Department of Economics

Single Equation Instrumental Variable Models

- Identification under Discrete Variation -

Konrad Smolinski

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Florence, January 2012

EUROPEAN UNIVERSITY INSTITUTE
Department of Economics

Single Equation Instrumental Variable Models

- Identification under Discrete Variation -

Konrad Smolinski

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Examining Board:

Professor Richard Spady, Johns Hopkins University (External Supervisor)
Professor Helmut Lütkepohl, European University Institute
Professor Stéphane Bonhomme, CEMFI
Professor Richard Smith, University of Cambridge

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Collaboration

I acknowledge that substatntial parts of this thesis result from an extensive and close colaboration with prof. Andrew Chesher all through my visit at UCL & IFS, London in 2009-2011. The influence of the intelectual circle around the Centre for Microdata Methods and Practices is also present throughout these pages.

The second and the third chapters of the thesis follow from joint projects with prof. Andrew Chesher and were published jointly in the CeMMAP Working Papers Series (CWP).

The second chapter was published under the title *IV Models of Ordered Choice* and turned up as CWP37/09. Revised and slightly extended version of this paper was accepted for publication in the Journal of Econometrics.

The third chapter appeared as CWP11/10 under the title *Sharp identified sets for discrete variable IV models.*

The last chapter is an independent work developed in a parallel to ideas presented in the CeMMAP Working Paper CWP06/11, *An instrumental variable models of multiple discrete choice*, by Andrew Chesher, Adam Rosen and myself.

Acknowledgments

A doctoral thesis is much more than doing research and producing pages. It is a unique journey through which the author emerges, bravely having navigated turbulent winds of personal limitations throughout. If anything this is what I learned during my time at the EUI in Florence and what I experienced in the following years at CeMMAP in London.

Along the way I have met remarkable people who added immensely to my understanding and to my life. Without their support, encouragement and challenge but also their simple presence and hearted openness, many ideas and insights in these pages would not have been possible. To all of them I express my deepest and heartfelt gratitude. However, it is possible to give particular mention to only some of them here. I owe an additional debt of gratitude to:

Richard H. Spady, my supervisor at the EUI in Florence who showed me widely opened doors of conteporary economic thoughts and challenged me in econometrics with courage and patience.

Andrew D. Chesher, my tutor at CeMMAP in London who presented me to the world of modern econometrics whom I wish to particularly acknowledge for encouragement, persistence in scientific inquiry and collaboration with this work and co-authorship of pages to follow.

My fellows from the famous Santa Rosa Community in Florence, Frank Betz, Stephane Tock and in particular my dearest friend Matej Avbelj, with whom I had the privilege to be associated and whose growth and development served to illuminate my own.

Georg Duernecker and Andrzej Siodmok for their scientific ingenuity and intellectual challenge, lasting support and courage to follow through and reach the end of this project.

Olga Milajewa, Malgorzata Renes and Aleksandra Wlodarczyk for inner strength, personal support, dedication and deep caring in the final year of work on the doctorate.

Mayssun El-Attair, Dominika Renes, Magdalena Silska, Rafal Silski and Agnieszka Sondej for their faith and hospitality. I thank you all for your friendship.

Colleagues at the EUI, CeMMAP and UCL for wonderful times and inspiration, countless discussions and endless debates.

Lastly, but by no means least, my deepest gratitude goes to all my dearest colleagues and friends who empowered my imagination and wonder or enriched me otherwise on the path towards the Doctorate. They had to bear with me at a time of possession by abstract concepts with very little patience for anything but a simplicity and sharpness of the ideas to follow. Above all, your encouragement, support and openness has made this venture possible.

I dedicate this thesis to my grandfather Wlodzimierz Ignatiuk.

Warsaw, January 13, 2012

Contents

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Chapter 1

Instead of Introduction

1.1 Preface

Over the last decade, substantial interest in theoretical econometrics and microeconometrics has been directed towards nonparametric models. Much work has been devoted to the development of novel identification and estimation technieques and in particular, to the identifying power of econometric models under various types of restrictions. Notable attention has been focused on the conditional independence restriction and instrumental variable methods for both continuous and discrete data problems. This immense effort has led to tremendous outcomes in terms of theoretical findings and most importantly, new empirical practices.

Nowadays, we face an apparent emphasis on minimal restrictions of nuisance parameters of the model, with a focus on specific structural features at the same time. New models permit the relaxation of implausible restrictions frequently superimposed unwillingly in empirical analysis of plain old econometric models.

In this spirit, recent developments in microeconometrics have given rise to increasing interest in partially identified models. In these models, for the credibility of claims, the feature of interest is bounded to a set rather then constituting of a point in the space of parameters or functions. This in turn has its own place in economic practice.

Among many appealing and commonly investigated economic circumstances, partial identification frequently arises in econometric inquiry when researchers are faced with discrete data, omnipresent in survey studies. Examples consider a very general class of the limited information discrete outcome models with endogeneity when very little is known about the

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

genesis of the process generating endogenous variable.

This thesis contributes to the aforementioned line of research and seeks to address a somewhat limited, but I believe important, range of issues in a great depth. These issues are concerned with the specification of identified sets in so-called single equation models with endogeneity. We achieve identification via instrumental variable restrictions and focus on discrete outcomes as well as discrete endogenous variables.

Our focus on discrete, ordered outcome models complements the vast majority of research on econometric design under continuous variation. The latter, even though theoretically sound, often becomes practically infeasible. We believe that this study provides a level of unity to the partial identification framework as a whole and makes steps forward in understanding some aspects of single equation instrumental variable models under discrete variation.

1.2 Content of the Thesis

The thesis is organized as follows. Chapter I studies single equation instrumental variable models of ordered choice in which explanatory variables may be endogenous. This chapter provides results on the properties of the identified set for the case in which potentially endogenous explanatory variables are discrete, with a sharpness result when the endogenous variable is binary. The results are used as the basis for the calculations showing the rate of shrinkage of the identified sets as the number of classes, in which the outcome is categorized, increases.

Chapter II discusses general characterization of the identified sets of structural functions when endogenous variables are discrete. Identified sets are unions of large numbers of convex sets and may not be convex or even be connected. Each of the component sets is a projection of a convex set, that resides in a much higher dimensional space, onto the space in which a structural function resides. We develop a symbolic expression for this projection and give a constructive demonstration that it is indeed the identified set. Also, we develop an expression for a set of structural functions for the case in which endogenous variables are continuous or mixed discrete-continuous and show that this set contains all structural functions in the identified set in the non-discrete case.

In Chapter III we introduce core determining partitions, indexes and sets for the models with discrete observables and continuous latent heterogeneity. Core determining indexes and sets give rise to the finite number of the core determining inequalities, i.e. the collection of the ultimate identification questions. We introduce an algorithm to deliver the core determining indexes and illustrate the method for the ordered outcome instrumental variable model studied in previous chapters.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Chapter 2

IV Models of Ordered Choice

This paper studies single equation instrumental variable models of ordered choice in which explanatory variables may be endogenous. The models are weakly restrictive, leaving unspecified the mechanism that generates endogenous variables. These incomplete models are set, not point, identifying for parametrically (e.g. ordered probit) or nonparametrically specified structural functions. The paper gives results on the properties of the identified set for the case in which potentially endogenous explanatory variables are discrete. The results are used as the basis for calculations showing the rate of shrinkage of identified sets as the number of classes in which the outcome is categorised increases.

2.1 Introduction

This paper studies single equation instrumental variables models for ordered outcomes in which explanatory variables may be endogenous. These models arise in structural econometric analysis of individuals' choices amongst ordered alternatives, or of individuals' attitudes arranged on an ordinal scale and they arise in many other settings in which data are interval censored continuous outcomes.

A common ploy when dealing with endogenous variation in a discrete response situation is to presume that the discrete response is generated in a recursive, triangular system along with the endogenous variable. Then, calling on some further restrictions, a control function method is used as the basis for identification and estimation. See for example Smith and Blundell

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

(1986), Rivers and Vuong (1988), Blundell and Powell (2003, 2004), Chesher (2003).[1]

Unfortunately this strategy does not generally work when endogenous variables are discrete.[2] And, as explained in Chesher(2009), the control function approach exploits strong restrictions concerning the process generating the endogenous variables, restrictions which may not be found plausible in many econometric settings. By contrast here we work with a model which is far less restrictive in this regard, imposing conditions *only* on the structural function generating the discrete response.

The model requires that a scalar ordered outcome Y, with $M \geq 2$ points of support, is determined by a structural function $h(X, U)$ which is weakly monotone in scalar unobserved U. The observed vector of explanatory variables, X, and U may not be independently distributed. However the model requires that U be distributed independently of instruments, Z. We call the model a *Single Equation Instrumental Variable* (SEIV) model. The SEIV model places no restrictions at all on the process generating the endogenous variable, X, and in this respect is incomplete.

Thinking about Manski's (2003) "Law of Decreasing Credibility" encourages us to take this approach. It allows one to see what is lost by relaxing the strong restrictions of the triangular control function model. It turns out that what is lost is point identification because the SEIV model is generally *set* not point identifying. Dropping the restrictions of the control function model leads to ambiguity.

This paper focusses on models with discrete endogenous variables, having K points of support, $\{x_1, \ldots, x_K\}$, and explores the identified sets the SEIV model delivers. The main results are now summarised.

Since the structural functions of a SEIV model are monotone in scalar U there is a threshold crossing representation in which U is normalised marginally uniformly distributed

[1] The control function approach is used quite widely in applied econometric practice. STATA, Statacorp(2007) and LIMDEP, Greene (2007), are examples of widely used proprietary software suites armed with commands to conduct control function estimation of models of binary responses.

[2] Chesher (2005) gives partial identification results for a control function model with discrete endogenous variables.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

on the unit interval.

$$
h(X,U) \equiv \begin{cases} 1 & , \quad 0 \leq \ U \ \leq h_1(X) \\ 2 & , \quad h_1(X) < \ U \ \leq h_2(X) \\ \vdots & \vdots \quad \quad \vdots \ \ \vdots \ \ \vdots \\ M & , \quad h_{M-1}(X) < \ U \ \leq 1 \end{cases}
$$

In the discrete endogenous variable case a nonparametrically specified structural function, h, is characterised by $N = K \times (M-1)$ parameters, denoted γ, which are the values of the $M-1$ threshold functions at the K values of X.

Let $\mathcal{H}^0(\mathcal{Z})$ denote the set of values of γ identified by the SEIV model given $F^0_{YX|Z}$, a probability distribution for Y and X conditional on Z, when Z takes values in a set $\mathcal{Z}$. Each structural function is characterised by a point in the unit N-cube and $\mathcal{H}^0(\mathcal{Z})$ is a subset of that space.

The identified set delivered by a nonparametric SEIV model is shown to be a union of convex sets each defined by a system of linear equalities and inequalities. The number of sets involved can be enormous in what at first sight seem to be small scale problems. For example when $M = K = 5$ there may be over 300 billion component sets. The result is generally not a convex set unless instruments are strong. We give examples in which the identified set is not convex and, indeed, not connected. Shape restrictions (e.g. monotonicity) or parametric restrictions can bring substantial simplification.

A system of inequalities given in Chesher (2008) defines an outer set, $\mathcal{C}^0(\mathcal{Z})$, within which the SEIV model's identified set lies. We develop expressions for these inequalities for the M outcome, discrete endogenous variable case. We propose a second system of inequalities defining a set of values of γ, $\mathcal{D}^0(\mathcal{Z})$, and show that the identified set resides in the intersection $\tilde{\mathcal{C}}^0(\mathcal{Z}) \equiv \mathcal{C}^0(\mathcal{Z}) \cap \mathcal{D}^0(\mathcal{Z})$.

When the outcome Y is binary $\mathcal{C}^0(\mathcal{Z})$ is a subset of $\mathcal{D}^0(\mathcal{Z})$ and, as shown in Chesher (2008), in that case $\mathcal{C}^0(\mathcal{Z})$ is the identified set $\mathcal{H}^0(\mathcal{Z})$. Here we show that when the *endogenous variable is binary* $\tilde{\mathcal{C}}^0(\mathcal{Z})$ is the identified set however many categories there are for Y.

Finally we examine the impact of *response discreteness* on the identified sets. The discrete response model studied here is a non-additive error model and the results for such models for continuous outcomes given in Chernozhukov and Hansen (2005) show that there can be

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

point identification in SEIV models when observed responses are continuous. So it is to be expected that as the number of categories observed rises there is reduction in ambiguity and an approach to point identification.

We investigate this in the context of a model with parametrically specified structural functions such as arise in ordered probit models. We find that in the cases considered identified sets for a parameter such as a coefficient in a linear index shrink at a rate approximately equal to the inverse of the square of the number of classes in which the outcome is categorised. In the example, when Y is categorised into 10 or more classes, the SEIV model delivers identified sets which are very small indeed.

The paper is organised as follows. Section 2.2 give a formal definition of the SEIV model and defines its identified set of structural functions.

Section 2.3 develops the main results for nonparametrically specified structural functions with discrete endogenous variables. In Section 2.3.1 a piecewise uniform system of conditional distributions of U given X and Z is introduced and conditions under which a structural function lies in the identified set are stated. The geometry of the identified set for nonparametrically specified structural functions is discussed in Section 2.3.2 and systems of inequalities obeyed by values of these functions that lie in the identified set are set out in Section 2.3.3 Proofs of propositions are given in an Annex.

Section 2.4 illustrates the results using a parametrically specified model which, in the absence of endogeneity, would be a conventional ordered probit model. This Section gives results on the rate of shrinkage of identified sets as the number of categories of the discrete outcome increases. Section 2.5 concludes.

2.2 An IV model for ordered outcomes

In the SEIV model a scalar ordered outcome Y is determined by observable X, which may be a vector, and unobserved scalar U. Restriction 1 defines admissible structural functions.

Restriction 1. *Y is determined by a structural function as follows:*

$$Y = h(X, U) \equiv \begin{cases} 1 & , \quad h_0(X) \leq \ U \ \leq h_1(X) \\ 2 & , \quad h_1(X) < \ U \ \leq h_2(X) \\ \vdots & \vdots \qquad\qquad \vdots \quad \vdots \quad \vdots \\ M & , \quad h_{M-1}(X) < \ U \ \leq h_M(X) \end{cases}$$

with, for all x, $h_0(x) = 0$ and $h_M(x) = 1$ and for all x and m, $h_m(x) > h_{m-1}(x)$. U is normalised to have a marginal uniform distribution on $[0, 1]$.

Specifying the values of Y to be the first M integers is an innocuous normalisation because Y is an ordered outcome.

U and X are not required to be independently distributed so the model allows elements of X to be endogenous. However U is required to be distributed independently of instrumental variables, Z, as set out in Restriction 2.

Restriction 2. *U and instrumental variables Z which take values in some set $\mathcal{Z}$ are independently distributed in the sense that the conditional distribution function of U given $Z = z$ satisfies $F_{U|Z}(u|z) = u$ for all $u \in [0, 1]$ and $z \in \mathcal{Z}$.*

Restriction 1 excludes the instrumental variables from the structural function. Neither restriction imposes any conditions on the process generating X. Now consider the identifying power of this model.

Let $F^0_{YX|Z}$ denote some distribution function of Y and X conditional on Z. Imagine a situation in which data are informative about this distribution for values of Z that lie in a set $\mathcal{Z}$. If this distribution function is compatible with the SEIV model then there exists (i) a structural function h^0 with threshold functions $\{h^0_m\}^M_{m=1}$ and (ii) a distribution function $F^0_{UX|Z}$, both admitted by the SEIV model and such that the following condition holds when $h = h^0$ and $F_{UX|Z} = F^0_{UX|Z}$.

$$F^0_{YX|Z}(m, x|z) = F_{UX|Z}(h_m(x), x|z), \qquad \text{for all: } z \in \mathcal{Z}, m \text{ and } x. \tag{2.1}$$

There may be more than one admissible structure $(h, F_{UX|Z})$ satisfying (2.1) because it may be possible to compensate for variations in the x-sensitivity of the threshold func-

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

tions $\{h_m\}_{m=1}^M$ by adjusting the u- and x-sensitivity of $F_{UX|Z}$ leaving the left hand side of (2.1) unchanged while respecting the independence Restriction 2. So the model is *partially identifying.*

For a distribution $F^0_{YX|Z}$ let $\mathcal{S}^0(\mathcal{Z})$ denote the set of structures identified by the model comprising Restrictions 1 and 2. This is the set of structures admitted by the SEIV model and satisfying (2.1). The set of *structural functions* identified by the model, denoted $\mathcal{H}^0(\mathcal{Z})$, is the set of structural functions h which are elements of structures lying in the identified set.

$$\mathcal{H}^0(\mathcal{Z}) \equiv \{h : \exists \, \text{admissible } F_{UX|Z} \text{ s.t. } (h, F_{UX|Z}) \in \mathcal{S}^0(\mathcal{Z})\}$$

The set $\mathcal{H}^0(\mathcal{Z})$ is a *projection* of the set $\mathcal{S}^0(\mathcal{Z})$.

This set is difficult to characterise and compute when X is continuously distributed because determining whether there exists a distribution function $F_{UX|Z}$ that can accommodate a particular structural function may require searching across an infinite dimensional space of functions.

However Chesher (2008) shows that when Y is *binary* the identified set is determined by a system of inequalities in which the distribution function $F_{UX|Z}$ does not appear. One of the contributions of this paper is a similar result for the case in which a scalar *endogenous explanatory variable* X is binary and Y takes any number of values.

When X is discrete, say with K points of support, the distribution function $F_{UX|Z}$ can be characterised by a finite number of parameters for each value of Z and the identified set can be computed when M and K are not too large. The remainder of the paper studies the case in which the explanatory variable, X, is discrete.

2.3 Identified sets with discrete endogenous variables

2.3.1 Identification

When X is discrete and K-valued with $X \in \{x_i\}_{i=1}^K$, the threshold functions are characterised by the parameters

$$\gamma_{mi} \equiv h_m(x_i), \qquad m \in \{0, \ldots, M\}, \quad i \in \{1, \ldots, K\}$$

 DOI: 10.2870/35896

of which $N \equiv (M-1)K$ are free, that is not restricted to be zero or one. Define $\gamma_i \equiv \{\gamma_{mi}\}_{m=0}^{M}$ and $\gamma \equiv \{\gamma_i\}_{i=1}^{K}$ with , for all $i \in \{1, \ldots, K\}$, $\gamma_{0i} \equiv 0$, $\gamma_{Mi} \equiv 1$.

In the discrete X case an identified set of structural functions is a set of values of γ, comprising a subset of the unit N-cube.

