Information and admissible sets

Jeff Rowley

28$^{\text{th}}$ July, 2014

Abstract

«Abstract here»

I explore the effect of incorporating information for a non-parametric binary choice model. The model permits endogenous variation in a scalar random variable that is due to non-random selection, and it is the average causal effect of this endogenous variable on the outcome variable that is of interest. The model embeds an exclusion restriction and an independence restriction that together define an instrumental variable but is silent as to the relationship between the endogenous variable and the instrumental variable. I restrict the relationship between the outcome variable and the endogenous variable up to a non-parametric threshold crossing function. The model is credible (Manski, 2013) in that it embeds only weak non-verifiable restrictions, but does not identify the average causal effect of the endogenous variable on the outcome variable.[1] Rather, the model partially identifies the average causal effect of the endogenous variable on the outcome variable.

I define information to be those additional characteristics of economic agents that are observable with the caveat that these characteristics be exogenous and relevant to the latent structure. It is convenient to think of such characteristics as being predetermined and immutable; characteristics that result from choices that are made jointly with the outcome variable are excluded by the definition. Accordingly, exogenous variables and instrumental variables are each regarded as information, and I distinguish between these classes of information. I study how the admissible set of values for the average causal effect of the endogenous variable on the outcome variable changes as each class of information is incorporated into the model separately.

It is useful to distinguish between classes of information since each class enters the latent structure in a different way. Exogenous variables are permitted to enter the structural equation for the outcome variable and to determine the endogenous variable. As such, exogenous variables can be seen to enrich both individual response and individual selection, respectively. An

---

[1]Assumptions that cannot be tested using data. The model does embed some non-trivial non-verifiable restrictions that might be relaxed.

important consequence is that the causal effect of the endogenous variable on the outcome variable depends upon the value of the exogenous variables when individual response is enriched. In contrast, instrumental variables are excluded from the structural equation for the outcome variable by definition and so only enrich individual selection. Given this, the effect of incorporating information is different depending upon the class of information that is being incorporated into the model.

Incorporating information of either class is generally sensible for a number of reasons. Firstly, incorporating information is known to be efficient; variation that is attributable to an observable variable is instead attributable to unobservable heterogeneity when that variable is omitted. Secondly, the effect of incorporating information for partially identifying models is not well-documented; one hypothesis is that incorporating information narrows bounds on admissible sets. Such an effect is not documented in identifying models precisely because such models deliver a point estimate (a set of length zero), but point estimates may shift as information is incorporated. A contribution that I make is in showing that incorporating information leads to narrower bounds on the admissible sets that are delivered by the model. A further reason to particularly favour incorporating exogenous variables is that the average causal effect of the endogenous variable on the outcome variable in identifiable sub-populations can be recovered. I name this structural characteristic the conditional average causal effect of the endogenous variable on the outcome variable, and index it by the conditioning value.[2] Understanding the effect of an intervention in sub-populations can be interesting if the intervention can be targeted or if the intervention is to be applied elsewhere in a population that differs according to its observable characteristics.

A relevant question is how to relate conditional causal effects to (unconditional) causal effects. More precisely, how does the average causal effect of the endogenous variable on the outcome variable relate to its conditional counterparts? I show that the average causal effect of the endogenous variable on the outcome variable can be expressed as a Minkowski summation of its conditional counterparts when the non-parametric binary choice model is augmented. I derive sharp bounds on the conditional average causal effect by applying random set theory. I employ the capacity (or containment) functional as in Chesher et al. (2013) as a matter of choice, rather than the Aumann expectation as in Beresteanu et al. (2012). As I show that the average causal effect of the endogenous variable on the outcome variable can be expressed as a Minkowski summation of its conditional counterparts, I derive sharp bounds on the average causal effect of the endogenous variable on the outcome variable.[3] I establish the conditions under which bounds on conditional causal effects can be informative about bounds on (unconditional) causal effects. That is, I establish the conditions under which bounds on conditional causal effects can be used to narrow bounds on (unconditional) causal effects, exploiting the mapping from one to the other.

