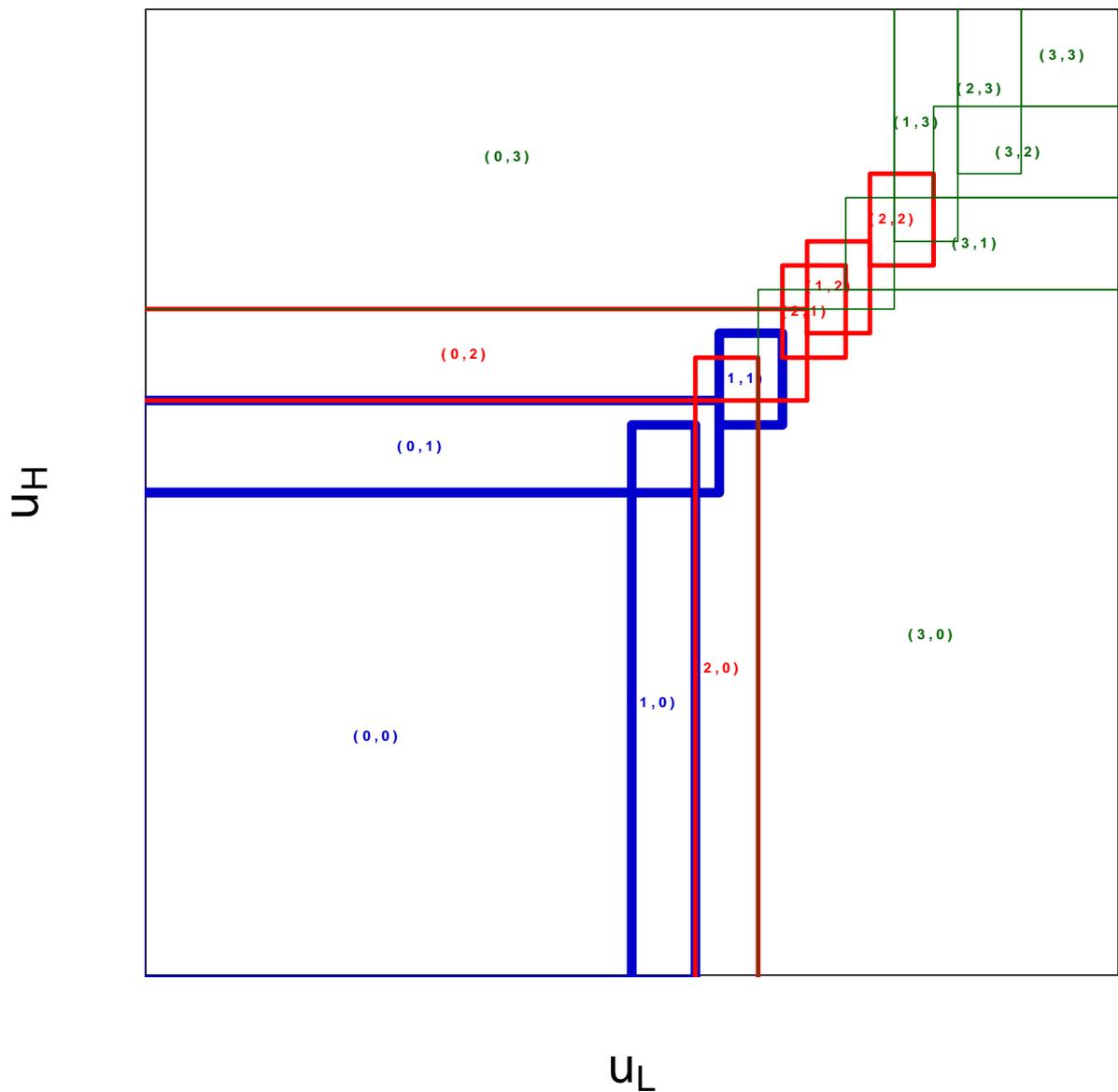
Figure 4: Residual sets for the simplied version of the M02 model used here with censoring at $Y_T = 3$, $T \in \{H, L\}$ and $\alpha_{LL} = -1.8$ $\alpha_{LL} = -1.3$ $\alpha_{HL} = -1.4$ $\alpha_{HH} = -1.8$. Labels (y_L, y_H) signify $\mathcal{U}((y_L, y_H), z; h)$ and are plotted at the vertical centre of a residual set. Residual set boundaries are drawn in various widths so that common boundaries are more easily identified.