When determining whether a structural function characterised by a value of γ lies in the identified set it is sufficient to search across distribution functions which, at each value z of the instrumental variables are characterised by the following parameters.

$$\beta_{mij}(z) \equiv F_{U|XZ}\left(\gamma_{mi}|x_j, z\right), \qquad m \in \{0, 1, \ldots, M\}, \quad (i, j) \in \{1, \ldots, K\}$$

Let $\beta(z)$ denote the list of values $\beta_{mij}(z)$, $m \in \{1, \ldots, M\}$, $(i, j) \in \{1, \ldots, K\}$ for some value z. For all $(i, j) \in \{1, \ldots, K\}$ define $\beta_{0ij}(z) \equiv 0$ and $\beta_{Mij}(z) \equiv 1$. Let $\beta(\mathcal{Z})$ denote the list of values of $\beta(z)$ generated as z varies across $\mathcal{Z}$.

Values $\beta_{mij}(z)$ with $i = j$ are relevant because observational equivalence requires that if γ lies in the identified set then for each $z \in \mathcal{Z}$, m and i the equality

$$F_{U|XZ}\left(\gamma_{mi}|x_i, z\right) = F^0_{Y|XZ}(m|x_i, z) \tag{2.2}$$

must hold. The conditional distribution $F^0_{X|Z}$ is identified so (2.2) is effectively the observational equivalence condition (2.1).

The independence restriction together with the uniform distribution normalisation of the marginal distribution of U requires that for each m, i and z the following condition holds:

$$E^0_{X|Z=z}[F_{U|XZ}\left(\gamma_{mi}|X, z\right)] \equiv \sum_{j=1}^{K} F_{U|XZ}\left(\gamma_{mi}|x_j, z\right) \mathrm{Pr}_0[X = x_j|Z = z] = \gamma_{mi} \tag{2.3}$$

so values of $\beta_{mij}(z)$ with $i \neq j$ are also relevant. Here $E^0_{X|Z=z}$ indicates expectation taken with respect to the distribution $F^0_{X|Z}$ with the conditioning variable Z taking the value z.

So, for each point x_j in the support of X the values of the conditional distribution functions, $F_{U|XZ}(u|x_j, z)$, at *each* value of $u \in \gamma$ are relevant when determining whether γ is in the identified set. Other values of u are not relevant because they play no role in the fulfillment of the observational equivalence condition (2.2) or the independence condition (2.3).

If γ_{mi} and $\gamma_{m'i'}$ are adjacent[3] values of the threshold parameters then the definition of $F_{U|XZ}$ for any values, x_j and z of the conditioning variables can be completed by connecting $F_{U|XZ}(\gamma_{mi}|x_j, z)$ and $F_{U|XZ}(\gamma_{m'i'}|x_j, z)$ with straight line segments delivering histogram-like piecewise uniform conditional distributions.[4]

Let $\Pr_0$ denote probabilities calculated using a particular distribution function $F^0_{YX|Z}$. Define conditional probabilities for X given Z:

$$\delta^0_i(z) \equiv \Pr_0[X = x_i|Z = z] \qquad i \in \{1, \ldots, K\}$$

and define $\delta^0(z) \equiv \{\delta^0_i(z)\}^K_{i=1}$. Let

$$\delta_i(z) \equiv \Pr[X = x_i|Z = z] \qquad i \in \{1, \ldots, K\}$$

be conditional probabilities of X given Z

Define conditional probabilities and cumulative probabilities of the outcome:

$$\alpha^0_{mi}(z) \equiv \Pr_0[Y = m|X = x_i, Z = z], \qquad m \in \{0, \ldots, M\}, \quad i \in \{1, \ldots, K\}$$

$$\bar{\alpha}^0_{mi}(z) \equiv \sum_{n=0}^{m} \alpha^0_{ni}(z), \qquad m \in \{0, \ldots, M\}, \quad i \in \{1, \ldots, K\}$$

with $\alpha^0_{0i}(z) \equiv 0$ for all i and z, and lists of conditional probabilities as follows.

$$\alpha^0_i(z) \equiv \{\alpha^0_{mi}(z)\}^M_{m=0} \qquad \alpha^0(z) \equiv \left\{\alpha^0_i(z)\right\}^K_{i=1}$$

$$\bar{\alpha}^0_i(z) \equiv \{\bar{\alpha}^0_{mi}(z)\}^M_{m=0} \qquad \bar{\alpha}^0(z) \equiv \left\{\bar{\alpha}^0_i(z)\right\}^K_{i=1}$$

Consider a *structure* characterised by

1. γ: a list of values of threshold functions,
2. $\beta(\mathcal{Z})$: a list of values of conditional distribution functions of U given X and Z obtained as Z takes values in $\mathcal{Z}$, and,

[3]If there is no $\gamma_{st} \in \gamma$ such that $\gamma_{mi} < \gamma_{st} < \gamma_{m'i'}$ then γ_{mi} and $\gamma_{m'i'}$ are adjacent.

[4]Using straight line segments ensures that the independence condition:

$$E^0_{X|Z=z}[F_{U|XZ}\,(u|X, z)] = u$$

holds for all $u \in (0, 1)$ and $z \in \mathcal{Z}$.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

3. $\delta(\mathcal{Z})$: a list of values of conditional probabilities of X given $Z = z$, $\delta(z) = \{\delta_i(z)\}_{i=1}^K$ where $\delta_i(z) \equiv \Pr[X = x_i|Z = z]$, obtained as z varies across $\mathcal{Z}$.

Such a structure lies in the set identified by the SEIV model associated with probabilities $\alpha^0(z)$ and $\delta^0(z)$ and a set of instrumental values $\mathcal{Z}$ if and only if the following three conditions hold for all $z \in \mathcal{Z}$.

I1. Observational equivalence. For $m \in \{1, \ldots, M\}$ and $i \in \{1, \ldots, K\}$

$$\beta_{mii}(z) = \bar{\alpha}^0_{mi}(z) \qquad \delta_i(z) = \delta^0_i(z)$$

I2. Independence. For $m \in \{1, \ldots, M\}$ and $i \in \{1, \ldots, K\}$

$$\sum_{j=1}^{K} \delta^0_j(z)\beta_{mij}(z) = \gamma_{mi}.$$

I3. Proper conditional distributions. For $(m, n) \in \{1, \ldots, M\}$ and $(i, j, k) \in \{1, \ldots, K\}$ if $\gamma_{mi} \leq \gamma_{nj}$ then $\beta_{mik}(z) \leq \beta_{njk}(z)$.

2.3.2 Geometry of the identified set

When determining whether a particular value of γ lies in the identified set, the ordering of the elements of γ is crucial in determining whether there exist distribution functions which satisfy condition I3.

There are $L \equiv (K(M - 1))!/((M - 1)!)^K$ admissible orderings of the N elements of γ which are not restricted to be zero or one.[5] For example, when $M = 3$ and $K = 2$, there are 6 of the possible 24 orderings that are admissible. The 18 inadmissible orderings have $\gamma_{11} > \gamma_{21}$ or $\gamma_{12} > \gamma_{22}$ or both.

Let l index the admissible orderings of γ. For each $l \in \{1, \ldots, L\}$ define sets $\mathcal{S}^0_l(z)$ and $\mathcal{H}^0_l(z)$ as follows.

$$\mathcal{S}^0_l(z) \equiv \{(\gamma, \beta(z), \delta(z)) : \gamma \text{ is in order } l \text{ and } (\gamma, \beta(z), \delta(z)) \text{ respects I1-I3}\}$$

[5]There are $(K(M - 1))!$ permutations of the free elements of γ. Amongst these only 1 in each $(M - 1)!$ have a sequence γ_i in ascending order and there are K such sequences to be considered so only 1 in each $((M - 1)!)^K$ have all these sequences in ascending order.

Table 2.1: Number of admissible orderings of gamma with (upper) and without (lower) monotonicity wth respect to X

M	Monotonicity with respect to X	K = 2	3	4	5
2	Yes	1	1	1	1
	No	2	6	24	120
3	Yes	2	5	14	42
	No	6	90	$2,520$	$113,400$
4	Yes	5	42	462	6006
	No	20	$1,680$	$369,600$	$168,168,000$
5	Yes	14	462	$24,024$	$1,662,804$
	No	70	$34,650$	$6,306,300$	$305,540,235,000$

$$\mathcal{H}_l^0(z) \equiv \left\{\gamma : \gamma \text{ is in order } l \text{ and } \exists\,(\beta(z), \delta(z)) \text{ s.t. } (\gamma, \beta(z), \delta(z)) \in \mathcal{S}_l^0(z)\right\}$$

The set $\mathcal{S}_l^0(z)$ is the set of structures admitted by the SEIV model that have γ in order l and deliver the distribution $F_{YX|Z}^0$ for $Z = z$. The set $\mathcal{H}_l^0(z)$ is the projection of this set onto the component γ, that is onto the structural function.

Since for any ordering, l, conditions I1-I3 comprise a system of linear equalities and inequalities, each set $\mathcal{S}_l^0(z)$ is convex, or empty. It follows, from consideration of the Fourier-Motzkin elimination algorithm[6], that the set $\mathcal{H}_l^0(z)$ is also defined by a system of linear equalities and inequalities, so it is also convex or empty.

The identified set of values of γ in order l obtained as z takes all values in the set of instrumental values $\mathcal{Z}$, denoted $\mathcal{H}_l^0(\mathcal{Z})$, is the following intersection of the sets $\mathcal{H}_l^0(z)$:

$$\mathcal{H}_l^0(\mathcal{Z}) \equiv \bigcap_{z \in \mathcal{Z}} \mathcal{H}_l^0(z)$$

which is convex or empty. The identified set of values of γ of all orders is the union of the sets $\mathcal{H}_l^0(\mathcal{Z})$, as follows.

$$\mathcal{H}^0(\mathcal{Z}) = \bigcup_{l=1}^{L} \mathcal{H}_l^0(\mathcal{Z})$$

Thus the identified set of values of γ, that is the identified set of structural functions, is a union of convex sets but that union may not itself be convex.

If there is a value l such that $\mathcal{H}_l(\mathcal{Z})$ contains values of γ in which no pair of elements have a common value and for more than one value of l there are sets $\mathcal{H}_l(\mathcal{Z})$ which are non-empty

[6]See Ziegler (2007).

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

then the identified set is not connected.

This is so because each set $\mathcal{H}_l(\mathcal{Z})$ lies in one of the $N!$ orthoschemes[7] of the unit N-cube and the orthoschemes have intersections only at their faces where there is equality of two or more elements of γ. In the example in Section 2.4 there are a number of cases in which the identified set is disconnected.

When instruments are strong or there are highly informative additional restrictions (for example parametric restrictions) the sets $\mathcal{H}_l(\mathcal{Z})$ may be empty for all but one value of l and then the identified set is convex. Otherwise the identified set may be very irregular and complex, composed of the union of a very large number of convex subsets of the identified set. With M and K as low as 4 the value of L is $369,600$ and as M or K increase the value of L quickly becomes astronomical.

Additional restrictions can bring some simplification. For example suppose the threshold functions are restricted to be monotone in a scalar explanatory variable X, with a common direction of dependence, say all non-decreasing.

The problem of finding the number of admissible orderings of γ under this restriction can be recast as the problem of finding the number of ways of filling a $(M-1) \times K$ matrix with the integers $\{1, 2, \ldots, (M-1)K\}$ such that the array increases both across columns and across rows. With $K = 2$ this is the Catalan number $\frac{1}{M+1}\binom{2(M-1)}{(M-1)}$ and the restriction of monotonicity with respect to$X$ brings an $(M-1)$-fold reduction in the number of admissible orderings.

Table 2.1 shows the value of L for values of M and K up to 5 together with the number of admissible orderings once monotonicity with respect to X is imposed.[8] The monotonicity restriction can bring large reductions in numbers of admissible orderings but when M or K are at all large there remain huge numbers of admissible orderings of γ.

[7]The orthoschemes of the unit cube are the regions within which points obeying a particular weak ordering of coordinate values lie. For example in a 3-cube within which lie (x, y, z) there are 6 orthoschemes defined by the inequalities $x \leq y \leq z$, $y \leq x \leq z$, etc. See Coxeter (1973).

[8]The row and column ascending matrices encountered here are special cases of Young Tableaux. The `NumberOfTableaux` command in the Combinatorica package (Pemmaraju and Skienka, 2003) of Mathematica (Wolfram Research, Inc., 2008) was used to compute those entries in Table 2.1 in which montonicity with respect to X is imposed.

2.3.3 Characterisation of the identified set

Chesher (2008) shows that all structural functions in the set identified by the SEIV model associated with a conditional distribution function $F^0_{YX|Z}$ and a set of instrumental values $\mathcal{Z}$ satisfy the following inequalities for all $u \in (0,1)$ and $z \in \mathcal{Z}$.

$$\text{Pr}_0[Y < h(X,u)|Z = z] < u \leq \text{Pr}_0[Y \leq h(X,u)|Z = z]$$

In terms of threshold functions these inequalities are as follows.

$$\sum_{m=1}^{M} \text{Pr}_0[(Y = m) \wedge (h_m(x) < u)|Z = z] < u \leq \sum_{m=1}^{M} \text{Pr}_0[(Y = m) \wedge (h_{m-1}(x) < u)|Z = z]$$

For the discrete endogenous variable case, there is the following representation.

$$\sum_{i=1}^{K} \sum_{m=1}^{M-1} \delta_i^0(z)\alpha_{mi}^0(z)1(\gamma_{mi} < u) < u \leq \sum_{i=1}^{K} \sum_{m=1}^{M} \delta_i^0(z)\alpha_{mi}^0(z)1(\gamma_{m-1,i} < u) \tag{2.4}$$

These inequalities have implications for γ as set out in the following Proposition which is proved in the Annex.

Proposition 1. *For any z, if the inequalities (2.4) hold for all $u \in (0,1)$ then for all $l \in \{1,\ldots,M\}$ and $s \in \{1,\ldots,K\}$ the following inequalities hold.*

$$\sum_{i=1}^{K} \sum_{m=1}^{M-1} \delta_i^0(z)\alpha_{mi}^0(z)1(\gamma_{mi} \leq \gamma_{ls}) \leq \gamma_{ls} \leq \sum_{i=1}^{K} \sum_{m=1}^{M} \delta_i^0(z)\alpha_{mi}^0(z)1(\gamma_{m-1,i} < \gamma_{ls}) \tag{2.5}$$

For any ordering l of γ let $\mathcal{C}_l^0(z)$ denote the set of values of γ that satisfy the inequalities (2.5) of Proposition 1. Since these inequalities define an intersection of halfspaces each set $\mathcal{C}_l^0(z)$ is convex or empty, as is its intersection

$$\mathcal{C}_l^0(\mathcal{Z}) = \bigcap_{z \in \mathcal{Z}} \mathcal{C}_l^0(z).$$

Define $\mathcal{C}^0(\mathcal{Z})$ as the set of values of γ of any ordering that satisfy the inequalities of Proposition 1 for all $z \in \mathcal{Z}$ when calculations are done using a distribution $F^0_{YX|Z}$. This is the union

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

of the sets $\mathcal{C}_l^0(\mathcal{Z})$:

$$\mathcal{C}^0(\mathcal{Z}) = \bigcup_{l=1}^{L} \mathcal{C}_l^0(\mathcal{Z})$$

and, like the identified set, $\mathcal{H}^0(\mathcal{Z})$, the set of values γ defined by the inequalities of Proposition 1, $\mathcal{C}^0(\mathcal{Z})$, is a union of convex sets. It may not itself be convex nor need it be connected.

Chesher (2008, 2009) shows that, when Y is binary, $\mathcal{C}^0(\mathcal{Z})$ is precisely the identified set, $\mathcal{H}^0(\mathcal{Z})$. When Y is not binary this may not be so.

This can be seen by considering Proposition 2, proved in the Annex. Proposition 2, which follows directly from conditions I1-I3, places restrictions on values of γ that lie in the identified set. It will be demonstrated in Section 2.4 that there can be values of γ which satisfy the inequalities of Proposition 1 and fail to satisfy the inequalities of Proposition 2.

Proposition 2. *If γ lies in the identified set associated with probabilities $\bar{\alpha}^0(z)$ and $\delta^0(z)$ for instrumental values, z, varying in $\mathcal{Z}$, then for all $(m,n) \in \{1,\ldots,M\}$ with $n > m$ and all $i \in \{1,\ldots,K\}$ there are the following inequalities, (i) for each $z \in \mathcal{Z}$:*

$$\gamma_{ni} - \gamma_{mi} \geq \delta_i^0(z)\left(\bar{\alpha}_{ni}^0(z) - \bar{\alpha}_{mi}^0(z)\right) \tag{2.6}$$

and (ii):

$$\gamma_{ni} - \gamma_{mi} \geq \max_{z\in\mathcal{Z}}\left(\delta_i^0(z)\left(\bar{\alpha}_{ni}^0(z) - \bar{\alpha}_{mi}^0(z)\right)\right). \tag{2.7}$$

Let $\mathcal{D}^0(\mathcal{Z})$ denote the set of values of γ that satisfy the system of inequalities (2.7) of Proposition 2. Since $\mathcal{D}^0(\mathcal{Z})$ is an intersection of halfspaces it is a convex set.

Values of γ that lie in the set identified by the SEIV model obey the inequalities of Proposition 1 and Proposition 2 so the identified set lies in the intersection of the sets defined by the inequalities of the two Propositions as stated in Proposition 3.

Proposition 3. *The identified set, $\mathcal{H}^0(\mathcal{Z})$, is a subset of $\tilde{\mathcal{C}}^0(\mathcal{Z}) \equiv \mathcal{C}^0(\mathcal{Z}) \cap \mathcal{D}^0(\mathcal{Z})$.*

Like $\mathcal{C}^0(\mathcal{Z})$ the set $\tilde{\mathcal{C}}^0(\mathcal{Z})$ is a union of convex sets as can be seen by expressing it as follows.

$$\tilde{\mathcal{C}}^0(\mathcal{Z}) = \bigcup_{l=1}^{L}\left(\mathcal{C}_l^0(\mathcal{Z}) \cap \mathcal{D}^0(\mathcal{Z})\right)$$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

When Y is *binary* the inequalities (2.6) of Proposition 2 reduce to the following.

$$\delta_i^0(z)\alpha_{1i}^0(z) \leq \gamma_{1i} \leq 1 + \delta_i^0(z)(1 - \alpha_{1i}^0(z)) \qquad i \in \{1, \ldots, K\} \tag{2.8}$$

The inequality (2.5) of Proposition 1 requires that

$$\sum_{j=1}^{i} \delta_j^0(z)\alpha_{1j}^0(z) \leq \gamma_{1i} \leq 1 + \sum_{j=i}^{K} \delta_j^0(z)(1 - \alpha_{1j}^0(z)) \qquad i \in \{1, \ldots, K\} \tag{2.9}$$

and it is clear that (2.8) is satisfied if (2.9) is satisfied. Therefore when Y is binary $\mathcal{C}^0(\mathcal{Z}) \subseteq \mathcal{D}^0(\mathcal{Z})$ so $\tilde{\mathcal{C}}^0(\mathcal{Z}) \equiv \mathcal{C}^0(\mathcal{Z})$ confirming the result of Chesher (2008) for the binary endogenous variable case: for binary Y, $\mathcal{C}^0(\mathcal{Z})$ *is* the identified set $\mathcal{H}^0(\mathcal{Z})$.