I demonstrate application of the non-parametric binary choice model, elucidating the practical difficulties that arise when estimating set identifying models (focusing on those issues that arise from incorporating information). As in Chesher and Rosen (2013), I estimate the average causal effect of childbirth on a mother's labour force participation using US census data. I extend Chesher and Rosen (2013) in a number of ways. Firstly, I report statistical uncertainty in the estimate of the average causal effect of childbirth on a mother's labour force participation using a method that is outlined in Chernozhukov et al. (2013). Secondly, I enrich the support of the instrumental variable and explore the effect that this has on the admissible set of values for the average causal effect of childbirth on a mother's labour force participation, and on its accompanying confidence region. Thirdly, I enrich individual response by permitting the structural equation for labour force participation to depend upon the age and other such predetermined and immutable characteristics of mothers. I discuss the complication of calculating statistical uncertainty when exogenous variables are permitted to enter the structural equation for labour force participation. With respect to the second and third extensions, it is necessary that I augment the model by embedding additional restrictions. In fact, Chesher and Rosen (2013) describe the augmented non-parametric binary choice model that I assume but simplify this model for application (by

---

[2] The conditioning value is specifically the value of the exogenous variables. Heckman and Vytlacil (2005) defines a parameter $ATE(x)$ that is equivalent to the conditional average causal effect of the endogenous variable on the outcome variable at the conditioning value x. Khan and Tamer (2010) and Abrevaya et al. (2013) instead refer to this parameter as the conditional average treatment effect and abbreviate this to $CATE(x)$.

[3] Molchanov (2005) is a useful companion in the study of random sets.

excluding exogenous variables from the structural equation for the outcome variable). I discuss how the augmented model relates to the simplified model in each case and the credibility of the additional restrictions that are embedded in the augmented model.

Related research

Other notable non-parametric binary choice models are described in Balke and Pearl (1997) and Shaikh and Vytlacil (2011), and general non-parametric models of choice are described in Chesher (2005), Kitagawa (2009) and Chesher (2010).

Balke and Pearl (1997) assumes a triangular model (the model embeds a structural equation for the outcome variable and a structural equation for the endogenous variable; see Strotz and Wold (1960) for a detailed discussion of triangular models) that relaxes separability of unobservable heterogeneity in the structural equation for the outcome variable. The cost is that the model is no longer silent as to the relationship between the endogenous variable and the instrumental variable. The model does not permit exogenous variables to enter the structural equation for the outcome variable. I discuss the credibility of separability of unobservable heterogeneity in the main text. Shaikh and Vytlacil (2011) assumes a triangular model but maintains separability of unobservable heterogeneity in the structural equation for the outcome variable. The model permits exogenous variables to enter the structural equation for the outcome variable.

Chesher (2005) and Kitagawa (2009) describe non-parametric models that permit continuous variation in the outcome variable. Chesher (2005) assumes a triangular model that relaxes separability of unobservable heterogeneity in the structural equation for the outcome variable. The model permits exogenous variables to enter the structural equation for the outcome variable, although local invariance of the structural equation for the outcome variable to variation in the exogenous variables is embedded. The model is uninformative when there is binary variation in the endogenous variable but is informative when there is discrete variation. Kitagawa (2009) extends Balke and Pearl (1997) to permit discrete and continuous variation in the outcome variable, and studies commonly invoked restrictions on covariation of the instrumental variable and unobservable heterogeneity.

Chesher (2010) describes an ordered choice model that permits discrete variation in the outcome variable. Chesher (2010) assumes a single equation model that relaxes separability of unobservable heterogeneity in the structural equation for the outcome variable, although monotonicity of the structural equation for the outcome variable in unobservable heterogeneity is embedded. The model permits binary or discrete variation in the endogenous variable.

Notation

I study a probability space $(\Omega, \Sigma, \mathbb{P})$. I define random variables on this probability space. I write random variables as upper case Latin letters, and I write realisations (or specific values) of random variables as lower case Latin letters. I write the support of A as $\mathcal{R}_A$. I write the counterfactual value of A when B has a causal interpretation and is externally fixed as $A(b)$. I write the average causal effect of B on A as $ACE(B \to A)$, and the conditional average causal effect of B on A given C as $ACE(B \to A|c)$.

I refer to Y as the outcome variable, to D as the endogenous variable, to X as the endogenous variable, to Z as the instrumental variable, and to U as unobservable heterogeneity. Despite the use of *the*, I permit (X, Z) to be vectors. I write the structural equation for the outcome variable as h, and the structural equation for the endogenous variable as g.

