Department of Economics, University College London

Information and admissible sets

Jeff Rowley

16th July, 2014

Abstract

«Abstract here»

Acknowledgements. *«Acknowledgements here» I acknowledge financial support from the Economic and Social Research Council (ESRC).*

I explore the effect of incorporating information for a non-parametric binary choice model. The model permits endogenous variation in a scalar random variable 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 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

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

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 aforementioned 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.

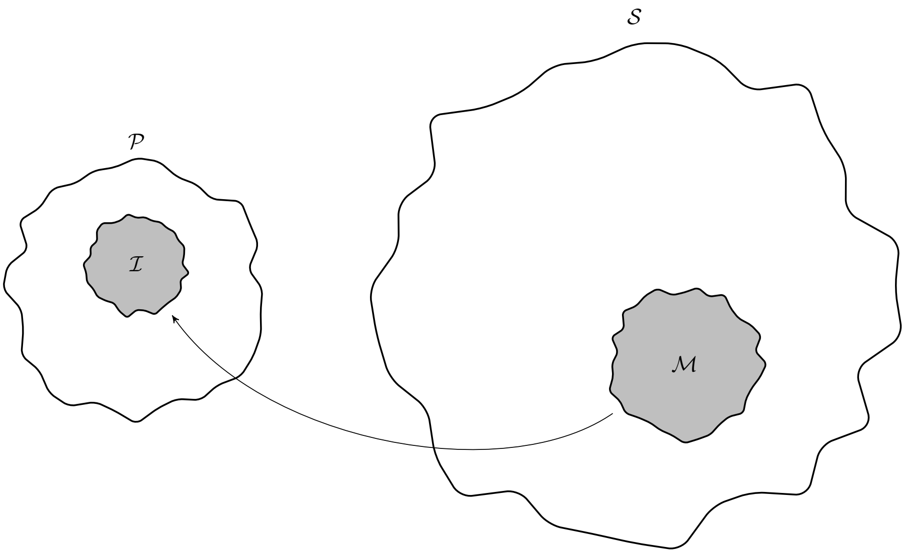
A relevant question that arises from the existence of conditional causal effects is how to relate these conditional effects to unconditional effects. More precisely, how does the average causal effect of the endogenous variable on the outcome variable relate to conditional average causal effects? I show that the average causal effect of the endogenous variable on the outcome variable can be expressed as a function of its conditional counterparts when the non-parametric binary choice model is enriched in a flexible way. I extend Chesher and Rosen (2013) in this respect.

Notation

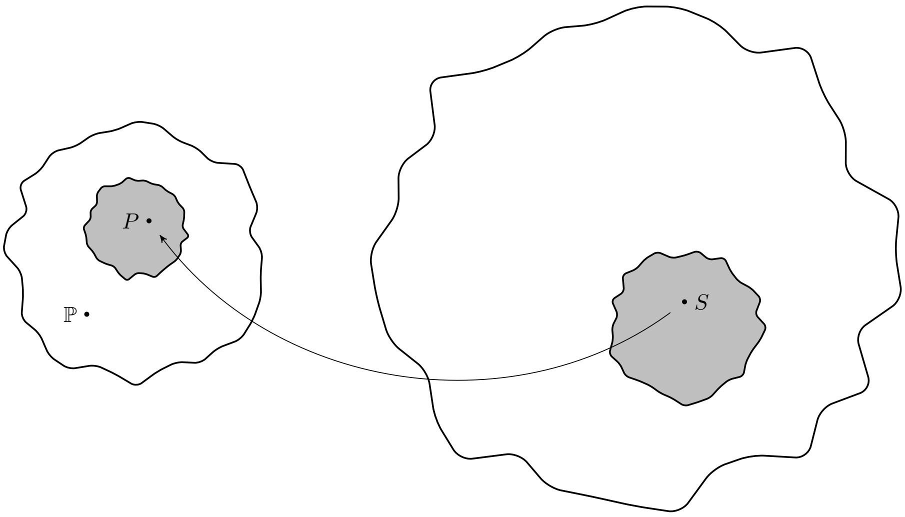
There is a probability space $(\Omega, \Sigma, \mathbb{P})$ on which are defined random variables (Y, D, X, Z, U). Here, (Y, D, X, Z) are observable with supports $(\mathcal{R}_Y, \mathcal{R}_D, \mathcal{R}_X, \mathcal{R}_Z)$, and U is unobservable with as yet unspecified support. I allow (X, Z, U) to be vectors, in which case the support is given by the Cartesian product of the supports of each element in the vector. I refer to Y as the outcome variable, to D as the endogenous variable, to X as the exogenous variable, to Z as the instrumental variable, and to U as unobservable heterogeneity. The logic of this naming convention will be made clear by the restrictions that are imposed upon these random variables in the main text. Lower case letters are used to represent specific values of these random variables.