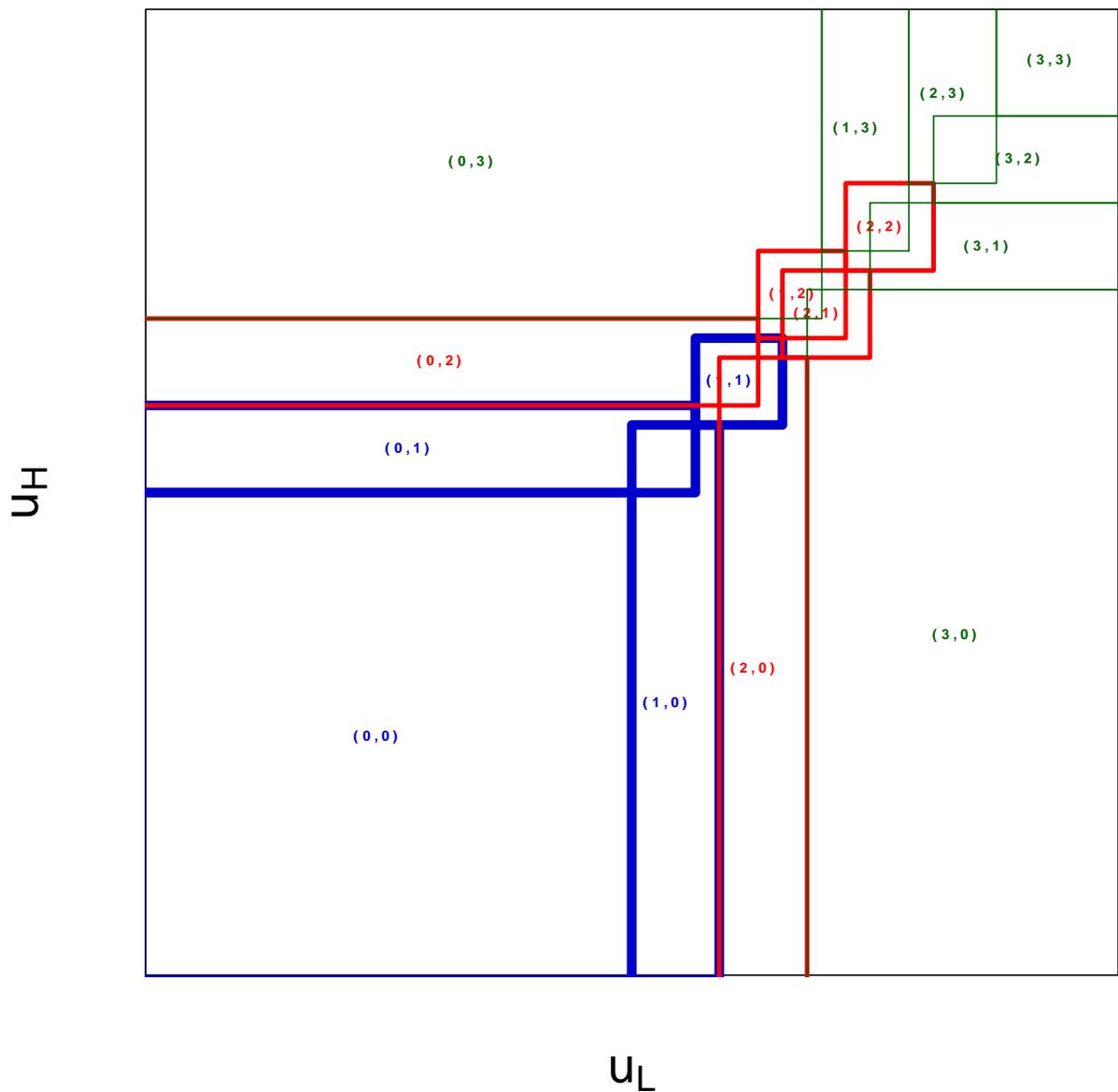
Figure 5: Projections of the identified set onto the spaces of pairs of structural function parameters in the KT16 model absent a parametric specification of the distribution of unobservables. The 3 exogenous variables each take 5 values and Z has 125 points of support each occurring with equal probability. The plotted points are the parameter values that generated the exact probabilities $\mathcal{F}_{Y|Z}$ used to calculate the identified set. The projections extend down to $-\infty$ for the parameters Δ_{LCC} and Δ_{OA}.