I write the expectation operator as $\mathbb{E}$, and the indicator function as $\mathbb{1}$. I write A is independent of B as $A \perp\!\!\!\perp B$. To distinguish between population and sample quantities, I subscript sample quantities by n.

I introduce further terminology and notation in Figure 1 through Figure 4. This specifically relates to models and structures, and is consistent with the approach that is formally laid out in Hurwicz (1950) and in Koopmans and Reiersøl (1950).

Application

I estimate the average causal effect of childbirth on a mother's labour force participation using United States census data. The data are obtainable from Angrist (2014), and are described in Angrist and Evans (1998). To summarise, the dataset consists of 254,654 households that were recorded as part of the 1980 United States census. The dataset specifically contains observations of married households with at least two children under the age of 18 years and where the mother is aged between 21 years and 35 years. For clarity, I translate each variable in the data into the mathematical notation that I employ.

$$Y \equiv \mathbb{1}[\text{Mother is employed in 1979}]$$
$$D \equiv \mathbb{1}[\text{Three or more children in the household}]$$

As (X, Z) are continually redefined in the main text, I do not define these variables as I do (Y, D). Instead, I note those variables that at some point or other form part of the definition of (X, Z). X is a function of mother's race or ethnicity (shortened to race).[4] Z is a function of whether the oldest two children in the household share the same gender, and whether the second pregnancy was a multiple pregnancy.

I refer to the application throughout the main text so as to illustrate how technical conditions on variables and on the relationship between variables restrict the behaviour of economic agents, in this case mothers.

1   A non-parametric model of binary choice

I begin by introducing the non-parametric binary choice model that is described in Chesher and Rosen (2013). The model constitutes the set of structures that are consistent with Restriction M1 through Restriction M4.

Axiom. *Economic agents are utility maximising, selecting between alternatives in a choice set according to the utility that they attach to that choice. Utility is perfectly observable by economic agents and is determined by a well-defined utility function for each choice. Each agent is permitted to value each choice differently.*

M1. Discrete support. (Y, D, X, Z) are observable and have discrete supports (with at least two points of support). Further, (Y, D) have binary supports and are normalised such that

(a.) $\mathcal{R}_Y = \{0, 1\}$ and

(b.) $\mathcal{R}_D = \{0, 1\}$,

respectively.

Restriction M1 is a verifiable restriction. (Y, D, X, Z) are observable and so it is trivial to verify that each variable satisfies its support restriction. (Y, D) are normalised to be consistent with the application, but any other supports $\{y_0, y_1\}$ and $\{d_0, d_1\}$ can be generated by an affine transformation of h and of g.

M2. Scalar U. U is an unobservable scalar such that $\mathcal{R}_U$ is an open subset of $\mathbb{R}$ with strictly positive Lebesgue measure.

Suppose that the utility of choice m is determined by the utility function

$$V_m(O, N)$$

where O is the vector of all observable determinants of utility, and N is the vector of all unobservable determinants. Then if the cardinality of supp is no greater than the cardinality of $\mathbb{R}$ then

---

[4] Angrist and Evans (1998) also treats mother's age, and mother's age at the time of her first birth (shortened to birthing age) as exogenous variables.

there exists a bijection $\text{supp}(N) \mapsto \mathbb{R}$, which has output U. It follows that there exists a utility function $\tilde{V}_m$ such that

$$V_m(O, N) \equiv \tilde{V}_m(O, U).$$

Restriction M2 can then be regarded as a normalisation rather than as a restriction in that U is simply a projection of a high-dimensional vector onto $\mathbb{R}$ that preserves uniqueness.

M3. Joint independence. $U \perp\!\!\!\perp (X, Z)$.

Restriction M2 is a non-verifiable restriction. The restriction is stronger than the marginal independence relations $U \perp\!\!\!\perp X$ and $U \perp\!\!\!\perp Z$ since it imposes extra constraints on the joint distribution $\mathbb{P}(u, x, z)$.