If the *explanatory* variable, X, is binary then $\tilde{\mathcal{C}}^0(\mathcal{Z})$ is the identified set, as stated in Proposition 4, which is proved in the Annex.

Proposition 4. *When X is binary $\mathcal{H}^0(\mathcal{Z}) = \tilde{\mathcal{C}}^0(\mathcal{Z})$ no matter how many points of support Y has.*

The inequalities defining $\tilde{\mathcal{C}}^0(\mathcal{Z})$ of Proposition 4 involve probabilities about which data is informative and the value γ that characterises a structural function. The values of the elements of $\beta(\mathcal{Z})$ that define the conditional distribution functions of U given X and Z do not appear in these inequalities. So Proposition 4 points the way to fast computation of the identified set. In Section 2.4 it provides the basis for computations that illustrate identified sets in a parametrically restricted ordered probit model with a binary endogenous variable and from $M = 2$ to $M = 130$ points of support for the ordered outcome Y.

2.4 Discreteness and identified sets in a parametric ordered probit model

2.4.1 Models

We now investigate the nature of the identified sets delivered by a parametric ordered probit model with a binary endogenous variable. In this model the structural function is paramet-

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

rically specified, as follows.

$$Y = \begin{cases} 1 & , & 0 \leq & U & \leq \Phi(s^{-1}(c_1 - a_0 - a_1 X)) \\ 2 & , & \Phi(s^{-1}(c_1 - a_0 - a_1 X)) < & U & \leq \Phi(s^{-1}(c_2 - a_0 - a_1 X)) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ M & , & \Phi(s^{-1}(c_{M-1} - a_0 - a_1 X)) < & U & \leq 1 \end{cases} \tag{2.10}$$

Here Φ denotes the standard normal distribution function, the constants $c_1, \ldots, c_{M-1}$ are threshold values defining cells within which a latent normal random variable is classified, and a_0, a_1 and s are constant parameters. Throughout X is binary with support $\{-1, +1\}$, There is the independence restriction: $U \perp Z$, U is normalised $Unif(0,1)$.

In one portfolio of illustrations (A) the model specifies the values of the threshold parameters $c_1, \ldots, c_{M-1}$ as known, and s as known and normalised to one. This leaves just two unknown parameters, a_0 and a_1, and it is easy to display the identified sets graphically. In these illustrations M, the number of levels of the outcome, is varied from 2 to 130.

In another illustration (B) M is held fixed at 3 and the model specifies the thresholds, c_1 and c_2, along with the slope coefficient, a_1, as unknown parameters. In these illustrations the values of a_0 and s are normalised to respectively 0 and 1.

In all cases the instrumental variable takes equally spaced values in the interval $[-1, 1]$.

There are a number of reasons for choosing this particular parametric model and set up for this exercise.

1. Many researchers doing applied work will base their analysis on parametric models and the ordered probit model is a leading case considered in practice.

2. When studying the impact of the discreteness of the outcome on identified sets it is convenient to have objects like the parameters a_0 and a_1 which remain stable with a common meaning as the discreteness of the outcome is varied.

3. The number of unknown objects in a fully nonparametric analysis is $N = K(M-1)$ and the identified set can be highly complex comprising the union of an enormous number of sets associated with each possible ordering of the N values delivered by the structural function - see Table 2.1. The parametric model severely restricts the number of feasible

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

orderings and, as explained below, it is not necessary to search across many possible orderings when determining the extent of the identified set.

2.4.2 Calculation procedures

The calculation of an identified set of parameter values for a particular distribution $F^0_{YX|Z}$ and set of instrumental values $\mathcal{Z}$ proceeds as follows.

A fine grid of values of the parameters (e.g. a_0 and a_1 in the illustrations in set A) is defined. A value, say (a_0^*, a_1^*) is selected from the grid and the value of γ, say γ^*, determined by (a_0^*, a_1^*) is calculated. Recall that γ is a list of values of the threshold functions defined by a model at the points of support of the discrete endogenous variable.

With a value γ^* to hand the ordering of its elements, say l^*, is determined and the linear equalities and inequalities defining the convex set $\mathcal{H}^0_{l^*}(\mathcal{Z})$ can be calculated. In all the illustrations, because X is binary, $\mathcal{H}^0_{l^*}(\mathcal{Z}) = \tilde{\mathcal{C}}^0_{l^*}(\mathcal{Z})$. If γ^* falls in this set then (a_0^*, a_1^*) is in the identified set, otherwise it is not.

Passing across the grid the identified set is computed. Care is required because the set may not be connected and sometimes component connected subsets of the identified set can be small. To avoid missing component subsets, dense grids of values are used in the calculations presented here.

2.4.3 Illustration A1

The probability distributions used in this illustration are generated by triangular Gaussian structures with structural equations as follows.

$$
\begin{aligned}
Y^* &= \alpha_1 X + W \\
X^* &= 0.5Z + V
\end{aligned}
$$

$$
Y = \begin{cases} 1 \;, & -\infty \leq Y^* \leq c_1 \\ 2 \;, & c_1 < Y^* \leq c_2 \\ \vdots \;\; \vdots & \quad\vdots \quad \vdots \quad \vdots \\ M \;, & c_{M-1} < Y^* \leq +\infty \end{cases} \qquad X = \begin{cases} -1 \;, & -\infty \leq X^* \leq 0 \\ +1 \;, & 0 < X^* \leq +\infty \end{cases}
$$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Table 2.2: Illustration A1: Threshold values

Number of Classes: M	Threshold Values (c_i)	Shading in Figure 1
2	$\{0.0\}$	red
4	$\{\pm 0.1, 0.0\}$	blue
6	$\{\pm 0.3, \pm 0.1, 0.0\}$	red
8	$\{\pm 0.7, \pm 0.3, \pm 0.1, 0.0\}$	blue
10	$\{\pm 1.1, \pm 0.7, \pm 0.3, \pm 0.1, 0.0\}$	red
12	$\{\pm 1.5, \pm 1.1, \pm 0.7, \pm 0.3, \pm 0.1, 0.0\}$	green
14	$\{\pm 1.8, \pm 1.5, \pm 1.1, \pm 0.7, \pm 0.3, \pm 0.1, 0.0\}$	black

The value of α_1 in this illustration is 1 and the distribution of (W, V) is specified to be Gaussian and independent of Z.

$$\begin{bmatrix} W \\ V \end{bmatrix} | Z \sim N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix} \right)$$

These structures are closely related to a special case of the parametric Gaussian models of discrete outcomes studied in Heckman (1978).

Expressed in terms of a random variable U which is uniformly distributed on the unit interval the structural functions are as follows.

$$h(X, U) = \begin{cases} 1 & , \quad 0 \leq \; U \; \leq \Phi(c_1 + X) \\ 2 & , \quad \Phi(c_1 + X) < \; U \; \leq \Phi(c_2 + X) \\ \vdots & \vdots \quad \vdots \quad \vdots \quad \vdots \\ M & , \quad \Phi(c_{M-1} + X) < \; U \; \leq 1 \end{cases}$$

There are 10 values in $\mathcal{Z}$ as follows.

$$\mathcal{Z} = \{\pm 1.0, \pm 0.777, \pm 0.555, \pm 0.333, \pm 0.111\}$$

In this illustration the number of classes in which Y is observed is increased from 2 through 14 with threshold values as set out in Table 2.2.

Identified sets for the two parameters, (a_0, a_1), are drawn in Figure 1. The sets are rhombuses arranged with edges parallel to 45° and 225° lines. Identified sets are superimposed one upon another.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

The largest rhombus drawn in Figure 1 is the identified set with $M = 2$. Because the outcome is binary this is the set $\mathcal{C}^0(\mathcal{Z})$.

The identified set with $M = 4$ is the rhombus comprising the lowest blue chevron and what lies above it but *excluding a narrow strip* on the edge of the two upper boundaries. This narrow strip (coloured dark blue) is the set $\mathcal{C}^0(\mathcal{Z})\cap\overline{\mathcal{D}^0(Z)}$. Notice that this does not extend all the way along the upper edges of the set because for the case $M = 2$, $\tilde{\mathcal{C}}^0(\mathcal{Z}) = \mathcal{C}^0(\mathcal{Z}) \subseteq \mathcal{D}^0(\mathcal{Z})$.

The identified set with $M = 6$ (respectively 8) is the rhombus comprising the second lowest red (respectively blue) chevron and all that lies above it apart from the narrow dark blue shaded strip on the edge of the two upper boundaries.

The identified set with $M = 10$ is disconnected and comprises the two small red shaded rhombuses in the upper part of the picture. The identified set when $M = 12$ is the small green shaded rhombus in the centre of the picture and the identified set when $M = 14$ is the tiny black shaded rhombus at the intersection of the horizontal and vertical dashed lines. Further increases in numbers of classes deliver sets which are barely distinguishable from points at the scale chosen for Figure 1.

As the number of classes rises the extent of the identified sets falls rapidly but the passage towards point identification is complex and even when the sets are quite small they can be disconnected.

2.4.4 Illustration A2

In this illustration the class of structures generating probability distributions is as in Illustration A1 and, as there, $\alpha_1 = 1$. But there are now 5 values in Z as follows

$$\mathcal{Z} = \{\pm 1.0, \pm 0.5, 0.0\}$$

and the number of classes is varied through the following sequence.

$$M \in \{2, 4, 6, 8, 10, 12, 14, 16, 18, 25, 50, 75\}$$

Threshold values are chosen to "cover" the main probability mass of the distribution of Y marginal with respect to X and Z. They are chosen as quantiles of a $N(0, (2.4)^2)$ distribution

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

associated with equally spaced probabilities in $[0, 1]$, e.g. $\{1/2\}$ for $M = 2$, $\{1/3, 2/3\}$ for $M = 3$. The identified sets are drawn in Figure 2-5.

Figure 2 shows identified sets for $M = 2$ (red), $M = 4$ (blue) and $M = 6$ (green). Notice that in the latter two cases the identified sets are disconnected comprising two rhombuses. On the upper edges of the upper rhombus in the case $M = 4$ is a narrow dark blue strip marking the intersection $\mathcal{C}^0(\mathcal{Z}) \cap \overline{\mathcal{D}^0(Z)}$ which does not lie in the identified set. This intersection is *empty* in the other cases shown in this Figure and in Figures 3 - 5.

Figure 3 shows identified sets for $M = 8$ (red), $M = 10$ (blue) and $M = 12$ (green). The identified set for $M = 10$ is disconnected. Notice that the scale is greatly expanded in this Figure - the identified sets are rapidly decreasing in size as the number of classes observed for Y increases. The outline unshaded rhombus in Figure 3 is the identified set for $M = 6$ copied across from Figure 2. Boxes formed by the dashed lines in Figure 2 show the region focussed on in Figure 3.

Figure 4 shows identified sets for $M = 14$ (red), $M = 16$ (blue) and $M = 18$ (green). Again the scale is greatly expanded relative to the previous Figure. The outline unshaded rhombus is the identified set for $M = 12$ copied across from Figure 3.

Figure 5 shows identified sets for $M = 25$ (red), $M = 50$ (blue) and $M = 75$ (green). Yet again the scale is greatly expanded relative to the previous Figure. The lower part of the identified set for $M = 18$ is drawn in outline. All the identified sets are connected and of very small extent. The situation is now very close to point identification. The identified set at $M = 100$ is not distinguishable from a point at the chosen scale.

The two panes of Figure 6 plot logarithm (base e) of the lengths of identified intervals for a_0 and a_1 against the logarithm of the number of classes in which Y is observed. Figure 7 plots the logarithm of he area of the identified set for a_0 and a_1 against the logarithm of the number of classes. In each case the points are quite tightly scattered around a negatively sloped linear relationships suggesting approach to point identification at a rate proportional to a power of the number of classes[9]. OLS estimates indicate that the lengths of the sets for a_0 and a_1 both fall at a rate proportional to $M^{-2.1}$ and that the area of the identified set for a_0 and a_1 falls at a rate proportional to $M^{-3.6}$.

[9]Where sets are disconnected the lengths of the identified sets for individual parameters are the calculated as the sum of the lengths of disjoint intervals and the area of the sets for a pair of parameters is calculated as the sum of the areas of the connected component sets.

The fine details of this approach and the geometry of the identified sets depends on fine details of the specification of the structures generating the probability distributions such as the precise spacing of the thresholds.

2.4.5 Illustration B1

The class of structures generating probability distributions is as in Illustration A1 and, as in that illustration there are 10 values in $\mathcal{Z}$, as follows.

$$\mathcal{Z} = \{\pm 1, \pm 0.777, \pm 0.555, \pm 0.333, \pm 0.111\}$$

In this illustration there are $M = 3$ classes throughout. The values of a_0 and s are normalised to respectively zero and one. The unknown parameters are the thresholds c_1 and c_2 and the slope coefficient a_1. This is the sort of set up one finds when modelling attitudinal data where threshold values are unknown parameters of considerable interest.

In the structure generating the probability distributions the values of the thresholds are as follows

$$(c_1, c_2) = (-0.667, +0.667)$$

and $\alpha_1 = 1$.

The identified set resides in a 3-dimensional square prism of infinite extent: $\mathbb{R} \times (0,1)^2$. Figures 8, 9 and 10 show slices taken through this at a sequence of values of a_1 showing at each chosen value of a_1 the associated identified set for (c_1, c_2). In all cases this is connected and resides in the upper orthoscheme of the unit square because the restriction $c_2 > c_1$ has been imposed.

In each case the rectangular regions (shaded red and green) indicate combinations of (c_1, c_2) which at the chosen value of a_1 lie in the set $\mathcal{C}^0(\mathcal{Z})$. The green shaded regions indicate combinations of (c_1, c_2) that at the chosen value of a_1 are in the intersection $\mathcal{C}^0(\mathcal{Z}) \cap \overline{\mathcal{D}^0(Z)}$. These combinations of (a_1, c_1, c_2) do not lie in the identified set. The red shaded regions indicate combinations of (c_1, c_2) that at the chosen value of a_1 are in the intersection $\tilde{\mathcal{C}}^0(\mathcal{Z}) = \mathcal{C}^0(\mathcal{Z}) \cap \mathcal{D}^0(\mathcal{Z})$. These combinations of (a_1, c_1, c_2) are in the identified set.

The extent of the regions in the $c_1 \times c_2$ plane grows as a_1 falls towards the value 1.0 and

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

then shrinks as a_1 falls further.

2.5 Concluding remarks

Single equation instrumental variable models for ordered discrete outcomes generally set identify structural functions or, if there are parametric restrictions, parameter values. Complete models, for example the triangular control function model, can be point identifying, but in applied econometric practice there may be no good reason to choose one point identifying model over another.

For any particular distribution of observable variables the sets delivered by the SEIV model give information about the variety of structural functions or parameter values that would be delivered by one or another of the point identifying models which are restricted versions of the SEIV model.

For the nonparametric case we have developed a system of equalities and inequalities that bound the identified sets of structural functions delivered by a SEIV model in the case when endogenous variables are discrete. We have shown that when either the outcome or the endogenous variable is binary the inequalities sharply define the identified set. The inequalities involve only probabilities about which data is informative and the identified sets can be estimated and inferences drawn using the methods set out in Chernozhukov, Lee and Rosen (2009). Some illustrative calculations for the binary outcome case are given in Chesher (2009).

Calculations in a parametric model suggest that the degree of ambiguity attendant on using the SEIV model reduces rapidly as the discreteness of the outcome is reduced. Research to determine the extent to which this is true in less restricted settings is one of a number of topics of current research.

References

Blundell R., and J. Powell (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in Dewatripont, M., L.P. Hansen, and S.J. Turnovsky, eds., *Advances* *in Economics and Econometrics: Theory and Applications, Eighth World Congress, Vol. II.* Cambridge: Cambridge University Press.

Blundell, R., and J. Powell (2004): "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, 71, 655-679.

Chernozhukov, V., And C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245-261.

Chernozhukov, V., S. Lee and A. Rosen (2009): "Intersection Bounds: Estimation and Inference," CeMMAP Working Paper 19/09.

Chesher, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405-1441.

Chesher, A. (2005): "Nonparametric Identification under Discrete Variation," *Econometrica*, 73, 1525-1550.

Chesher, A. (2008): "Instrumental Variable Models for Discrete Outcomes," CeMMAP Working Paper 30/08, forthcoming in *Econometrica*.

Chesher, A. (2009): "Single Equation Endogenous Binary Response Models," CeMMAP Working Paper 23/09.

Coxeter, H. (1973): *Regular Polytopes.* New York: Dover.

Greene, W. (2007): *LIMDEP 9.0 Reference Guide.* Econometric Software, Inc., New York.

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

Ihaka, R., and R. Gentleman (1996): "R: A Language for Data Analysis and Graphics," *Journal of Computational and Graphical Statistics*, 5, 299-314.

Manski, C. (2003): *Partial Identification of Probability Distributions.* New York: Springer-Verlag.

Pemmaraju, S., and S. Skienka (2003); *Combinatorial Discrete Mathematics.* New York: Cambridge University Press.

Rivers, D. and Q. Vuong (1988): "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics* 39, 347-366.

Smith, R., and R. W. Blundell (1986): "An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply," *Econometrica* 54, 679-685.

Statacorp (2007): "Stata Statistical Software: Release 10." College Station, TX: StataCorp LP.

Wolfram Research Inc. (2008): *Mathematica Edition: Version 7.0.* Champaign IL: Wolfram Research Inc.

Ziegler, G. (2007): *Lectures on Polytopes.* (Updated seventh printing of the first edition). New York: Springer Science.

Annex: Proofs of Propositions

Proof of Proposition 1. Consider some arrangement of the elements of γ in which two elements, $\gamma_{kr} < \gamma_{ls}$ are adjacent so that there is no element $\gamma_{qt} \in \gamma$ satisfying $\gamma_{kr} < \gamma_{qt} < \gamma_{ls}$. Consider $u \in (\gamma_{kr}, \gamma_{ls}]$ and the right hand side of (2.4), reproduced here.

$$u \leq \sum_{i=1}^{K} \sum_{m=1}^{M} \delta_i^0(z)\alpha_{mi}^0(z)1(\gamma_{m-1,i} < u)$$

This inequality must hold for all u in $(\gamma_{kr}, \gamma_{ls}]$ and so must hold at the supremum of the interval which is its maximal value, γ_{ls}, and so there is:

$$\gamma_{ls} \leq \sum_{i=1}^{K} \sum_{m=1}^{M} \delta_i^0(z)\alpha_{mi}^0(z)1(\gamma_{m-1,i} < \gamma_{ls})$$

which is the right hand side of (2.5).