I denote by $Y(d)$ the counterfactual value of Y when D is externally fixed, and by $D(z)$ the counterfactual value of D when Z is externally fixed. I denote by $\mathbb{E}$ the expectation operator, and by $\mathbb{1}$ the indicator function. Related to these concepts are the average causal effects $ACE(D \to Y)$ and $ACE(Z \to D)$ that are defined as $\mathbb{E}[Y(d_1) - Y(d_0)]$ and $\mathbb{E}[D(z_1) - D(z_0)]$ that are well-defined when D and Z are binary, respectively. To distinguish between population and sample quantities, I subscript sample quantities by n.

Further terminology and notation is introduced 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). Following Hurwicz (1950) I also adopt the notation $S :. P$ that signifies that a structure S generates a probability distribution (of observable variables) P, and $P .: G$ that signifies that P is generated by S.

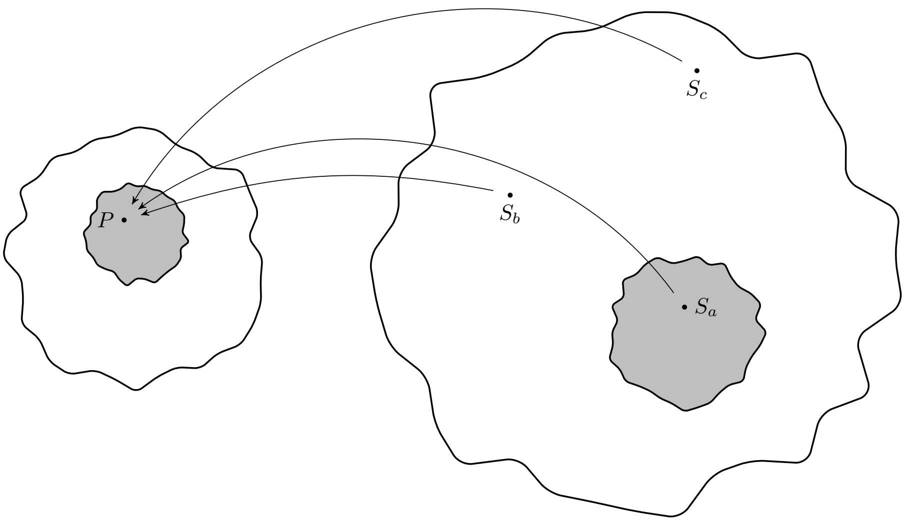


(a) A model $\mathcal{M}$ is a set of structures that forms a proper subset of the class of all structures $\mathcal{S}$. Each structure in $\mathcal{M}$ generates a probability distribution in the class of all probability distributions (of observable variables) $\mathcal{P}$. Then the image $\mathcal{I}$ is the set of all probability distributions that are generated by structures in $\mathcal{M}$.

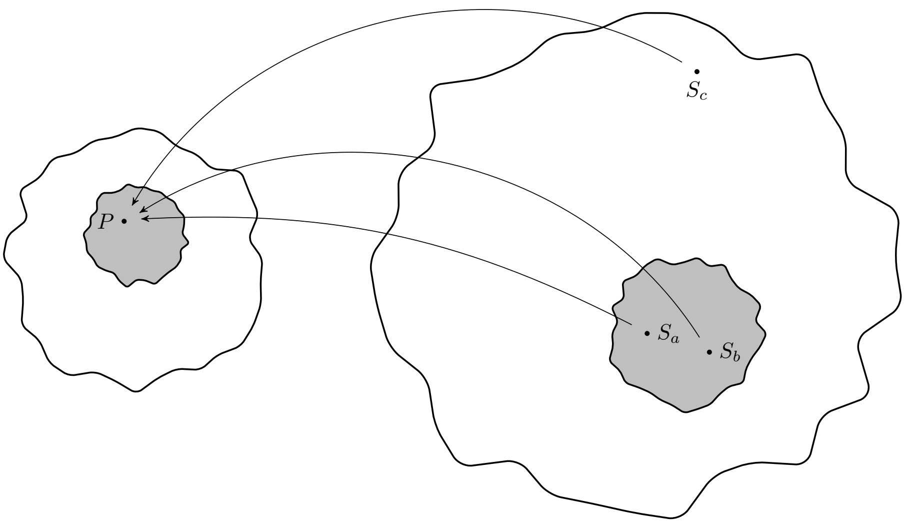


(b) A structure S is incompatible with data if it generates a probability distribution (of observable variables) P that is distinct from a realised probability distribution $\mathbb{P}$. If all structures in $\mathcal{M}$ are incompatible with data then $\mathcal{M}$ is said to be observationally restrictive, and is falsified. This condition is equivalent to $\mathbb{P} \in \mathcal{P} \setminus \mathcal{I}$.

Figure 1: Structures, models, probability distributions (of observable variables), and falsifiability.