2 Credibility in economic modelling

3 Incorporating information

3.1 Enriching the support of the instrumental variable

3.2 Enriching individual response

4 A non-parametric model of binary choice

I begin by introducing the non-parametric binary choice model that is studied in Chesher and Rosen (2013). I discuss the credibility of the restrictions that are embedded in the model and the limitations of these restrictions. I then augment the model and incorporate information. Firstly, I enrich the support of the instrumental variable. Secondly, I enrich the structural function for the outcome variable by allowing individual response to depend upon exogenous variables as well as on the endogenous variable. In both cases, I discuss the credibility of the restrictions that are embedded in the augmented model and the limitations of these restrictions. I discuss the effect of incorporating both classes of information on the admissible set of values for the average causal effect of the endogenous variable on the outcome variable.

## 4.1 A baseline model

Chesher and Rosen (2013)

$$Y = \mathbb{1}[p(D) > U]$$
$$U \perp\!\!\!\perp Z$$

References

Abrevaya, J., Y.-C. Hsu, and R. P. Lieli (2013). Estimating conditional average treatment effects. Technical report, Working paper.

Angrist, J. D. (2014). Angrist data archive. [Online; accessed 25-July-2014].

Angrist, J. D. and W. N. Evans (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review 88* (3), $450 - 477$.

Balke, A. and J. Pearl (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association 92* (439), $1171 - 1176$.

Beresteanu, A., I. Molchanov, and F. Molinari (2012). Partial identification using random set theory. *Journal of Econometrics 166* (1), $17 - 32$.
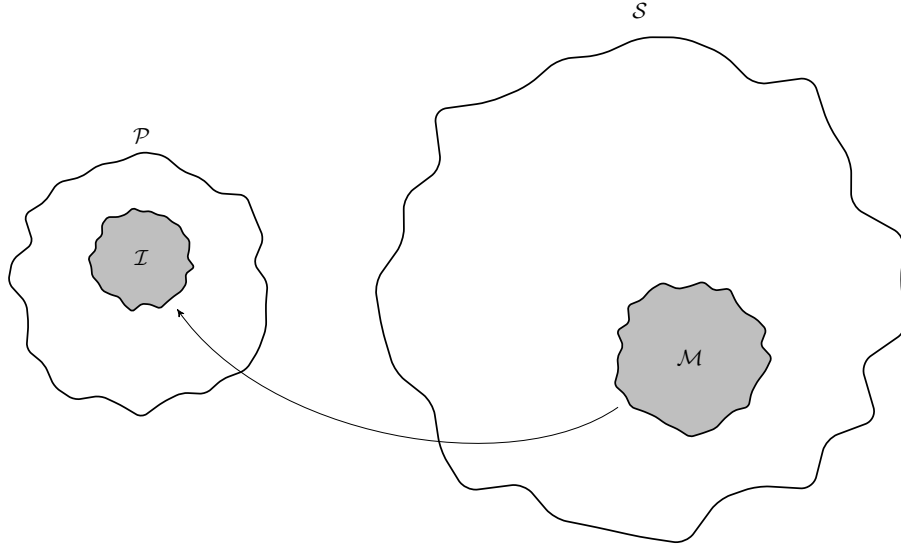
(a) A model $\mathcal{M}$ is a set of structures that forms a proper subset of the class of all
structures $\mathcal{S}$. Each structure in $\mathcal{M}$ generates a probability distribution in the class
of all probability distributions (of observable variables) $\mathcal{P}$. Then the image $\mathcal{I}$ is the
set of all probability distributions that are generated by structures in $\mathcal{M}$.
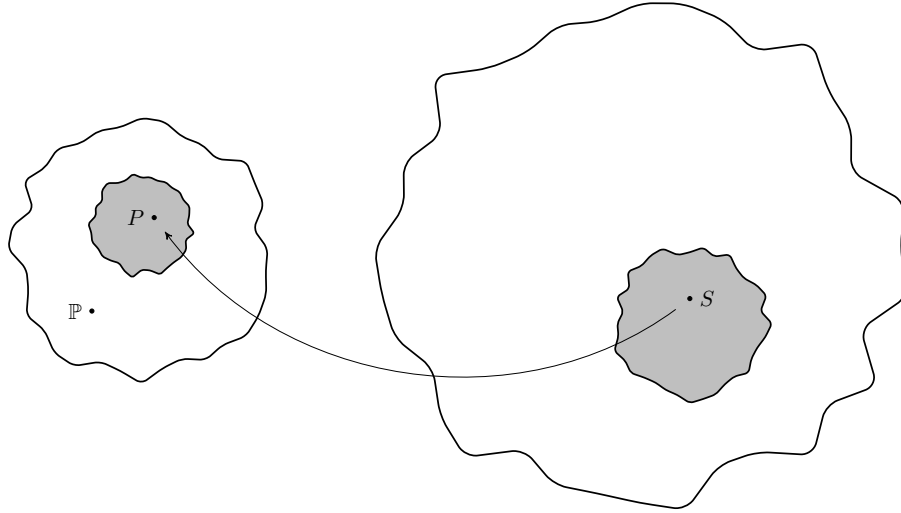
(b) A structure S is incompatible with data if it generates a probability distribution
(of observable variables) P that is distinct from a realised probability distribution $\mathbb{P}$.
If all structures in $\mathcal{M}$ are incompatible with data then $\mathcal{M}$ is said to be observationally
restrictive, and is falsified. This condition is equivalent to $\mathbb{P} \in \mathcal{P} \smallsetminus \mathcal{I}$.