Now consider some arrangement of the elements of γ in which two elements, $\gamma_{ls} < \gamma_{pr}$ are adjacent so that there is no element γ_{qt} satisfying $\gamma_{ls} < \gamma_{qt} < \gamma_{pr}$. Consider $u \in (\gamma_{ls}, \gamma_{pr}]$ and the left hand side of (2.4), reproduced here.

$$\sum_{i=1}^{K} \sum_{m=1}^{M-1} \delta_i^0(z)\alpha_{mi}^0(z)1(\gamma_{mi} < u) < u$$

This inequality must hold for all u in $(\gamma_{ls}, \gamma_{pr}]$ and so must hold as in (2.4) with *strong* inequalities at every value of u *in* the interval and so with *weak* inequalities at the *infimum* of the interval which is γ_{ls}, and so there is:

$$\sum_{i=1}^{K} \sum_{m=1}^{M-1} \delta_i^0(z)\alpha_{mi}^0(z)1(\gamma_{mi} \leq \gamma_{ls}) \leq \gamma_{ls}$$

which is the left hand side of (2.5). $\square$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Proof of Proposition 2. Since γ is in the identified set for each $z \in \mathcal{Z}$ there exists a distribution function characterised by $\beta(z)$ satisfying conditions I1-I3. Conditions I1 and I2 imply that:

$$\begin{aligned}\gamma_{ni} &= \delta_i^0(z)\bar{\alpha}_{ni}^0(z) + \sum_{j \neq i} \delta_j^0(z)\beta_{nij}(z) \\ \gamma_{mi} &= \delta_i^0\bar{\alpha}_{mi}^0(z) + \sum_{j \neq i} \delta_j^0(z)\beta_{mij}(z)\end{aligned}$$

and the result (i) follows on subtracting and noting that the properness condition I3 ensures that for, each i and j, $\beta_{nij}(z) \geq \beta_{mij}(z)$ because $n > m$. The result (ii) follows directly on intersecting the intervals obtained at each value $z \in \mathcal{Z}$. $\square$

Proof of Proposition 4. Consider candidate structural functions, that is, values of γ_{m1} and γ_{m2}, $m \in \{1, \ldots, M-1\}$. Define $\beta(\mathcal{Z})$ so that conditions I1 and I2 are satisfied for all $z \in \mathcal{Z}$. There is only one way to do this: for each m, to satisfy Condition I1:

$$\beta_{m11}(z) = \bar{\alpha}_{m1}^0(z) \qquad \beta_{m22}(z) = \bar{\alpha}_{m2}^0(z) \tag{2.11}$$

and to satisfy Condition I2:

$$\begin{aligned}\delta_1^0(z)\beta_{m11}(z) + \delta_2^0(z)\beta_{m12}(z) &= \gamma_{m1} \\ \delta_1^0(z)\beta_{m21}(z) + \delta_2^0(z)\beta_{m22}(z) &= \gamma_{m2}\end{aligned}$$

and, on combining these results, for $m \in \{1, \ldots, M\}$ there are the following expressions

$$\beta_{m12}(z) = \frac{\gamma_{m1} - \delta_1^0(z)\bar{\alpha}_{m1}^0(z)}{\delta_2^0(z)} \qquad \beta_{m21}(z) = \frac{\gamma_{m2} - \delta_2^0(z)\bar{\alpha}_{m2}^0(z)}{\delta_1^0(z)} \tag{2.12}$$

It is now shown that for every $\gamma \in \tilde{\mathcal{C}}^0(\mathcal{Z})$ the value of $\beta(\mathcal{Z})$ defined by (2.11) and (2.12) as z varies across $\mathcal{Z}$ satisfies the properness condition I3. It follows that $\tilde{\mathcal{C}}^0(\mathcal{Z}) \subseteq \mathcal{H}^0(Z)$ and Proposition 3 states that $\mathcal{H}^0(Z) \subseteq \tilde{\mathcal{C}}^0(\mathcal{Z})$, so it must be that $\mathcal{H}^0(Z) = \tilde{\mathcal{C}}^0(\mathcal{Z})$ in this binary endogenous variable case.

To proceed, consider the distribution function characterised by $\beta_{mj1}(z)$ for $m \in \{1, \ldots, M-1\}$ and $j \in \{1, 2\}$ and any $z \in \mathcal{Z}$. Here conditioning is on $X = x_1$ and $Z = z$. The argument

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

when conditioning is on $X = x_2$ goes on similar lines and can be worked through by exchange of indices in what follows.

Condition I3 is satisfied if for every adjacent pair of values $\gamma_{si} < \gamma_{tj}$:

$$\beta_{si1}(z) \leq \beta_{tj1}(z)$$

and there are four possibilities to consider as follows.

A1 $i = 1$, $j = 1$. In this case $t = s + 1$ because $\gamma_{s1} < \gamma_{t1}$ are adjacent. Properness requires that $\beta_{s11} \leq \beta_{s+1,11}$ but (2.11) ensures that this holds because $\beta_{s11} = \bar{\alpha}^0_{s1}(z) \leq \bar{\alpha}^0_{s+1,1}(z) = \beta_{s+1,11}$.

A2 $i = 1$, $j = 2$. Properness requires that $\beta_{s11} \leq \beta_{t21}$ which, on using (2.11) and (2.12), requires that:

$$\bar{\alpha}^0_{s1}(z) \leq \frac{\gamma_{t2} - \delta^0_2(z)\bar{\alpha}^0_{t2}(z)}{\delta^0_1(z)}$$

which is written as follows.

$$\delta^0_1(z)\bar{\alpha}^0_{s1}(z) + \delta^0_2(z)\bar{\alpha}^0_{t2}(z) \leq \gamma_{t2} \tag{2.13}$$

If $\gamma \in \mathcal{C}^0(z)$ then the inequality (2.5) holds and, on its left hand side, replacing γ_{ls} by γ_{t2} there is:

$$\sum_{i=1}^{K} \sum_{m=1}^{M-1} \delta^0_i(z)\alpha^0_{mi}(z)1(\gamma_{mi} \leq \gamma_{t2}) \leq \gamma_{t2} \tag{2.14}$$

and since $\gamma_{s1} < \gamma_{t2}$ and the values are adjacent the left hand side of (2.14) as follows:

$$\delta^0_1(z)\sum_{m=1}^{s} \alpha^0_{m1}(z) + \delta^0_2(z)\sum_{m=1}^{t} \alpha^0_{m2}(z) = \delta^0_1(z)\bar{\alpha}^0_{s1}(z) + \delta^0_2(z)\bar{\alpha}^0_{t2}(z)$$

and so (2.13) holds.

A3 $i = 2$, $j = 1$. Properness requires that $\beta_{s21} \leq \beta_{t11}$ which, on using (2.11) and (2.12), requires that:

$$\frac{\gamma_{s2} - \delta^0_2(z)\bar{\alpha}^0_{s2}(z)}{\delta^0_1(z)} \leq \bar{\alpha}^0_{t1}(z)$$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

which is written as follows.

$$\gamma_{s2} \leq \delta_1^0(z)\bar{\alpha}_{t1}^0(z) + \delta_2^0(z)\bar{\alpha}_{s2}^0(z) \tag{2.15}$$

If $\gamma \in \mathcal{C}^0(z)$ then the inequality (2.5) holds and, on its right hand side, replacing γ_{ls} by γ_{s2} there is:

$$\gamma_{s2} \leq \sum_{i=1}^{K}\sum_{m=1}^{M} \delta_i^0(z)\alpha_{mi}^0(z)1(\gamma_{m-1,i} < \gamma_{s2}) \tag{2.16}$$

and since $\gamma_{s2} < \gamma_{t1}$ and the values are adjacent the right hand side of (2.16) is as follows:

$$\delta_1^0(z)\sum_{m=1}^{t}\alpha_{m1}^0(z) + \delta_2^0(z)\sum_{m=1}^{2}\alpha_{m2}^0(z) = \delta_1^0(z)\bar{\alpha}_{t1}^0(z) + \delta_2^0(z)\bar{\alpha}_{s2}^0(z)$$

and so (2.15) holds.

A4 $i = 2$, $j = 2$. It must be that $t = s+1$ because $\gamma_{s2} < \gamma_{t2}$ are adjacent. Properness requires that $\beta_{s21} \leq \beta_{s+1,21}$ which, on using (2.12), requires that:

$$\frac{\gamma_{s2} - \delta_2^0(z)\bar{\alpha}_{s2}^0(z)}{\delta_1^0(z)} \leq \frac{\gamma_{s+1,2} - \delta_2^0(z)\bar{\alpha}_{s+1,2}^0(z)}{\delta_1^0(z)}$$

which is written as follows.

$$\delta_2^0(z)\alpha_{s+1,2}^0(z) \leq \gamma_{s+1,2} - \gamma_{s2} \tag{2.17}$$

If $\gamma \in \mathcal{D}^0(z)$ then the inequality (2.6) of Proposition 2 holds and replacing γ_{ni} and γ_{mi} by respectively $\gamma_{s+1,2}$ and γ_{s2} gives the following:

$$\gamma_{s+1,2} - \gamma_{s2} \geq \delta_2^0(z)\left(\bar{\alpha}_{s+1,2}^0(z) - \bar{\alpha}_{s2}^0(z)\right) = \delta_2^0(z)\alpha_{s+1,2}^0(z)$$

and so (2.17) holds.

It has been shown that for any $z \in \mathcal{Z}$ and for all $\gamma \in \tilde{\mathcal{C}}^0(z) = \mathcal{C}^0(z) \cap \mathcal{D}^0(z)$ there are conditional distribution functions characterised by $\beta(z)$ defined as in (2.11) and (2.12) such that conditions I1, I2 and I3 hold.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Let $\beta(\mathcal{Z})$ be the conditional distribution functions generated using the definitions (2.11) and (2.12) as z varies within $\mathcal{Z}$. Since $\tilde{\mathcal{C}}^0(\mathcal{Z}) = \bigcap_{z \in \mathcal{Z}} \tilde{\mathcal{C}}^0(z)$, values $\gamma \in \tilde{\mathcal{C}}^0(\mathcal{Z})$ lie in every set $\tilde{\mathcal{C}}^0(z)$ and so for each such value of γ there are conditional distribution functions in $\beta(\mathcal{Z})$ such that conditions I1, I2 and I3 are satisfied. It follows that $\tilde{\mathcal{C}}^0(\mathcal{Z}) \subseteq \mathcal{H}^0(Z)$ and since $\mathcal{H}^0(Z) \subseteq \tilde{\mathcal{C}}^0(\mathcal{Z})$, it follows that $\mathcal{H}^0(Z) = \tilde{\mathcal{C}}^0(\mathcal{Z})$. $\square$

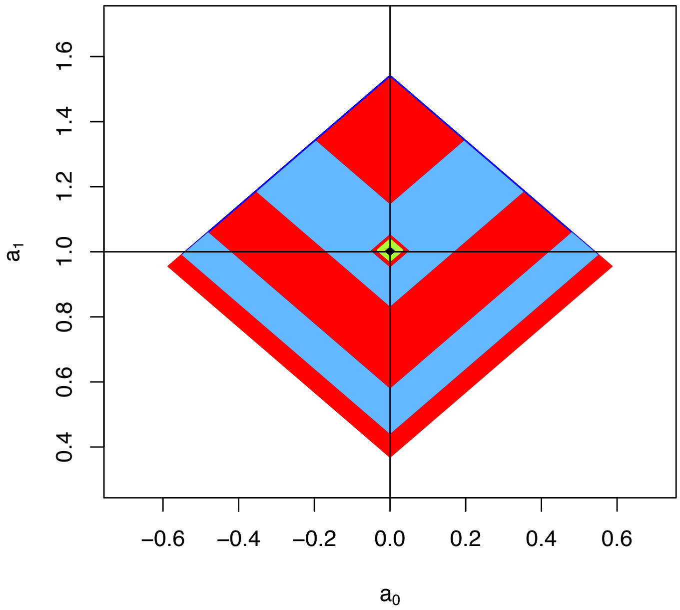


Figure 2.1: **Illustration A1.** Outer sets and identified sets in a binary endogenous variable SEIV model with a parametric ordered probit structural function with threshold functions of the form $\Phi(c_i - a_0 - a_1 x)$ as the number of categories of the outcome varies from 2 to 10. The dark blue strip at the upper margin of the rhombuses is not part of the identified sets.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

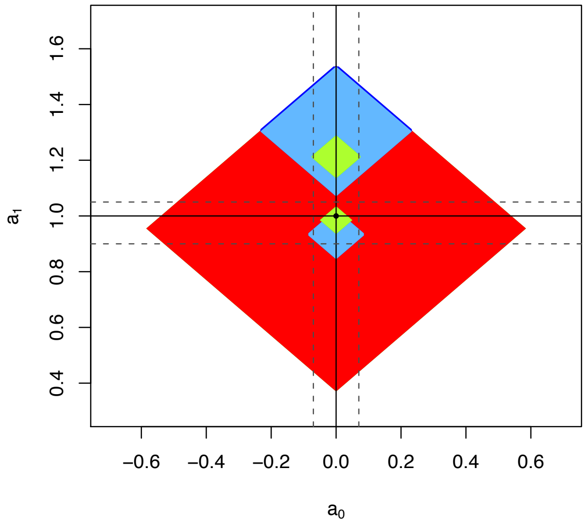


Figure 2.2: **Illustration A2.** Outer sets and identified sets delivered by a binary endogenous variable SEIV model with a parametric ordered probit structural function, intercept a_0 and slope a_1. Number of categories of the otucome, M: 2(red), 4(blue) and 6(green). The dark blue strip at the upper margin is not in the identified sets.

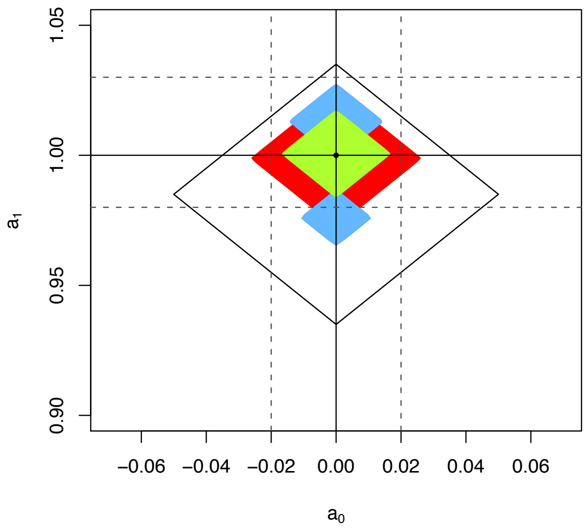


Figure 2.3: **Illustration A2.** Identified sets delivered by a binary endogenous variable SEIV model with a parametric ordered probit structural function, intercept a_0 and slope a_1. Number of categories of the outcome, M: 8(red), 10(blue) and 12(green).

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

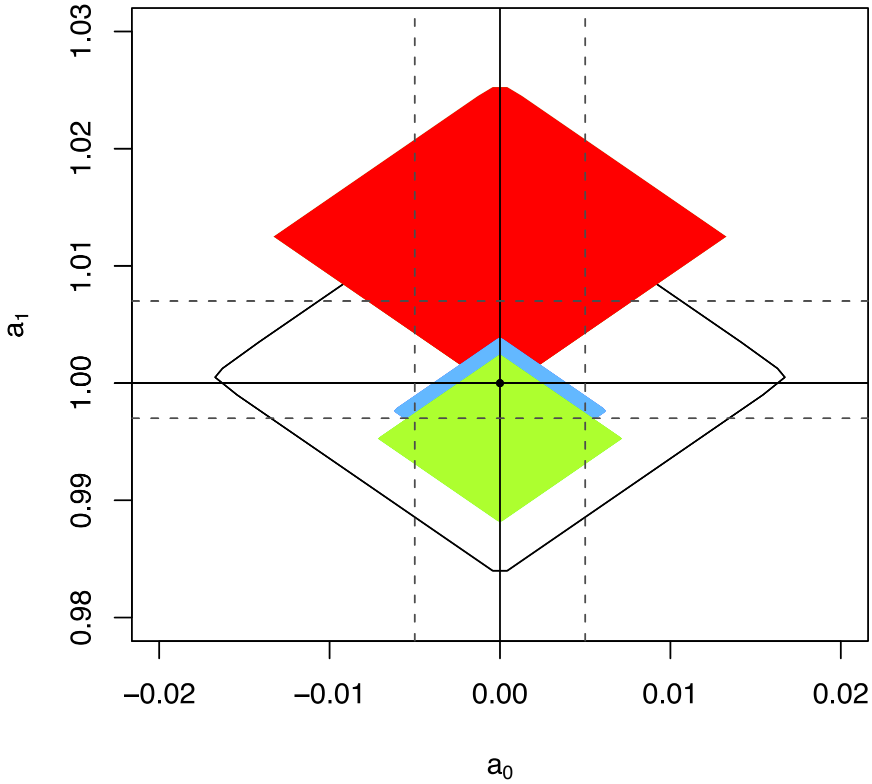


Figure 2.4: **Illustration A2.** Identified sets delivered by a binary endogenous variable SEIV model with a parametric ordered probit structural function, intercept a_0 and slope a_1. Number of categories of the outcome, M:14(red), 16(blue) and 18(green).