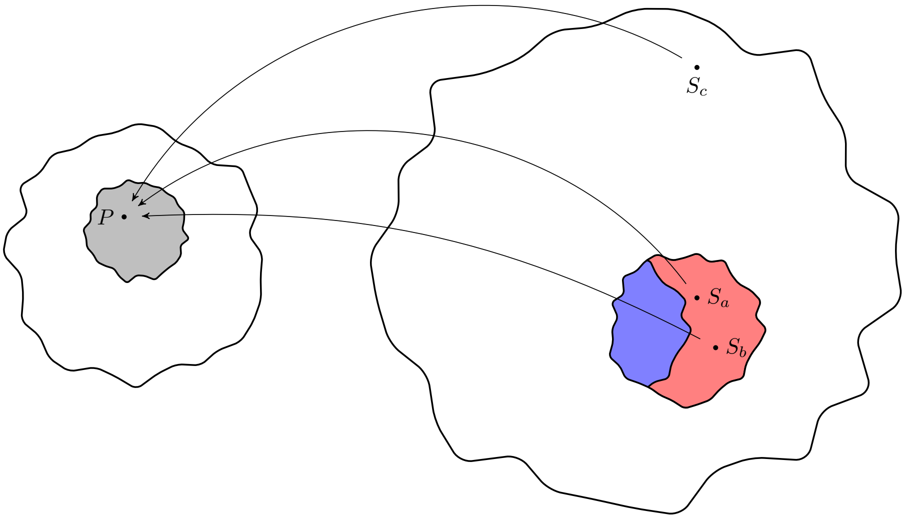


(a) A model $\mathcal{M}$ is said to identify a structure S if the probability distribution (of observable variables) P that is generated by S is distinct from those generated by other structures in $\mathcal{M}$. The structures S_a, S_b and S_c are said to be observationally equivalent as they all generate 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.

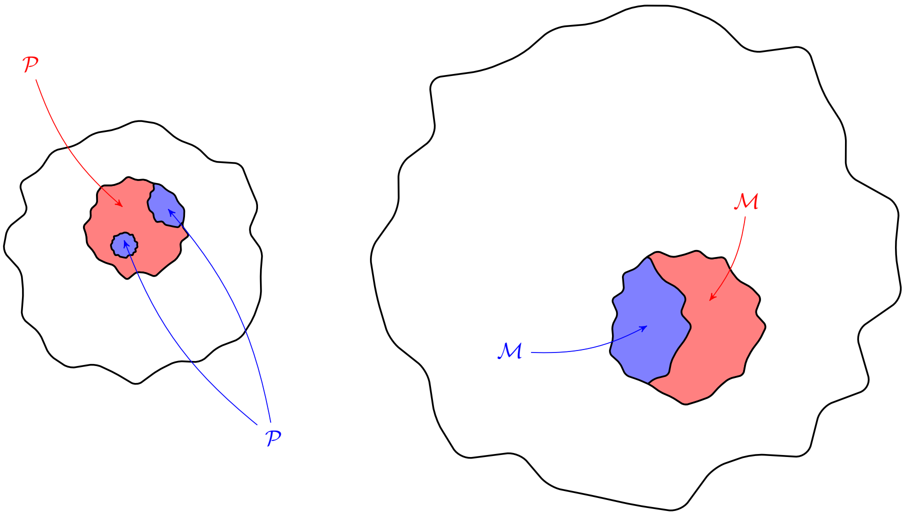


(b) As S_a and S_b are observationally equivalent and are both admitted by $\mathcal{M}$ then $\mathcal{M}$ does not identify either S_a or S_b. Nonetheless, as $\mathcal{M}$ restricts the set of observationally equivalent structures that generate P to S_a and S_b then $\mathcal{M}$ partially identifies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



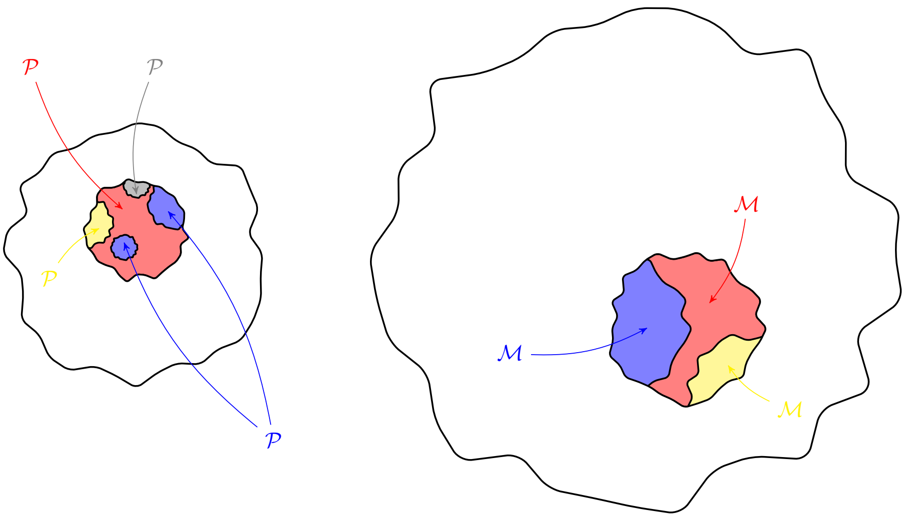
(a) A structural characteristic χ is a function of a structure S. A model $\mathcal{M}$ can be partitioned such that structures in a partition deliver the same value for χ. Structures in the red partition $\mathcal{M}$ deliver the value a for χ, and structures in the red partition $\mathcal{M}$ deliver 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.

1 A threshold crossing model

References

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

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

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.

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.

[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 as $CATE(x)$.