Figure 1: Structures, models, probability distributions (of observable variables), and falsifiability.
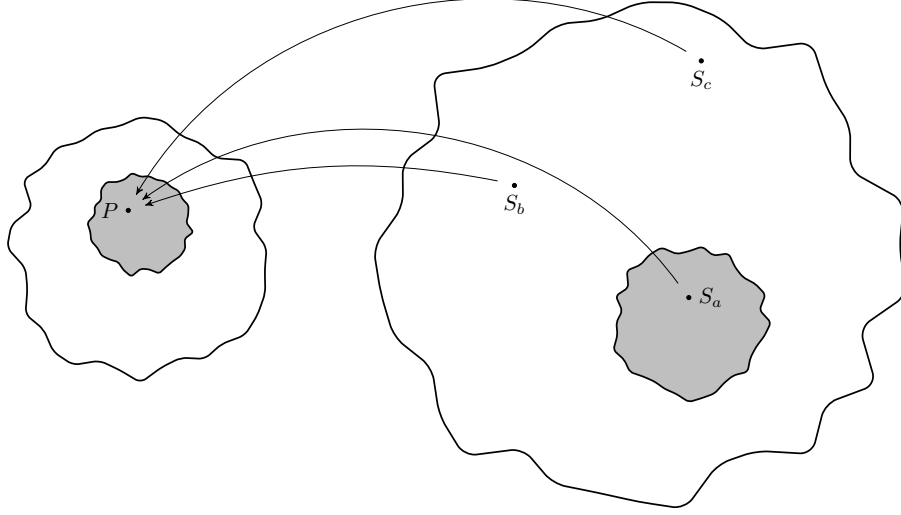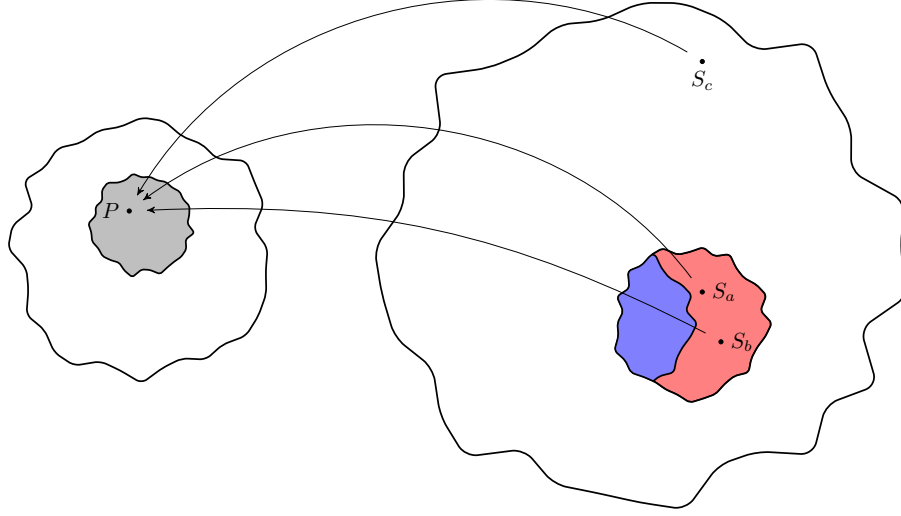
(a) A model $\mathcal{M}$ is said to identify a structure S if the probability distribution (of observable variables) P that is generated by S is distinct from those generated by other structures in $\mathcal{M}$. The structures S_a, S_b and S_c are said to be observationally equivalent as they all generate P but S_b and S_c are not admitted by $\mathcal{M}$. As S_a is the only structure that is admitted by $\mathcal{M}$ and that generates P, S_a is identified by $\mathcal{M}$. For completeness, $\mathcal{M}$ is said to be uniformly identifying if it identifies each structure that it admits.
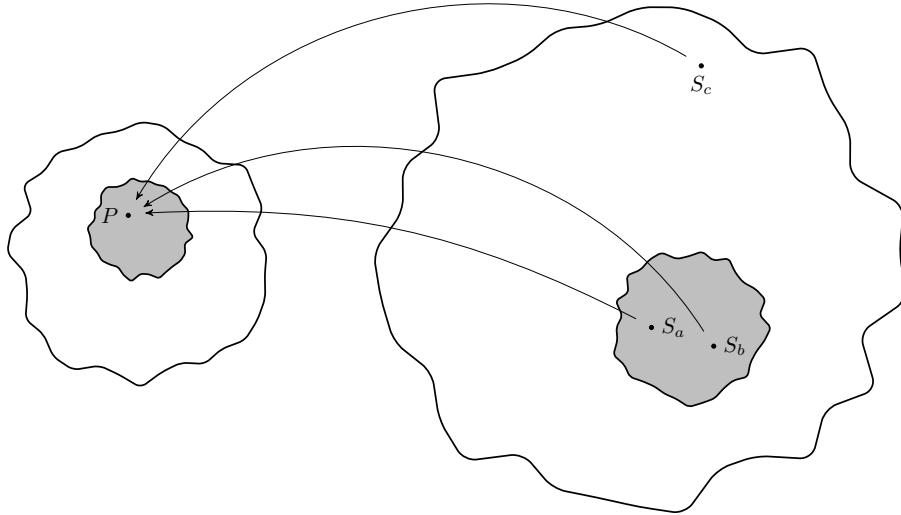