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

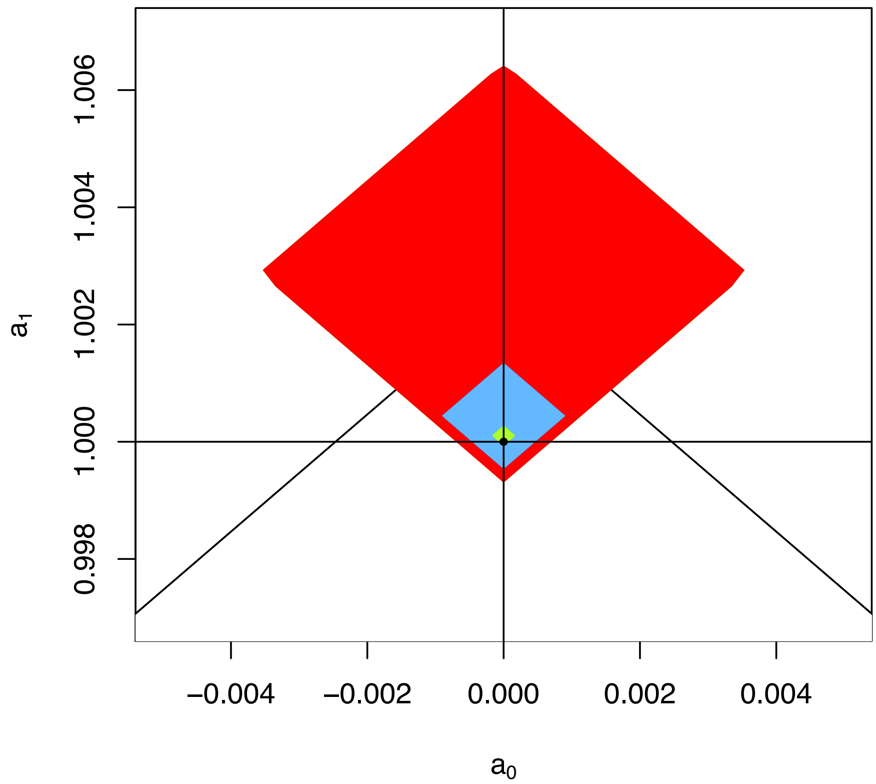
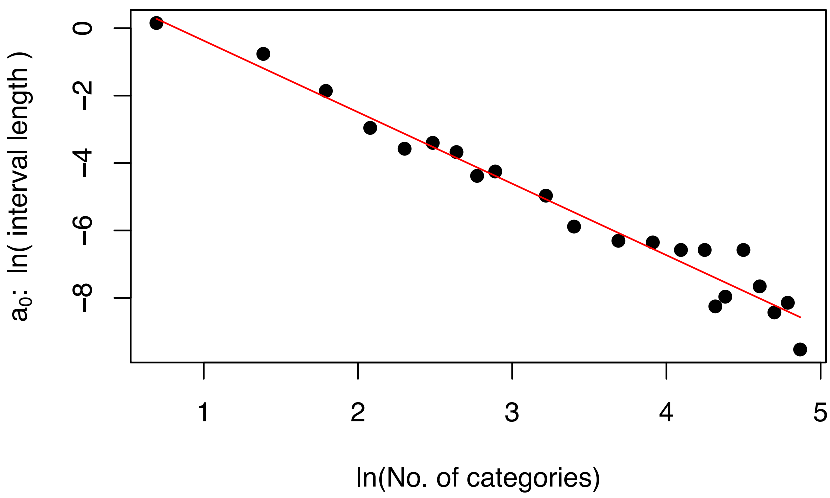


Figure 2.5: **Illustration A2.** Identified sets delivered by a binary endogenous variable SEIV model with a parametric ordered probit structural function, intercept a_0 and slope a_1. Number of categories of the outcome, M: 25(red), 50(blue) and 75(green).

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896



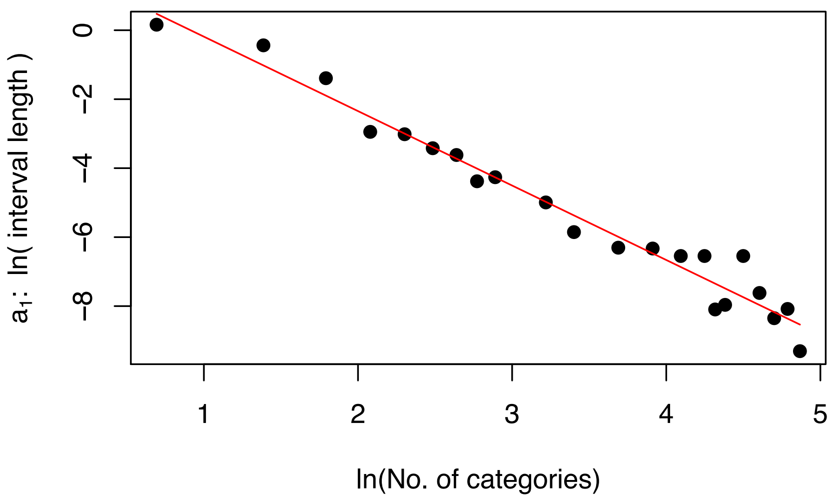


Figure 2.6: **Illustration A2.** Reduction of identified set as the number of outcome categories increases: (upper pane) logarithm of length of the identified interval for a_0 plotted against logarithm of number of categories of the outcome, Y, (lower pane) logarithm of length of the identified interval for a_1 plotted against logarithm of number of categories of the outcome, Y.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

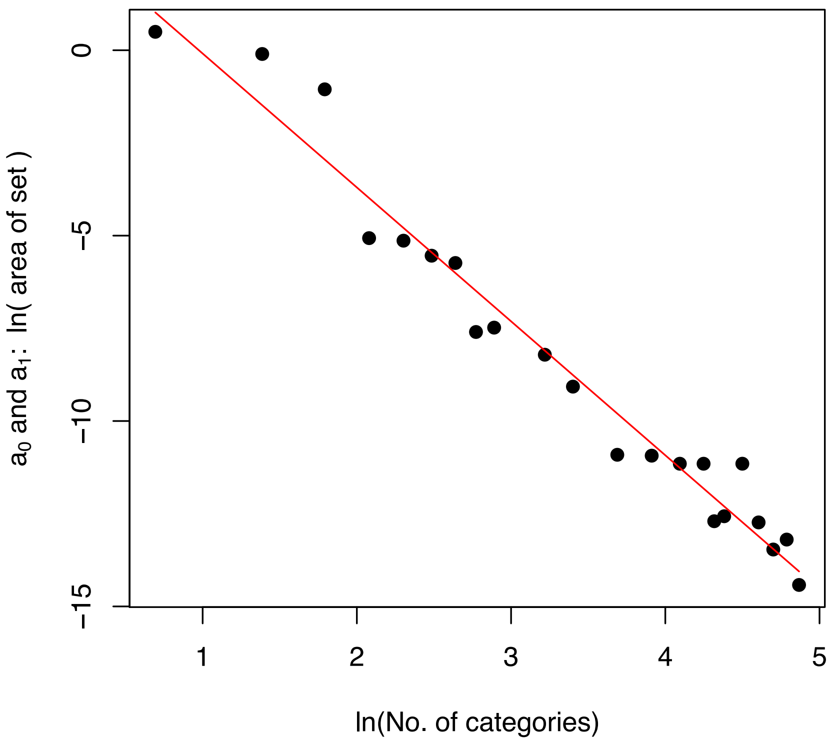


Figure 2.7: **Illustration A2.** Reduction of identified set as the number of outcome categories increases. Logarithm of area of the identified set plotted against logarithm of number of categories of the outcome, Y.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

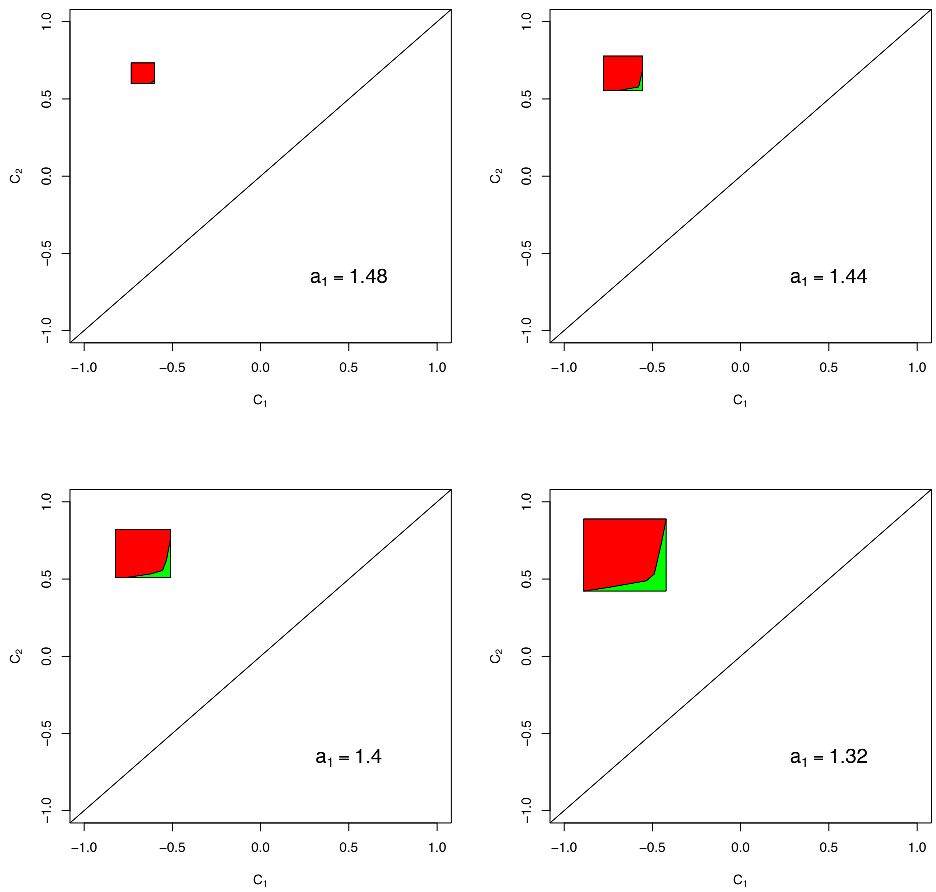


Figure 2.8: **Illustration B1.** Three class ordered probit model with unknown threshold parameters c_1 and c_2 and slope coefficient a_1. Cross-section of the identified set (red) and outer set (red and green) for c_1, c_2 and a_1 at selected values of a_1.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

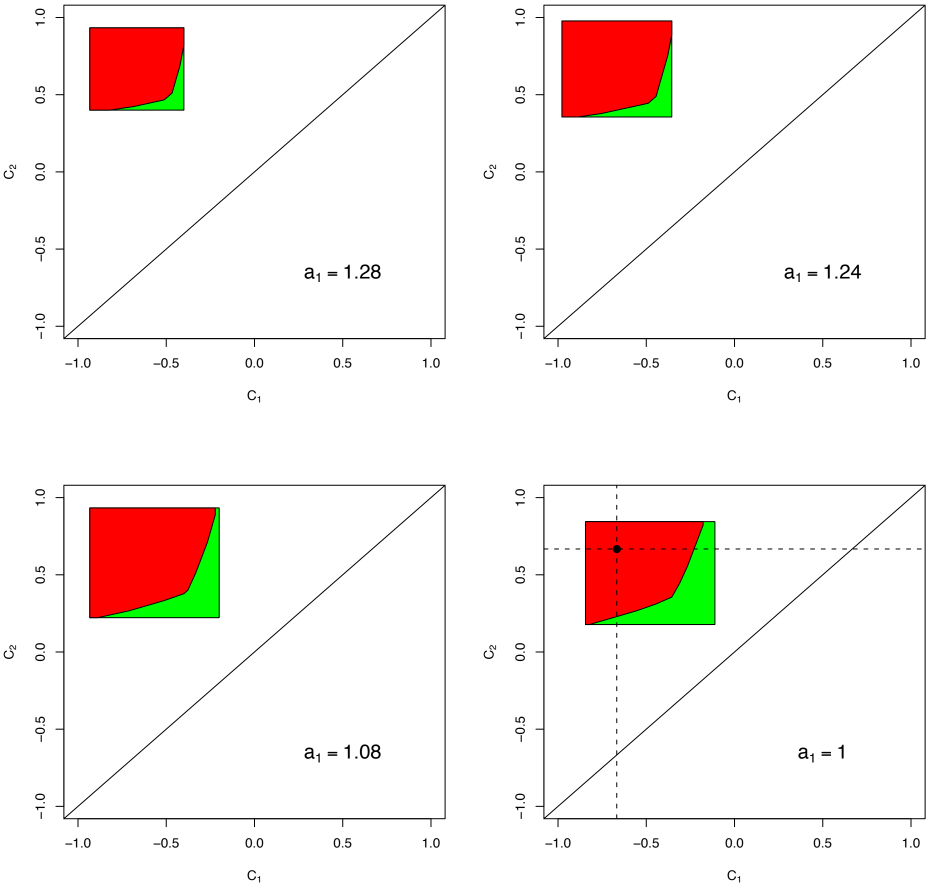


Figure 2.9: **Illustration B1.** Three class ordered probit model with unknown threshold parameters c_1 and c_2 and slope coefficient a_1. Cross-section of the identified set (red) and outer set (red and green) for c_1, c_2 and a_1 at selected values of a_1.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

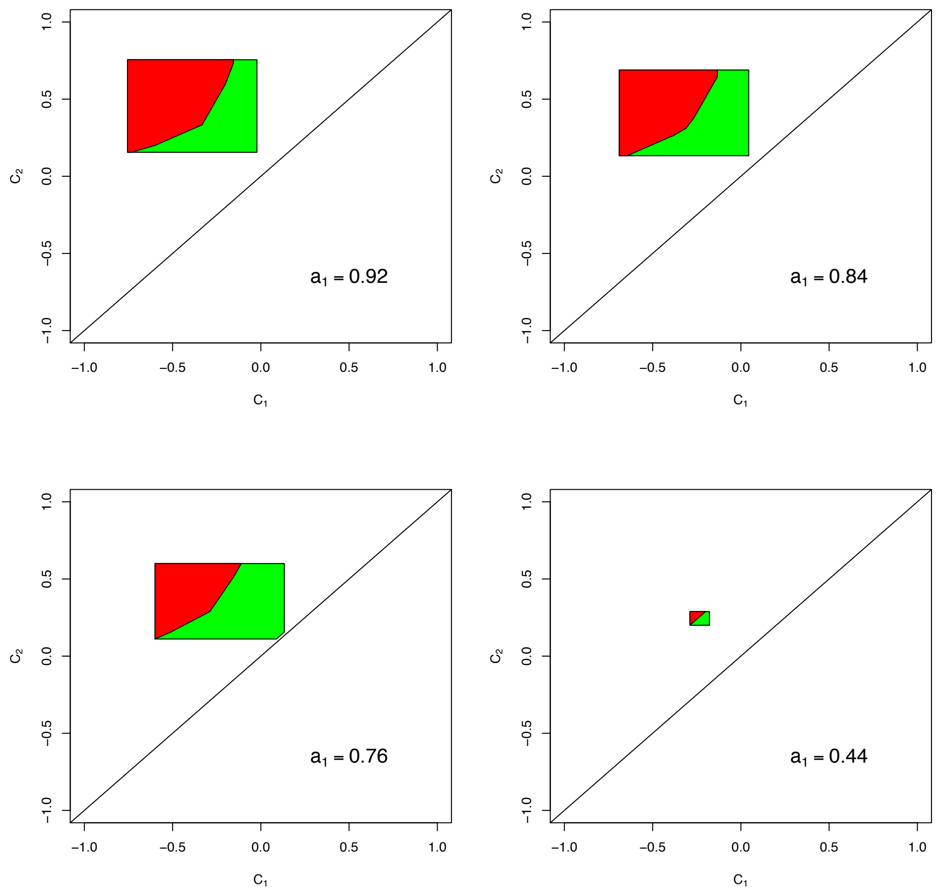


Figure 2.10: **Illustration B1.** Three class ordered probit model with unknown threshold parameters c_1 and c_2 and slope coefficient a_1. Cross-section of the identified set (red) and outer set (red and green) for c_1, c_2 and a_1 at selected values of a_1.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Chapter 3

Sharp identified sets for discrete variable IV models

Instrumental variable models for discrete outcomes are set, not point, identifying. The paper characterises identified sets of structural functions when endogenous variables are discrete. Identified sets are unions of large numbers of convex sets and may not be convex nor even connected. Each of the component sets is a projection of a convex set that resides in a much higher dimensional space onto the space in which a structural function resides. The paper develops a symbolic expression for this projection and gives a constructive demonstration that it is indeed the identified set. We provide a `Mathematica™` notebook which computes the set symbolically. We derive properties of the set, suggest how the set can be used in practical econometric analysis when outcomes and endogenous variables are discrete and propose a method for estimating identified sets under parametric or shape restrictions. We develop an expression for a set of structural functions for the case in which endogenous variables are continuous or mixed discrete-continuous and show that this set contains all structural functions in the identified set in the non-discrete case.

3.1 Introduction

This paper gives new results on the identifying power of single equation instrumental variable (SEIV) models in which both the outcome of interest and potentially endogenous explanatory variables are discrete. These models generally set rather than point identify structural

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

functions.[1] The paper derives the sharp identified set for the general case in which there is an M-valued outcome and there are endogenous variables with K points of support.

The discrete outcome, discrete endogenous variable case studied here arises frequently in applied econometrics practice. Examples of settings in which the results of the paper are useful include situations in which a binary or ordered probit, or a logit or a count data analysis or some semiparametric or nonparametric alternative would be considered and explanatory variables are endogenous. We study nonparametric models but, as we show, characterizations of identified sets for nonparametric models are very useful in constructing identified sets in parametric cases.

In the instrumental variable model studied here an M-valued outcome, Y, is determined by a structural function characterised by $M-1$ threshold functions of possibly endogenous variables X. Instrumental variables, Z, are excluded from these threshold functions. The instrumental variables and the stochastic term whose value relative to the threshold functions determines the value of Y are independently distributed. When endogenous variables have K points of support the structural function is characterised by $N = K(M-1)$ parameters: the values of the $M-1$ threshold functions at the K values of X. A conventional parametric model, for example an ordered probit model, places restrictions on these N objects.

The model studied here places no restrictions on the process generating values of the potentially endogenous variables X. It is in this sense that it is a *single equation* model.

By contrast the commonly employed control function approach to identification employs a more restrictive triangular model which places restrictions on the process generating the potentially endogenous variables.[2] That model generally fails to deliver point identification when endogenous variables are discrete so the SEIV model is a leading contender for application in practice.

The single equation approach taken in this paper has some other points to recommend it. For example, structural simultaneous equations models for discrete endogenous variables throw up coherency issues, first studied in Heckman (1978) and subsequently discussed in, for example, Lewbel (2007). These can be neglected in a single equation analysis. Economic models involving simultaneous determination of values of discrete outcomes can involve mul-

[1] See Chesher (2010).

[2] See for example Blundell and Powell (2003, 2004), Chesher (2003), Imbens and Newey (2009).

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

tiple equilibria. See for example Tamer (2003). Taking a single equation approach one is free to leave the equilibrium selection process unspecified.

After normalisation the structural function in the discrete-outcome, discrete-endogenous variable case is characterised by a point in the unit N-cube. The set that is identified by a single equation instrumental variable model is a subset of this space. We show that the identified set is the union of many convex sets each of which is an intersection of linear half-spaces. The faces of these component convex polytopes are arranged either parallel to or at $45°$ angles to faces of the N-cube. The identified set may not be convex or even connected.