(b) As S_a and S_b are observationally equivalent and are both admitted by $\mathcal{M}$ then $\mathcal{M}$ does not identify either S_a or S_b. Nonetheless, as $\mathcal{M}$ restricts the set of observationally equivalent structures that generate P to S_a and S_b then $\mathcal{M}$ partially identifies S_a (and S_b to within $\{S_a, S_b\}$).

Figure 2: Identification and non-identification of a structure, and partial identification of a structure.

(a) A structural characteristic χ is a function of a structure S. A model $\mathcal{M}$ can be partitioned such that structures in a partition deliver the same value for χ. Structures in the red partition $\mathcal{M}$ deliver the value a for χ, and structures in the red partition $\mathcal{M}$ deliver the value b for χ. If χ is constant across all observationally equivalent structures that $\mathcal{M}$ admits then $\mathcal{M}$ is said to identify χ. As $\chi(S_a)$ is equal to $\chi(S_b)$ (is equal to a) $\mathcal{M}$ identifies χ.
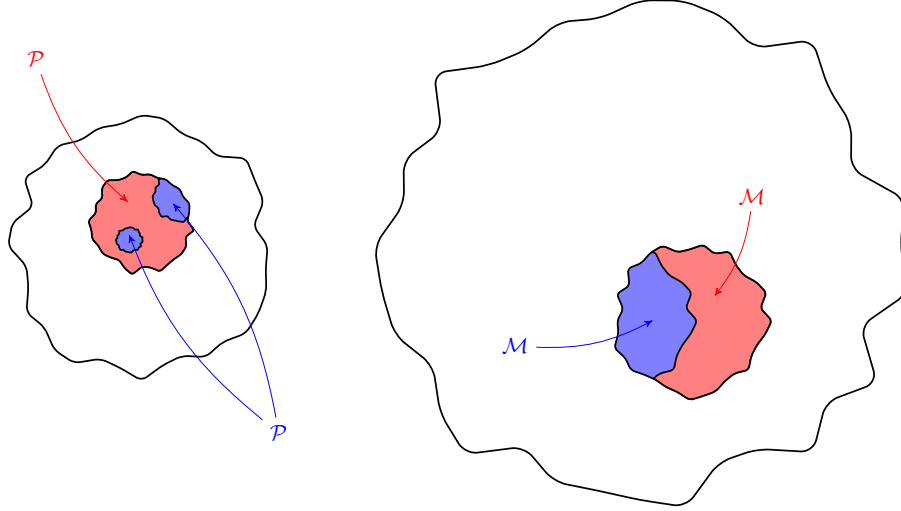
(b) If $\mathcal{M}$ identifies χ for all structures in $\mathcal{M}$ then $\mathcal{M}$ is said to uniformly identify χ. The class of all probability distributions (of observable variables) is partitioned into the blue partition $\mathcal{P}$ and into the red partition $\mathcal{P}$. Probability distributions in $\mathcal{P}$ are generated by (potentially many) structures in $\mathcal{M}$, and probability distributions in $\mathcal{P}$ are generated by (potentially many) structures in $\mathcal{M}$. It is important that the number of partitions in $\mathcal{M}$ and in $\mathcal{P}$ are equal, although that number can be countably infinite. In the context of Figure 3b $\mathcal{M}$ uniformly identifies χ since observationally equivalent structures that $\mathcal{M}$ admits are in the same colour of $\mathcal{M}$. More conveniently, whether $\mathcal{M}$ uniformly identifies χ can be determined by the existence of an identifying correspondence G, a functional. P is a probability distribution in $\mathcal{P}$, and P is a probability distribution in $\mathcal{P}$. Then $\mathcal{M}$ uniformly identifies χ if the value of $G(P)$ is a and if the value of $G(P)$ is b, holding for any such P and P. Notice that if $\mathcal{M}$ uniformly identifies all χ then $\mathcal{M}$ also uniformly identifies structures.

Figure 3: The identification of structural characteristics, and identifying correspondences.

(a) A structural characteristic χ is a function of a structure S. A model $\mathcal{M}$ can be partitioned such that structures in a partition deliver the same value for χ. Structures in the red partition $\mathcal{M}$ deliver the value a for χ, structures in the blue partition $\mathcal{M}$ deliver the value b for χ, and structures in the yellow partition $\mathcal{M}$ deliver the value c for χ. The class of all probability distributions (of observable variables) $\mathcal{P}$ is partitioned into the red partition $\mathcal{P}$, into the blue partition $\mathcal{P}$, into the yellow partition $\mathcal{P}$ and into the grey partition $\mathcal{P}$. Probability distributions in a colour of $\mathcal{P}$ are generated by (potentially many) structures in the same colour of $\mathcal{M}$; the exception is probability distributions in $\mathcal{P}$ which are generated by (potentially many) structures in $\mathcal{M}$ and in $\mathcal{M}$. P is a probability distribution in $\mathcal{P}$ with probability distributions defined similarly for each colour in $\mathcal{P}$.
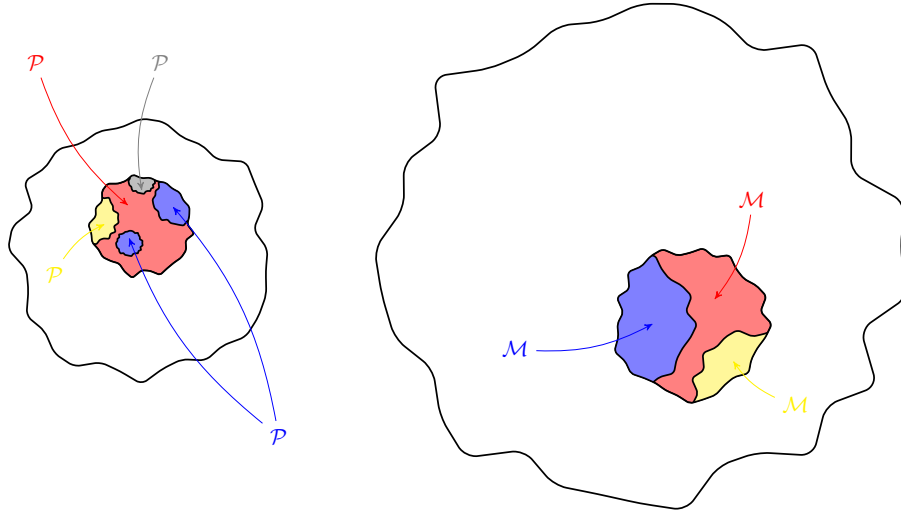
(b) That probability distributions in $\mathcal{P}$ are generated by structures in $\mathcal{M}$ and in $\mathcal{M}$ creates a complication; the value of χ is not constant across observationally equivalent structures that $\mathcal{M}$ admits and that generate a probability distribution in $\mathcal{P}$. So $\mathcal{M}$ does not uniformly identify χ. Consideration of the identifying correspondence G determines that this corresponds to there being structures in $\mathcal{M}$ for which G does not deliver the value of χ when applied to the probability distributions that these structures generate. Nonetheless, if $\mathcal{M}$ restricts the set of values of χ for any probability distribution in $\mathcal{P}$ then $\mathcal{M}$ does have some non-trivial identifying power for χ. Then $\mathcal{M}$ is said to uniformly partially identify χ if $\mathcal{M}$ and $\mathcal{P}$ can each be partitioned into countably many disjoint subsets and that a probability distribution in a partition of $\mathcal{P}$ is not generated by a structure in at least one partition of $\mathcal{M}$, holding for any such partition of $\mathcal{P}$. In the context of Figure 4 $\mathcal{M}$ identifies χ up to $\{a, c\}$, $\mathcal{M}$ identifies χ uniquely to b, and $\mathcal{M}$ identifies χ up to $\{a, c\}$. Each partition of $\mathcal{P}$ includes probability distributions that are generated by structures in at least one partition of $\mathcal{M}$. Equivalently, if G is permitted to be a multivalued functional (or one-to-many) then $\mathcal{M}$ uniformly partially identifies χ if G exists and if $G(P)$ contains the set of values of χ that are delivered by structures that generate P, holding for all such P. A caveat must be applied here; G cannot be trivial in the sense that it is constant across all such P. Clearly this definition of G does not exclude the possibility that there is multiplicity of identifying correspondences that satisfy this property. Sharpness is a desirable property in such circumstances; a functional G that can be shown to deliver smaller sets according to some well-defined distance measure across all possible P (and that satisfies the properties above) should be preferred to any alternative identifying correspondence.