The convex components of the identified sets are projections of high-dimensional sets onto the space in which the structural function resides. Direct computation of these sets is challenging. Calculation for small scale problems can be done using the method of Fourier-Motzkin elimination. However for M or K larger than 4 the computations are prohibitively time consuming because of the very large number of inequalities produced during the process of projection. Almost all of these are redundant, but determining which are redundant is computationally demanding. The key to solving this problem is to make use of the structure placed on the problem by the SEIV model.

We consider probability distributions for Y and X conditional on $Z = z$ for values of z in some set of instrumental values $\mathcal{Z}$. We develop a system of inequalities which must be satisfied by the N values that characterise a structural function for all structural functions that are elements of structures admitted by the SEIV model which generate these probability distributions. The identified set of structural functions must be a subset of the set defined by these inequalities. We show using a constructive proof that the set is precisely the identified set.

Calculation of the convex components of an identified set using the expressions we present here is very easy. The remaining, non-trivial, computational challenge is to deal with the very large number of convex components that arises when M or K is large. This problem disappears if sufficiently strong shape restrictions can be invoked. Parametric models are useful in providing these. An alternative is to employ shape constrained sieve approximations.

We show how recently developed results on set estimation and inference when sets are defined by intersection bounds can be used to operationalise the results given here.

Finally we extend the analysis to the case in which endogenous variables are continuous or

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

mixed discrete-continuous and derive a set of structural functions within which all functions in the identified set lie. We conjecture that this is the identified set. A constructive proof remains to be completed.

The restrictions of the SEIV model are now set out and then the results given here are set in the context of earlier work.

3.1.1 The single equation instrumental variable model

In the SEIV model a scalar discrete outcome, Y, is determined by a structural function h as follows.

$$Y = h(X, U) \tag{3.1}$$

Here U is a scalar unobservable continuously distributed random variable and X is a list of explanatory variables. These explanatory variables may be endogenous in the sense that U and X may not be independently distributed. The focus is on identification of the structural function h.

In practice there may be variables appearing in h that are restricted to be exogenous (distributed independently of U) and the results of the paper are easily extended to accommodate these but for simplicity we proceed with the structural function specified as in equation (3.1).

The structural function h is restricted to be monotone in U for all values of X. It is normalized weakly increasing in what follows and the marginal distribution of U is normalized uniform on the closed unit interval $[0, 1]$. The support of X is denoted by $\mathcal{X}$.

The discrete outcome Y has M fixed points of support and without loss of generality these are taken to be the integers $1, \ldots, M$. Since h varies monotonically with U there is the following threshold crossing representation of the structural function: for $m \in \{1, \ldots, M\}$:

$$h(x, u) = m \text{ if and only if } h_{m-1}(x) < u \leq h_m(x)$$

with $h_0(x) = 0$ and $h_M(x) = 1$ for all $x \in \mathcal{X}$.

In this set-up a standard parametric probit model for $Y \in \{1, 2\}$ would have threshold functions as follows:

$$h_0(x) = 0 \qquad h_1(x) = \Phi(\alpha_0 + \alpha_1 x) \qquad h_2(x) = 1$$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

where $\Phi(\cdot)$ is the standard normal distribution function. A standard logit model would have $h_1(x) = (1 + \exp(\alpha_0 + \alpha_1 x))^{-1}$.

In this paper, except in Section 3.4, we study the case in which X is discrete with a finite number, K, of points of support: $\mathcal{X} = \{x_1, \ldots, x_K\}$. The objects whose identification is considered are the values of the $M-1$ threshold functions at the K values taken by X.

If the model restricted X to be exogenous then it would identify the threshold functions at each point in the support of X because in that case $\Pr[Y \leq m|X = x] = h_m(x)$.

The SEIV model does not require X to be exogenous but admits instrumental variables, one or many, discrete or continuous, arranged in a vector Z which takes values in a set $\mathcal{Z}$. The instrumental variables Z and U are independently distributed and Z is excluded from the structural function.[3] The model set identifies the structural function.

3.1.2 Relation to earlier work

The SEIV model studied here is an example of the sort of nonseparable model studied in Chernozhukov and Hansen (2005), Blundell and Powell (2003, 2004), Chesher (2003) and Imbens and Newey (2009).

All but the first of these papers study complete models which specify triangular equation systems in which there are structural equations for endogenous explanatory variables as well as for the outcome of interest. When endogenous variables are continuous these models can point identify structural functions but when endogenous variables are discrete they may not. Dealing with the discrete endogenous variable case, Chesher (2005) introduces an additional restriction on the nature of the dependence amongst unobservables providing a set identifying triangular model with discrete endogenous variables. Jun, Pinkse and Xu (2009) provide some refinements. Point identification can be achieved under parametric restrictions such as those used in Heckman (1978).

Discreteness of endogenous variables is not a problem for SEIV models, indeed it brings simplifications - for example eliminating the "ill posed inverse problem" which arises when endogenous variables are continuous. This is shown clearly in Das (2005) where *additive error* nonparametric models with discrete endogenous variables and instrumental variable

[3] At no point in the development is Z required to be a random variable. It could for example be a variable whose values are set by an experimenter. The key requirement is that the conditional distribution of U given $Z = z$ be invariant with respect to changes in z within the set $\mathcal{Z}$.

 DOI: 10.2870/35896

restrictions are considered. Because of the additive error restrictions this construction is not well suited to modelling discrete outcomes which sit more comfortably in the nonseparable error setting studied here.

Chernozhukov and Hansen (2005) study a nonadditive-error SEIV model like that considered here, focussing on the case in which the outcome is *continuous*. The identification results of that paper are built around the following equality which, when Y is continuous, holds for all $\tau \in (0,1)$ and all $z \in \mathcal{Z}$.

$$\Pr[Y = h(X,\tau)|Z = z] = \tau \tag{3.2}$$

Additional (completeness) conditions are provided under which the model point identifies the structural function.

The condition (3.2) does not hold when Y is discrete. Instead, as shown in Chernozhukov and Hansen (2001), there are the following inequalities which hold for all $\tau \in (0,1)$ and $z \in Z$.

$$\Pr[Y < h(X,\tau)|Z = z] < \tau \leq \Pr[Y \leq h(X,\tau)|Z = z]$$

These imply that the inequalities:

$$\max_{z\in\mathcal{Z}} \Pr[Y < h(X,\tau)|Z = z] < \tau \leq \min_{z\in\mathcal{Z}} \Pr[Y \leq h(X,\tau)|Z = z] \tag{3.3}$$

hold for all $\tau \in (0,1)$ as shown in Chesher (2007, 2010). The result is that the SEIV model generally fails to point identify the structural function when the *outcome* Y is discrete. However the model can be informative about the structural function as long as $\mathcal{Z}$ is not a singleton.

To see this suppose that for some value m and two values in $\mathcal{Z}$, z_1 and z_2, $\Pr[Y \leq m|Z = z_1] \neq \Pr[Y \leq m|Z = z_2]$. The restrictions of the model imply that in this case $h_m(x)$ is *not constant* for variations in x in admissible structures which generate the probability distribution under consideration. This is so because if $h_m(x)$ were constant, equal say to h^*_m, then *for all* $z \in \mathcal{Z}$, $\Pr[Y \leq m|Z = z] = h^*_m$ so *any* variation in $\Pr[Y \leq m|Z = z]$ with z rules

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

out the possibility that $h_m(x)$ is constant for variations in x.[4] At least the set of structural functions identified by the SEIV model excludes structures with constant threshold functions if the outcome and the instruments are not independently distributed.

The set identifying power of the SEIV model when the outcome is discrete was first studied in Chesher (2007). Let $\mathcal{H}(\mathcal{Z})$ denote the identified set of structural functions associated with some probability distribution $F_{YX|Z}$ for Y and X given $Z = z \in \mathcal{Z}$.[5] Chesher (2007, 2010) develops a set, denoted here by $\mathcal{C}(\mathcal{Z})$, based on the inequalities (3.3). It is shown that, when Y is binary and X is continuous, $\mathcal{H}(\mathcal{Z}) = \mathcal{C}(\mathcal{Z})$ and $\mathcal{C}(\mathcal{Z})$ provides tight set identification. In other cases it is an outer set in the sense that it can be that $\mathcal{H}(\mathcal{Z}) \subset \mathcal{C}(\mathcal{Z})$.

Chesher (2009) studies the binary outcome case, proving $\mathcal{H}(\mathcal{Z}) = \mathcal{C}(\mathcal{Z})$ when endogenous variables are discrete, considering the impact of parametric restrictions and shape restrictions, and giving some results on estimation under shape restrictions employing results on inference using intersection bounds given in Chernozhukov, Lee and Rosen (2009).

In Chesher and Smolinski (2009) a refinement[6] to $\mathcal{C}(\mathcal{Z})$, denoted $\tilde{\mathcal{D}}(\mathcal{Z})$, is developed. This delivers the identified set when there is a single binary endogenous variable no matter how many points of support the outcome Y has. The results are used in an investigation of the nature of the reduction in extent of the identified set as the number of points of support of Y increases in an endogenous parametric ordered probit example.

This paper studies the general finite M-outcome, K-point of support discrete endogenous variable case and develops a further refinement[7] to $\mathcal{C}(\mathcal{Z})$, denoted $\mathcal{E}(\mathcal{Z})$ and shows that $\mathcal{E}(\mathcal{Z})$ is precisely the identified set, $\mathcal{H}(\mathcal{Z})$.

---

[4] There is the following.

$$\begin{aligned}
\Pr[Y \leq m|Z = z] &= \sum_k \Pr[U \leq h_m^*|X = x_k, Z = z] \Pr[X = x_k|Z = z] \\
&= \Pr[U \leq h_m^*|Z = z] \\
&= h_m^*
\end{aligned}$$

Since the model excludes Z from the structural function h and requires U and Z to be independent the only way in which Z can affect the distribution of Y is through its effect on X and then only if h is sensitive to variations in X.

[5] It would be clearer to give a distinctive symbol to the probability distribution under consideration, e.g. $F^0_{Y X|Z}$ and label the various sets accordingly thus: $\mathcal{H}^0(\mathcal{Z})$, $\mathcal{C}^0(\mathcal{Z})$ and so forth. We do not do this here because the notation quickly becomes cumbersome. However it is important to keep in mind that each of the sets under discussion is associated with a particular probability distribution.

[6] By a refinement we mean that $\tilde{\mathcal{D}}(\mathcal{Z}) \subseteq \mathcal{C}(\mathcal{Z})$ with the possibility that $\tilde{\mathcal{D}}(\mathcal{Z}) \subset \mathcal{C}(\mathcal{Z})$

[7] Here, by a refinement we mean that $\mathcal{E}(\mathcal{Z}) \subseteq \tilde{\mathcal{D}}(\mathcal{Z}) \subseteq \mathcal{C}(\mathcal{Z})$.

3.1.3 Plan of the paper

Section 3.2 defines the set identified by the SEIV model and reviews its characteristics.

Section 3.3 develops the new set, $\mathcal{E}(\mathcal{Z})$, shows that it contains the identified set, and then gives a constructive proof that $\mathcal{E}(\mathcal{Z})$ is the identified set. This is done by proposing an algorithm for construction of a proper distribution for the unobservable U and endogenous X conditional on values of instrumental variables which for any value of γ in $\mathcal{E}(\mathcal{Z})$ delivers the probabilities used to construct the set while respecting the restriction that U and Z be independently distributed.

Section 3.3.1 develops some properties of the identified set. Section 3.3.2 presents the algorithm for constructing the distribution of U and X given Z which is used to demonstrate that $\mathcal{E}(\mathcal{Z})$ is the identified set. Section 3.3.3 gives an alternative derivation of the set $\mathcal{E}(\mathcal{Z})$ which is useful in linking the results of this paper with earlier results.

Section 3.3.4 sets out properties of the set $\mathcal{E}(\mathcal{Z})$ when the outcome is binary. Section 3.3.5 shows how the set $\mathcal{E}(\mathcal{Z})$ is related to the set defined in Chesher (2010). Section 3.3.6 gives alternative expressions for the inequalities defining the new set which help clarify its relationship to the set defined by the inequalities (3.3).

Section 3.4 derives a set $\mathcal{E}^A(\mathcal{Z})$ which is shown to contain all structural functions in the identified set for cases in which X is continuous or mixed discrete-continuous and is equal to the set $\mathcal{E}(\mathcal{Z})$ when X is discrete.

Section 3.5 gives some illustrative calculations, describes a `Mathematica` notebook which does symbolic calculation of the convex components of the identified set and discusses estimation and inference.

Section 3.6 concludes.

3.2 The identified set

In this Section the set identified by the SEIV model is defined and notation is introduced.

We consider situations in which X, which may be a scalar or a vector, is discrete and takes values in the set $\mathcal{X} = \{x_k\}_{k=1}^{K}$. In this case the structural function h is characterized

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

by $N \equiv K(M-1)$ parameters as follows,

$$\gamma_{mk} \equiv h_m(x_k), \qquad m \in \{1,\ldots,M-1\}, \quad k \in \{1,\ldots,K\}$$

which are arranged in a vector γ, as follows.[8]

$$\gamma \equiv [\gamma_{11},\ldots,\gamma_{1K},\gamma_{21},\ldots,\gamma_{2K},\ldots,\gamma_{M-1,1},\ldots,\gamma_{M-1,K}]$$

Identification of the vector γ is studied in this paper. Each element of γ lies in the unit interval so each value of γ is a point in the unit N-cube. The identified set is a subset of the unit N-cube. There are the restrictions $\gamma_{lk} < \gamma_{mk}$ for $l < m$ and all k. Henceforth "for all k" means for $k \in \{1,\ldots,K\}$.

Consider a particular probability distribution for Y and X given $Z = z \in \mathcal{Z}$. The identified set of values of γ associated with this distribution contains all and only values of γ for which there exist admissible conditional probability distributions of U and X given $Z = z$ for all values of z in $\mathcal{Z}$ such that the resulting structures deliver the probability distribution under consideration. Notation for that probability distribution is now introduced.

For values $z \in \mathcal{Z}$, for $m \in \{1,\ldots,M\}$ and $k \in \{1,\ldots,K\}$, there are the following point probabilities:

$$\rho_{mk}(z) \equiv \Pr[Y = m \wedge X = x_k | Z = z]$$

and cumulative probabilities:

$$\bar{\rho}_{mk}(z) \equiv \Pr[Y \leq m \wedge X = x_k | Z = z]$$

and probabilities marginal with respect to Y:

$$\delta_k(z) \equiv \Pr[X = x_k | Z = z] = \bar{\rho}_{Mk}(z).$$

Data are informative about these probabilities.

[8] If Γ is a matrix with (m,k) element equal to γ_{mk} then $\gamma \equiv vec(\Gamma')$. Considering γ_r, the rth element of γ, there are the following relationships.

$$r = (m-1)K + k$$

$$k = r \text{ modulo } K \qquad m = (r-k)/K + 1$$

For all k define: $\gamma_{0k} = 0$, $\gamma_{Mk} = 1$, $\rho_{0k}(z) = \bar{\rho}_{0k}(z) = 0$. In what follows "for all m" means for $m \in \{0, \dots, M\}$.

Associated with a particular value of γ and each value $z \in \mathcal{Z}$, define a piecewise uniform conditional distribution for U and X given Z, such that for all m, k and k':

$$\bar{\eta}_{mkk'}(z) \equiv \Pr[U \leq \gamma_{mk} \wedge X = x_{k'} | Z = z]$$

and let $\bar{\eta}(z)$ denote the complete list of $(M+1)K^2$ such terms.[9]

A list of values of $(\gamma, \bar{\eta}(z))$ produced as z varies in $\mathcal{Z}$ characterizes a *structure* which is *admissible* if it satisfies the following *independence* and *properness* conditions.

[1]. Independence. For all $z \in \mathcal{Z}$ and for all m and k the following equalities hold.[10]

$$\sum_{k'=1}^{K} \bar{\eta}_{mkk'}(z) = \gamma_{mk} \tag{3.4}$$

[2]. Properness. For all $z \in \mathcal{Z}$ and for all j, k, l, m and k', $\bar{\eta}_{ljk'}(z) \leq \bar{\eta}_{mkk'}(z)$ if and only if $\gamma_{lj} \leq \gamma_{mk}$. For all $z \in \mathcal{Z}$ and for all k, k', $\bar{\eta}_{0kk'}(z) = 0$. For all $z \in \mathcal{Z}$ and for all k' $\sum_{k=1}^{K} \bar{\eta}_{Mkk'}(z) = 1$.

If in addition the following *observational equivalence* condition is satisfied then the structure generates the probability distribution under consideration.

[3]. Observational equivalence. For all $z \in \mathcal{Z}$ and for all m and k the following equalities hold.

$$\bar{\eta}_{mkk}(z) = \bar{\rho}_{mk}(z) \tag{3.5}$$

All and *only* structures that obey conditions [1], [2] and [3] are in the set of structures identified by the model for the probabilities considered. Let $\mathcal{S}(\mathcal{Z})$ denote that set.

[9]Between each pair of adjacent knots, γ_{mk}, each conditional density function for Y given X and Z is uniform. The construction is justified in Chesher (2009). The conditional density functions have a histogram-like appearance.

[10]The left hand side is $\Pr[U \leq \gamma_{mk} | Z = z]$ which the independence restriction requires to be free of z. The value γ_{mk} on the right hand side arises because of the uniform distribution normalisation of the marginal distribution of U. See Chesher (2010).

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

The identified set of structural *functions*, $\mathcal{H}(\mathcal{Z})$, is the set of values of γ for which there are values of $\bar{\eta}(z)$ for $z \in \mathcal{Z}$ such that the resulting structure is in the identified set, $\mathcal{S}(\mathcal{Z})$. The identified set for γ, $\mathcal{H}(\mathcal{Z})$, is the *projection* of $\mathcal{S}(\mathcal{Z})$ onto the unit N-cube within which all values of γ lie.

The geometry of these sets is considered in Chesher and Smolinski (2009). A brief account is given here. Because of the properness condition [2] the order in which the elements of γ lie is an important consideration.