Figure 4: Partial identification of a structural characteristic.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013). Intersection bounds: Estimation and inference. *Econometrica 81* (2), 667 − 737.

Chesher, A. (2005). Nonparametric identification under discrete variation. *Econometrica 73* (5), 1525 − 1550.

Chesher, A. (2010). Instrumental variable models for discrete outcomes. *Econometrica 78* (2), 575 − 601.

Chesher, A. and A. M. Rosen (2013). What do instrumental variable models deliver with discrete dependent variables?. *American Economic Review 103* (3), 557 − 562.

Chesher, A., A. M. Rosen, and K. Smolinski (2013). An instrumental variable model of multiple discrete choice. *Quantitative Economics 4* (2), 157 − 196.

Heckman, J. J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation1. *Econometrica 73* (3), 669 − 738.

Hurwicz, L. (1950). Generalization of the concept of identification. *Statistical Inference in Dynamic Economic Models (T. Koopmans, ed.). Cowles Commission, Monograph 10*, 245 − 257.

Khan, S. and E. Tamer (2010). Irregular identification, support conditions, and inverse weight estimation. *Econometrica 78* (6), 2021 − 2042.

Kitagawa, T. (2009). Identification region of the potential outcome distributions under instrument independence. *Econometrica (Revise and resubmit)*.

Koopmans, T. C. and O. Reiersøl (1950). The identification of structural characteristics. *Annals of Mathematical Statistics 21* (2), 165 − 181.

Manski, C. F. (2013). *Public Policy in an Uncertain World: Analysis and Decisions.* Cambridge and London: Harvard University Press.

Molchanov, I. (2005). *Theory of Random Sets.* Springer.

Shaikh, A. M. and E. J. Vytlacil (2011). Partial identification in triangular systems of equations with binary dependent variables. *Econometrica 79* (3), 949 − 955.

Strotz, R. H. and H. O. Wold (1960). Recursive vs. nonrecursive systems: An attempt at synthesis (part i of a triptych on causal chain systems). *Econometrica: Journal of the Econometric Society*, 417 − 427.