There are

$$T \equiv (K(M-1))!/\left((M-1)!\right)^K \tag{3.6}$$

admissible arrangements of the elements of γ.[11]

In each arrangement, $t \in \{1, \ldots, T\}$, the set of admissible observationally equivalent structures defined by [1], [2] and [3], denoted by $\mathcal{S}_t(\mathcal{Z})$, is either empty or a convex polytope because it is an intersection of bounded linear half spaces. The identified set of structures is the union of the sets obtained under each admissible arrangement.

$$\mathcal{S}(\mathcal{Z}) = \bigcup_{t=1}^{T} \mathcal{S}_t(\mathcal{Z})$$

In each arrangement, t, the identified set of structural functions obtained by projecting away $\bar{\eta}(z)$ for $z \in \mathcal{Z}$, denoted $\mathcal{H}_t(\mathcal{Z})$, is also either empty or a convex polytope. The complete identified set of structural functions is the union of these convex sets. The result may not itself be convex, nor even connected.

$$\mathcal{H}(\mathcal{Z}) = \bigcup_{t=1}^{T} \mathcal{H}_t(\mathcal{Z})$$

Direct computation of an identified set of structural functions is difficult when M and K are at all large. A head on attack would consider each admissible arrangement in turn and

[11] Arrangements in which there is a pair of indices m and m' with $m > m'$ such that for some k, $\gamma_{mk} \leq \gamma_{m'k}$ are inadmissible. The formula for T arises as follows. There are $\binom{N}{M-1}$ ways of placing $\gamma_{11}, \gamma_{21}, \ldots, \gamma_{M-1,1}$ in the $N = (M-1)K$ places available and only one order in which those values can lie. There are then $\binom{N-(M-1)}{M-1}$ ways of placing $\gamma_{12}, \gamma_{22}, \ldots, \gamma_{M-1,2}$ in the remaining $N-(M-1)$ places. Continuing in this way it is clear that there are

$$\prod_{k=1}^{K} \binom{N-(k-1)(M-1)}{M-1}$$

admissible arrangements of γ which on simplification yields the formula (3.6) for T.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

use the method of Fourier-Motzkin elimination[12] to project away the $(M-1)K^2$ elements in $\overline{\eta}(z)$ for each $z \in \mathcal{Z}$ but this is computationally infeasible when M and K are large.

In the next Section we develop easy-to-compute sets which are shown to be precisely the identified set of structural functions, $\mathcal{H}(\mathcal{Z})$.

3.3 Sharp set identification of the structural function

Conditions [1] - [3] place restrictions on values of $(\gamma, \overline{\eta}(z))$ for z varying in $\mathcal{Z}$. They define the identified set of *structures*: $\mathcal{S}(\mathcal{Z})$.

In this Section we develop implications of these restrictions for admissible values of γ, that is for values of γ that lie in the identified set of *structural functions*: $\mathcal{H}(\mathcal{Z})$. The result is a list of inequalities that define a set denoted $\mathcal{E}(\mathcal{Z})$. We show that this is the identified set $\mathcal{H}(\mathcal{Z})$.

The ordering of the elements of γ is important. The set $\mathcal{E}(\mathcal{Z})$ is a union of convex sets, $\mathcal{E}_t(\mathcal{Z})$, one associated with each admissible arrangement, t, of γ.

$$\mathcal{E}(\mathcal{Z}) = \bigcup_{t=1}^{T} \mathcal{E}_t(\mathcal{Z})$$

Each set $\mathcal{E}_t(\mathcal{Z})$ is defined as an intersection of linear half spaces.

We proceed to develop a definition of a set $\mathcal{E}_t(\mathcal{Z})$ obtained *under a particular arrangement*, t, of the elements of γ. First it is necessary to develop notation and functions for dealing with arrangements.

Let $\gamma_{[n]}$ be the nth largest value in an arrangement. Recall there are $N \equiv (M-1)K$ elements in γ. We adopt the notation used in the literature on order statistics to denote the ordered values of γ:

$$\gamma_{[1]} \leq \gamma_{[2]} \cdots \leq \gamma_{[N-1]} \leq \gamma_{[N]}$$

and we define $\gamma_{[0]} \equiv 0$ and $\gamma_{[N+1]} \equiv 1$.[13]

[12] See Zeigler (2007).

[13] A more precise notation would carry an identifier of the arrangement t under consideration and when stating formal results we do employ such a notation, for example denoting the ordered elements of an arrangement t by $\gamma^t_{[1]}, \ldots, \gamma^t_{[N]}$. During the exposition, while it is clear that a particular arrangement is under consideration, we simplify notation and do not make dependence on the arrangement under consideration explicit.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Define functions $m(n)$ and $k(n)$ such that $\gamma_{m(n)k(n)} = \gamma_{[n]}$. Define $m(0) = 0$. With $M = 3$ and $K = 3$, for which $N = 6$ and

$$\gamma = [\gamma_{11}, \gamma_{12}, \gamma_{13}, \gamma_{21}, \gamma_{22}, \gamma_{23}]$$

and for the arrangement

$$[\gamma_{11}, \gamma_{12}, \gamma_{21}, \gamma_{13}, \gamma_{23}, \gamma_{22}] \tag{3.7}$$

the functions $m(\cdot)$ and $k(\cdot)$ are as shown below. We will work with this example throughout this Section.

n	$m(n)$	$k(n)$
1	1	1
2	1	2
3	2	1
4	1	3
5	2	3
6	2	2

Figure 1 shows a configuration of threshold functions that is consonant with this arrangement. In this case $M = 3$ so there are two threshold functions, $h_1(x)$ and $h_2(x)$.

Values of X are measured along the horizontal axis in Figure 1 and three points of support, x_1, x_2 and x_3 are marked. Values of threshold functions are measured along the vertical axis which is the unit interval $[0, 1]$. This axis also measures values of the unobservable variable U.

At any value of x, values of U falling on or below the lowest threshold function deliver the value 1 for Y, values of U falling between the two threshold functions or on the highest threshold function deliver the value 2 for Y and values of U falling above the highest threshold deliver the value 3 for Y. Notice that the upper threshold function is not monotone in x reflecting the inequality $\gamma_{23} < \gamma_{22}$.

Now define an *inverse* function $n(m, k)$ such that $\gamma_{mk} = \gamma_{[n(m,k)]}$ and note that $n(m(n), k(n)) = n$. For all k define $n(0, k) = 0$.

For the arrangement (3.7) considered above the function $n(\cdot, \cdot)$ delivers values as shown

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

below.

	$k=1$	$k=2$	$k=3$
$m=1$	1	2	4
$m=2$	3	6	5

The functions $m(\cdot)$, $k(\cdot)$ and $n(\cdot,\cdot)$ are specific to the particular arrangement under consideration and we could make this dependence explicit by writing $m_t(\cdot)$, $k_t(\cdot)$ and $n_t(\cdot,\cdot)$ but for the most part this is not done in order to avoid excessively complex notation.

We use an abbreviated notation as follows: $\bar{\rho}_{[n]}$ denotes $\bar{\rho}_{m(n)k(n)}$, thus:

$$\bar{\rho}_{[n]} \equiv \bar{\rho}_{m(n)k(n)} = \Pr[Y \leq m(n) \wedge X = x_{k(n)}|Z = z]$$

and $\bar{\eta}_{[n]k'}$ denotes $\bar{\eta}_{m(n)k(n)k'}$ thus.

$$\bar{\eta}_{[n]k'} \equiv \bar{\eta}_{m(n)k(n)k'} = \Pr[U \leq \gamma_{m(n)k(n)} \wedge X = x_{k'}|Z = z]$$

There are associated non-cumulative probabilities as follows.[14]

$$\rho_{[n]} = \Pr[Y = m(n) \wedge X = x_{k(n)}|Z = z]$$

$$\eta_{[n]k'} = \Pr[U \in (\gamma_{[n-1]}, \gamma_{[n]}] \wedge X = x_{k'}|Z = z]$$

It is important to understand that $\bar{\eta}_{[n]k'} = \sum_{j=1}^{n} \eta_{[j]k'}$ but in general $\bar{\rho}_{[n]} \neq \sum_{j=1}^{n} \rho_{[j]}$ rather:

$$\bar{\rho}_{[n]} = \sum_{j=1}^{m(n)} \rho_{jk(n)} = \sum_{j=1}^{m(n)} \rho_{[n(j,k(n))]}.$$

All these probabilities depend on the instrumental value z under consideration but this dependence is not made explicit in the notation for the moment. Define $\rho_{[0]} = \bar{\rho}_{[0]} = 0$, $\eta_{[n]k} = \bar{\eta}_{[n]k} = 0$.

It is helpful to extend the definitions to cover probability masses associated with the Mth

[14] Here too we could make dependence on the arrangement under consideration explicit in the notation, e.g. writing $\rho^t_{[n]}$ and $\eta^t_{[i]k}$, but do not do so until we come to formal statements of results.

point of support of Y so for all k define

$$n(M,k) = N + k$$

and

$$\begin{aligned} m(N+k) &= M \\ k(N+k) &= k \end{aligned}$$

which lead to

$$\rho_{[N+k]} = \Pr[Y = M \wedge X = x_k | Z = z]$$

with associated cumulative probabilities

$$\bar{\rho}_{[N+k]} = \Pr[Y \leq M \wedge X = x_k | Z = z] = \Pr[X = x_k | Z = z] = \delta_k.$$

Table 3.1 exhibits the probability masses $\eta_{[n]k}$ for a general case with an M-valued outcome and endogenous variables with K points of support and $N = K(M-1)$. All values are non-negative, values in column k sum to δ_k and the sum of all $K(N+1)$ probability masses in the table is 1. We will make extensive reference to tables like this in what follows.

We consider a particular value $z \in \mathcal{Z}$ and construct a set $\mathcal{E}_t(z)$, defining $\mathcal{E}_t(\mathcal{Z})$ as the intersection of such sets for z varying in $\mathcal{Z}$.

$$\mathcal{E}_t(\mathcal{Z}) \equiv \bigcap_{z \in \mathcal{Z}} \mathcal{E}_t(z)$$

To avoid cumbersome notation we do not make the dependence of probabilities on the chosen value z explicit in the notation for the moment.

Each set $\mathcal{E}_t(z)$ is obtained by considering restrictions that Conditions [1] - [3] place on the elements of γ when they are in arrangement t. The restrictions arise because for all values of γ (see the final column of the Table) that lie in the identified set there exist values of the probability masses $\eta_{[n]k}$ such that:

1. the sum of probability masses lying in rows 1 through n is equal to $\gamma_{[n]}$, equivalently,

	Value of x				Ordered
n	x_1	x_2	$\cdots$	x_K	values of γ
0	0	0	$\cdots$	0	$\gamma_{[0]}$
1	$\eta_{[1]1}$	$\eta_{[1]2}$	$\cdots$	$\eta_{[1]K}$	$\gamma_{[1]}$
2	$\eta_{[2]1}$	$\eta_{[2]2}$	$\cdots$	$\eta_{[2]K}$	$\gamma_{[2]}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
n	$\eta_{[n]1}$	$\eta_{[n]2}$	$\cdots$	$\eta_{[n]K}$	$\gamma_{[n]}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
N	$\eta_{[N]1}$	$\eta_{[N]2}$	$\cdots$	$\eta_{[N]K}$	$\gamma_{[N]}$
$N+1$	$\eta_{[N+1]1}$	$\eta_{[N+1]2}$	$\cdots$	$\eta_{[N+1]K}$	$\gamma_{[N+1]}$

Table 3.1: Piece-wise uniform joint distribution of U and X conditional on a value of Z arranged by ordered values of γ (rows) and points of support of the endogenous variable X (columns).

the sum of probability masses in row n is equal to $\gamma_{[n]} - \gamma_{[n-1]}$,

$$\sum_{i=1}^{n}\sum_{k=1}^{K}\eta_{[i]k} = \gamma_{[n]}$$

$$\sum_{k=1}^{K}\eta_{[i]k} = \gamma_{[n]} - \gamma_{[n-1]} \equiv \Delta\gamma_{[n]}$$

2. all probability masses are non-negative,

3. probability masses sum over appropriate blocks of cells in the table to deliver the observed probabilities $\rho_{[1]}, \ldots, \rho_{[n]}$.

Table 3.2 exhibits the blocks of cells over which probability masses must be aggregated for the arrangement shown in equation (3.7). Table 3.3 shows the values that must be achieved when summing within blocks if the observational equivalence condition is to be satisfied. For example in the arrangement considered there is $m(4) = 1$, $k(4) = 3$ and

$$\rho_{[4]} = \Pr[Y = 1 \wedge X = x_3|Z = z] = \sum_{i=1}^{4}\eta_{[i]3} = \Pr[U \leq \gamma_{[4]} \wedge X = x_3|Z = z]$$

must hold if the observational equivalence restriction is to be satisfied. Another example: $m(7) = 3$, $k(7) = 1$ and so observational equivalence requires that the following equalities

	Value of x			Differences of ordered	Ascending list
n	x_1	x_2	x_3	values of γ	of elements in γ
0					0
1	$\eta_{[1]1}$	$\eta_{[1]2}$	$\eta_{[1]3}$	$\gamma_{[1]} - \gamma_{[0]}$	γ_{11}
2	$\eta_{[2]1}$	$\eta_{[2]2}$	$\eta_{[2]3}$	$\gamma_{[2]} - \gamma_{[1]}$	γ_{12}
3	$\eta_{[3]1}$	$\eta_{[3]2}$	$\eta_{[3]3}$	$\gamma_{[3]} - \gamma_{[2]}$	γ_{21}
4	$\eta_{[4]1}$	$\eta_{[4]2}$	$\eta_{[4]3}$	$\gamma_{[4]} - \gamma_{[3]}$	γ_{13}
5	$\eta_{[5]1}$	$\eta_{[5]2}$	$\eta_{[5]3}$	$\gamma_{[5]} - \gamma_{[4]}$	γ_{23}
6	$\eta_{[6]1}$	$\eta_{[6]2}$	$\eta_{[6]3}$	$\gamma_{[6]} - \gamma_{[5]}$	γ_{22}
7	$\eta_{[7]1}$	$\eta_{[7]2}$	$\eta_{[7]3}$	$\gamma_{[7]} - \gamma_{[6]}$	1

Table 3.2: Conditional mass function values arranged by ordered values of γ and values of the conditioning variable X showing blocks of cells whose mass must be aggregated when considering the observational equivalence condition.

	Value of x			Differences of ordered	List of elements of γ
n	x_1	x_2	x_3	values of γ	in ascending order
0					0
1	$\rho_{[1]}$			$\gamma_{[1]} - \gamma_{[0]}$	γ_{11}
2		$\rho_{[2]}$		$\gamma_{[2]} - \gamma_{[1]}$	γ_{12}
3	$\rho_{[3]}$			$\gamma_{[3]} - \gamma_{[2]}$	γ_{21}
4			$\rho_{[4]}$	$\gamma_{[4]} - \gamma_{[3]}$	γ_{13}
5			$\rho_{[5]}$	$\gamma_{[5]} - \gamma_{[4]}$	γ_{23}
6		$\rho_{[6]}$		$\gamma_{[6]} - \gamma_{[5]}$	γ_{22}
7	$\rho_{[7]}$	$\rho_{[8]}$	$\rho_{[9]}$	$\gamma_{[7]} - \gamma_{[6]}$	1

Table 3.3: Sums of probability masses in blocks of cells must aggregate to the indicated probabilities if the observational equivalence condition is to be satisfied.

hold.

$$\rho_{[7]} = \Pr[Y = 3 \wedge X = x_1 | Z = z] = \sum_{i=4}^{7} \eta_{[i]1} = \Pr[\gamma_{[3]} < U \leq \gamma_{[7]} \wedge X = x_1 | Z = z].$$

In general there are M blocks in each column of the table. In each row exactly one block terminates. The block of cells in which the mass $\rho_{[n]}$ must lie is in the column of the table associated with $x_{k(n)}$ and in the rows that end at $n(m(n), k(n)) = n$ and start at $n(m(n)-1, k(n))+1$. So the observational equivalence restriction requires that the conditions:

$$\rho_{[n]} = \sum_{i=n(m(n)-1,k(n))+1}^{n} \eta_{[i]k(n)}$$

hold for $n = 1, \ldots, N + K$.

A particular value of γ in arrangement t can only support an allocation of probability

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

mass satisfying Conditions (1) - (3) if the elements $\gamma_{[1]}, \ldots, \gamma_{[N]}$ are spaced sufficiently far apart to permit the allocation of probability mass in the required amounts in the blocks of cells that arise in the arrangement. For example, in the arrangement (3.7), considering Table 3.3, $\gamma_{[1]}$ must be at least equal to $\rho_{[1]}$, $\gamma_{[2]}$ must be at least equal to $\rho_{[1]} + \rho_{[2]}$ and so forth.

There are additional restrictions. For example $\gamma_{[5]} - \gamma_{[3]}$ must be at least equal to $\rho_{[5]}$ and $\gamma_{[7]} - \gamma_{[4]} = 1 - \gamma_{[4]}$ must be at least $\rho_{[5]} + \rho_{[8]} + \rho_{[9]}$. We now develop a complete characterisation of these inequalities which determine the spacing between elements of γ under a particular arrangement such that the allocation of probability mass to blocks of cells that is required to deliver observational equivalence is feasible.

To proceed we introduce the idea of the *active indexes in a row.* The active indexes in row n are K distinct elements of the list $\{1, 2, \ldots, N + K\}$. These are the indexes, i, of probabilities $\rho_{[i]}$ to which cells in row n contribute. The active index for column k of row n is defined as follows.

$$a_{nk} \equiv \min \{i : (n \leq i \leq N + K) \wedge (k(i) = k)\}$$

The active index list for row n is defined thus: $a_n \equiv \{a_{n1}, \ldots, a_{nK}\}$. Clearly $a_{mk} \leq a_{nk}$ for all $m \leq n$ and k. For all k define $a_{0k} = 0$. Each active index list a_n has n as a member and it is always the smallest member.

For the arrangement given in (3.7) there are, on considering Table 3.3, the following active index lists.

n	a_n
0	$\{0, 0, 0\}$
1	$\{1, 2, 4\}$
2	$\{3, 2, 4\}$
3	$\{3, 6, 4\}$
4	$\{7, 6, 4\}$
5	$\{7, 6, 5\}$
6	$\{7, 6, 9\}$
7	$\{7, 8, 9\}$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

We now introduce the idea of *last-discharged indexes for a row.* The last discharged index for column k in row n is the index, i, of the probability $\rho_{[i]}$ falling in column k whose block of cells was most recently completed at row n. The last-discharged index for column k in row n is defined for all k and $n \in \{1, \ldots, N\}$ as follows.

$$d_{nk} \equiv \max \{i : (0 \leq i \leq n) \wedge (k(i) = k)\}$$

For all k define $d_{0k} \equiv 0$ and $d_{N+1,k} \equiv N + k$. Clearly $d_{mk} \leq d_{nk}$ for all $m \leq n$ and k.

The row n last-discharged index list is defined as $d_n \equiv \{d_{n1}, \ldots, d_{nK}\}$. Each list d_n has n as a member and, except in row $N + 1$, it is the largest member. For the arrangement (3.7) there are, on considering Table 3.3, the following last-discharged index lists.

n	d_n
0	$\{0, 0, 0\}$
1	$\{1, 0, 0\}$
2	$\{1, 2, 0\}$
3	$\{3, 2, 0\}$
4	$\{3, 2, 4\}$
5	$\{3, 2, 5\}$
6	$\{3, 6, 5\}$
7	$\{7, 8, 9\}$

For a pair of indexes, $(r, s) \in \{0, 1, \ldots, N + 1\}$ with $r < s$ there is a minimal probability mass required to fall between $\gamma_{[r]}$ and $\gamma_{[s]}$ if observational equivalence is to be achieved.

This minimal mass is calculated as follows. In a column, k, there is a probability mass equal to $\bar{\rho}_{[d_{sk}]}$ required to lie below $\gamma_{[s]}$ because d_{sk} is the discharged index associated with row s and column k. From this must be removed any probability mass associated with the active index in column k of row r, a_{rk}, and any probability mass associated with active indexes for rows prior to r in column k (and so discharged by row r). This mass is given by $\bar{\rho}_{[a_{rk}]}$. This can exceed $\bar{\rho}_{[d_{sk}]}$ so the minimal probability mass required to fall in the interval $(\gamma_{[r]}, \gamma_{[s]}]$ associated with $X = x_k$ is $\max(0, \bar{\rho}_{[d_{sk}]} - \bar{\rho}_{[a_{rk}]})$ and the total (across all values of X) minimal probability mass required to fall in the interval $(\gamma_{[r]}, \gamma_{[s]}]$ is $\sum_{k=1}^{K} \max(0, \bar{\rho}_{[d_{sk}]} - \bar{\rho}_{[a_{rk}]})$.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

By way of example consider the arrangement (3.7) used before and the cases considered earlier.

1. $\gamma_{[1]} - \gamma_{[0]}$. The last-discharged indexes in row 1 are $\{1, 0, 0\}$ and the active indexes in row 0 are $\{0, 0, 0\}$. Only column 1 delivers a positive value and, noting that $\rho_{[0]} = 0$ there is:

$$\gamma_{[1]} - \gamma_{[0]} = \gamma_{[1]} \geq \bar{\rho}_{[1]} = \rho_{[1]}.$$

2. $\gamma_{[2]} - \gamma_{[0]}$. The last-discharged indexes in row 2 are $\{1, 2, 0\}$ and the active indexes in row 0 are $\{0, 0, 0\}$. Columns 1 and 2 deliver a positive value and, there is:

$$\gamma_{[2]} - \gamma_{[0]} = \gamma_{[2]} \geq \bar{\rho}_{[1]} + \bar{\rho}_{[2]} = \rho_{[1]} + \rho_{[2]}.$$

3. $\gamma_{[5]} - \gamma_{[3]}$. The last-discharged indexes in row 5 are $\{3, 2, 5\}$ and the active indexes in row 3 are $\{3, 6, 4\}$. Notice that $d_{52} = 2 < a_{32} = 6$ so column 2 produces no positive contribution. Only column 3 delivers a positive value and there is:

$$\gamma_{[5]} - \gamma_{[3]} \geq \bar{\rho}_{[5]} - \bar{\rho}_{[4]} = \rho_{[5]}.$$

4. $\gamma_{[7]} - \gamma_{[4]}$. The last-discharged indexes in row 7 are $\{7, 8, 9\}$ and the active indexes in row 4 are $\{7, 6, 4\}$. Columns 2 and 3 produce positive values and there is:

$$\gamma_{[7]} - \gamma_{[4]} = 1 - \gamma_{[4]} \geq \left(\bar{\rho}_{[8]} - \bar{\rho}_{[6]}\right) + \left(\bar{\rho}_{[9]} - \bar{\rho}_{[4]}\right) = \rho_{[8]} + \rho_{[5]} + \rho_{[9]}.$$

From the argument so far it follows that for every pair of indexes

$$(r, s) \in \{0, 1, \dots, N+1\}$$

with $r < s$ the following inequality must hold if the value of γ in the arrangement under consideration is to allow the allocations of non-negative probability mass required to satisfy the observational equivalence restriction.

$$\gamma_{[s]} - \gamma_{[r]} \geq \sum_{k=1}^{K} \max(0, \bar{\rho}_{[d_{sk}]} - \bar{\rho}_{[a_{rk}]}) \tag{3.8}$$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

This system of $(N+1)(N+2)/2$ inequalities defines a set of values of γ denoted by $\mathcal{E}_t(z)$ associated with arrangement t and instrumental value z.

All values of γ in $\mathcal{H}_t(z)$ must satisfy these inequalities, so $\mathcal{H}_t(z) \subseteq \mathcal{E}_t(z)$.

We can now give a formal statement regarding the convex components of the identified set of structural functions. At this point we make explicit in the notation the dependence of objects on the arrangement under consideration, t, and on the instrumental value, z.

THEOREM 1: *Consider an arrangement t of*

$$\gamma \equiv [\gamma_{11}, \ldots, \gamma_{1K}, \gamma_{21}, \ldots, \gamma_{2K}, \ldots, \gamma_{M-1,1}, \ldots, \gamma_{M-1,K}]$$

with ith largest element $\gamma^t_{[n]}$ such that

$$0 \equiv \gamma^t_{[0]} \leq \gamma^t_{[1]} \leq \cdots \leq \gamma^t_{[N]} \leq \gamma^t_{[N+1]} \equiv 1$$

where $N \equiv K(M-1)$. The correspondence between elements of the ordered and unordered lists is given by arrangement-specific functions $m_t(\cdot)$, $k_t(\cdot)$ and an inverse function $n_t(\cdot,\cdot)$ which are such that for $n \in \{1, \ldots, N\}$, $k \in \{1, \ldots, K\}$ and $m \in \{1, \ldots, M-1\}$:

$$\begin{aligned} \gamma^t_{[n]} &= \gamma_{m_t(n)k_t(n)} \\ \gamma_{mk} &= \gamma^t_{[n_t(m,k)]}. \end{aligned}$$

For all arrangements define $m_t(0) \equiv 0$ and for all k: $m_t(N+k) \equiv M$, $k_t(N+k) \equiv k, n_t(M,k) \equiv N+k$, $n_t(0,k) \equiv 0$. For $n \in \{1, \ldots, N+K\}$ define:

$$\bar{\rho}^t_{[n]}(z) \equiv \Pr[Y \leq m_t(n) \wedge X = x_{k_t(n)} | Z = z]$$

with $\bar{\rho}^t_{[0]}(z) \equiv 0$. For all k and *$n \in \{1, \ldots, N\}$ define*

$$a^t_{nk}(z) \equiv \min\{i : (n \leq i \leq N+k) \wedge (k_t(i) = k)\}$$

$$d^t_{nk}(z) \equiv \max\{i : (0 \leq i \leq n) \wedge (k_t(i) = k)\}$$

and for all k define $d_{0k} \equiv 0$ and $d_{N+1,k} \equiv N+k$. Define a set of values of γ, $\mathcal{E}_t(z)$,

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

determined by the intersection of the following $(N+1)(N+2)/2$ *linear half spaces:*

$$\gamma^t_{[s]} - \gamma^t_{[r]} \geq \sum_{k=1}^{K} \max(0, \bar{\rho}^t_{[d_{sk}(z)]}(z) - \bar{\rho}^t_{[a_{rk}(z)]}(z)) \tag{3.9}$$

with $(r,s) \in \{0,1,\ldots,N+1\}$ *and* $s > r$.

Then:

1. *The set* $\mathcal{H}_t(z)$ *is a subset of* $\mathcal{E}_t(z)$.

2. *For every* $\gamma \in \mathcal{E}_t(z)$ *there exists a distribution of* U *and* X *given* $Z = z$ *which is piecewise uniform for variations in* U *that:*

   (a) *is proper,*

   (b) *satisfies the independence condition: for all* $i \in \{1,\ldots,N\}$

   $$\sum_{j=1}^{i}\sum_{k=1}^{K} \eta^t_{[j]k}(z) = \gamma^t_{[i]}$$

   (c) *delivers the probabilities* $\bar{\rho}^t_{[i]}(z)$ *for all* $i \in \{1,\ldots,N\}$ *and so satisfies the observational equivalence property.*

Result 1 of the Theorem has already been demonstrated to be true because we showed that all $\gamma \in H_t(z)$ satisfy the inequalities that define the polytope $\mathcal{E}_t(z)$. It remains to show how to construct the distribution referred to in Result 2. That is the subject of Section 3.3.2. First two corollaries are stated and proved and simple upper and lower bounds on the elements of γ are derived.

Corollary 1: *For all* t *and* z, $\mathcal{E}_t(z) = \mathcal{H}_t(z)$.

Proof: Result 1 of the Theorem states that $\mathcal{E}_t(z) \subseteq \mathcal{H}_t(z)$ and Result 2 implies that $\mathcal{H}_t(z) \subseteq \mathcal{E}_t(z)$, from which it follows that $\mathcal{E}_t(z) = \mathcal{H}_t(z)$.

Corollary 2: *The set of values of the structural function identified by the SEIV model is as follows.*

$$\mathcal{H}(\mathcal{Z}) = \bigcup_{t=1}^{T}\left(\bigcap_{z\in\mathcal{Z}} \mathcal{E}_t(z)\right)$$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Proof: This follows directly from Corollary 1 on noting the composition of the identified set, $\mathcal{H}(\mathcal{Z})$.

$$\mathcal{H}(\mathcal{Z}) = \bigcup_{t=1}^{T} \left(\bigcap_{z \in \mathcal{Z}} \mathcal{H}_t(z) \right)$$

For a particular arrangement many of the inequalities defining a set $\mathcal{E}_t(z)$ will be redundant. The inequality given by setting $s = N + 1$ and $r = 0$ is always redundant[15] so there are at most $(N+1)(N+2)/2 - 1$ inequalities defining the polytope $\mathcal{E}_t(z)$ and often far fewer. This is investigated further in Section 3.5.

The identified set is determined by a large number of elementary inequalities which either place upper or lower bounds on elements of γ or lower bounds on differences of pairs of elements of γ. The convex polytope within which identified values of γ lie in any particular arrangement is a facetted N-orthotope lying in the unit N-cube with all facets taken at angles of 45° to the faces of the unit N-cube.

3.3.1 Upper and lower bounds

The inequalities (3.9) deliver simple upper and lower bounds on elements of γ specific to an arrangement t and an instrumental value.

Suppressing dependence on the arrangement, t, and the instrumental value, z, and setting $r = 0$ in (3.8) and noting that $\gamma_{[0]} = 0$ and for all k, $a_{0k} = 0$ and $\bar{\rho}_{[0]} = 0$, there is for all $s \in \{1, \ldots, N+1\}$ the lower bound:

$$\gamma_{[s]} \geq \sum_{k=1}^{K} \bar{\rho}_{[d_{sk}]} \tag{3.10}$$

which can be expressed in terms of non-cumulative probabilities as follows.

$$\gamma_{[s]} \geq \sum_{i=1}^{s} \rho_{[i]}.$$

Setting $s = N + 1$ in (3.8) and noting that $\gamma_{[N+1]} = 1$ there is for all $r \in \{1, \ldots, N+1\}$:

$$1 - \gamma_{[r]} \geq \sum_{k=1}^{K} \left(\delta_k - \bar{\rho}_{[a_{rk}]} \right)$$

[15] In this case the inequality (3.8) is $1 \geq 1$.

and so the following upper bound.

$$\gamma_{[r]} \leq \sum_{k=1}^{K} \bar{\rho}_{[a_{rk}]} \tag{3.11}$$

These are the bounds given in Chesher (2007, 2010).

3.3.2 Construction of a joint distribution of U and X

We propose an algorithm for constructing a joint distribution for U and X given $Z = z$ for *any* value of γ that lies in a set $\mathcal{E}_t(z)$ constructed using a sequence of probabilities $\rho_{[1]}, \ldots, \rho_{[N]}$. We then prove Result 2 of Theorem 1 by showing that the distribution has the required properties, namely that it is proper, that it satisfies observational equivalence, delivering the probabilities $\rho_{[1]}, \ldots, \rho_{[N]}$ that determine $\mathcal{E}_t(z)$, and that it satisfies the independence restriction as expressed in (3.4).

While setting up notation and giving the details of the workings of the algorithm dependence of objects such as $\rho_{[n]}$, $\gamma_{]n]}$, $\eta_{[n]k}$, a_{nk} and d_{nk} on the arrangement under consideration and the instrumental value is suppressed in the notation.[16] In what follows sums from a to b, $\sum_{i=a}^{b}(\cdot)_i$, with $b < a$ are by convention equal to zero.

We have introduced the active index lists a_n and we now make use of *ordered* active index lists $a_n^o \equiv \{a_{n[1]}, \ldots, a_{n[K]}\}$ where:

$$\min\{a_{n1}, \ldots, a_{nK}\} \equiv a_{n[1]} < a_{n[2]} < \cdots < a_{n[K]} \equiv \max\{a_{n1}, \ldots, a_{nK}\}.$$

Note that for all $n \in \{1, \ldots, N+1\}$, $a_{n[1]} = n$. The ordered active index list for the

[16]Of course the IV restriction ensures that $\gamma_{[n]}$ does not vary with the instrumental value z.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

arrangement (3.7) is as follows.

n	a_n^o
0	$\{0,0,0\}$
1	$\{1,2,4\}$
2	$\{2,3,4\}$
3	$\{3,4,6\}$
4	$\{4,6,7\}$
5	$\{5,6,7\}$
6	$\{6,7,9\}$
7	$\{7,8,9\}$

For i passing through the sequence: $1, 2, \ldots, N+1$ and, at each value of i, for j passing through the ascending sequence: $1, 2, \ldots, K$ the algorithm produces values $\eta_{[i]k}$ calculated recursively as follows:

$$\eta(i,j) = \min(G(i,j), \max(0, R(i,j))). \tag{3.12}$$

with component objects defined as follows.

$$\eta(i,j) \equiv \eta_{[i]k(a_{i[j]})}$$

$$G(i,j) \equiv \Delta\gamma_{[i]} - \sum_{j'=1}^{j-1} \eta(i,j')$$

$$R(i,j) \equiv \rho_{[a_{i[j]}]} - \sum_{i'=n(m(a_{i[j]})-1,k(a_{i[j]}))+1}^{i-1} \eta_{[i']k(a_{i[j]})}$$

It will shortly be shown that for every value of γ that lies in the set $\mathcal{E}_t(z)$ this algorithm fills the cells of a table with probability masses which are (i) non-negative while (ii) delivering total masses within groups of cells that respect the observational equivalence restriction and (iii) allocating a total mass of exactly $\Delta\gamma_{[i]}$ in row i for each $i \in \{1, \ldots, N+1\}$.

In each row $G(i,1)$ has the value $\Delta\gamma_{[i]}$ and for $j \in \{2, \ldots, K+1\}$ there is the following recursion.

$$G(i,j) = G(i,j-1) - \eta(i,j-1) \tag{3.13}$$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Note that

$$G(i, K+1) = \Delta\gamma_{[i]} - \sum_{j=1}^{K} \eta(i, j) = \Delta\gamma_{[i]} - \sum_{k=1}^{K} \eta_{[i]k}$$

so when $G(i, K+1) = 0$ for all i the algorithm delivers probabilities that satisfy the independence restriction.

In each row i the first cell to be addressed is the one with the smallest active index in that row. This is the cell that completes in row i the block of cells in column $k(i)$ which must contain probability mass $\rho_{[i]}$ if the observational equivalence condition is to be satisfied. $R(i, 1)$ is the mass to be allocated to that cell to bring the total to $\rho_{[i]}$. The next cell to be addressed is the one in the column $k(a_{i[2]})$ corresponding to the next active index to be discharged. Up to an amount $R(i, 2)$ is allocated in this cell. This is the mass which, if allocated to that cell, would bring the mass in the block of cells in which the cell appears up to $\rho_{[a_{i[2]}]}$. The process proceeds with j incrementing until $j = K$ with probability mass being allocated to cells until all the mass $\Delta\gamma_{[i]}$ has been allocated after which (as will be shown) zero values appear in the cells of row i.

We now show that when γ lies in the set $\mathcal{E}_t(z)$ the algorithm delivers probability masses in each cell such that the *observational equivalence* condition is satisfied. Then we prove that the independence and properness conditions are satisfied.

If γ lies in the set $\mathcal{E}_t(z)$ then *by construction* there is sufficient probability mass available between every pair of values $\gamma_{[r]}$ and $\gamma_{[s]}$ to permit the allocation of probability masses $\rho_{[1]}, \ldots, \rho_{[N]}$ in their appropriate locations. The probability mass $\rho_{[i]}$ is equal to $\Pr[Y = m(i) \wedge X = x_{k(i)} | Z = z]$. The cells in which this mass must be allocated lie in the column associated with $x_{k(i)}$ and terminate in row i. They start in the row given by $n(m(i) - 1, k(i)) + 1$. The observational equivalence condition is therefore: for all $i \in \{1, \ldots, N\}$

$$\sum_{t=n(m(i)-1,k(i))+1}^{i} \eta_{[t]k(i)} = \rho_{[i]}.$$

The proposed algorithm fill blocks of cells in index order, $\rho_{[1]}$ first, $\rho_{[2]}$ second and so on. At each step of the process the algorithm allocates as much probability mass as possible to the blocks of cells associated with probabilities $\rho_{[i]}$ which have the lowest values of i accessible at

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

that point.[17] The algorithm delivers the required allocations of probability mass for values of γ that lie in $\mathcal{E}_t(z)$ because such values have elements that are sufficiently separated to permit the required allocation of probability masses.

It is shown in the Proposition below that *whether or not* the value of γ lies in $\mathcal{E}_t(z)$ the algorithm (i) allocates non-negative probability mass in every cell and (ii) never allocates more than an amount $\Delta\gamma_{[i]}$ in row i, for $i \in \{1, \dots, N\}$. When the value of γ lies in $\mathcal{E}_t(z)$, a total mass of 1 is allocated by the algorithm because the observational equivalence condition is satisfied. Since, as shown below, an allocation exceeding $\Delta\gamma_{[i]}$ cannot occur for any i and $\sum_{i=1}^{N+1} \Delta\gamma_{[i]} = 1$, when the value of γ lies in $\mathcal{E}_t(z)$ the algorithm must place a probability mass *exactly* equal to $\Delta\gamma_{[i]}$ in each row $i \in \{1, \dots, N\}$, thus satisfying the independence condition. The properness conditions is satisfied because (i) all probability masses allocated are non-negative and (ii) since the observational equivalence condition is satisfied a total mass of 1 is allocated.

Here is the Proposition setting out some properties of the algorithm. These obtain *whether or not* $\gamma \in \mathcal{E}_t(z)$.

Proposition 1:

1. *For all i and j:*

   (a) $\eta(i, j) \leq G(i, j)$,

   (b) $G(i, j) \geq 0$,

   (c) $\eta(i, j) \geq 0$,

   (d) $G(i, j)$ *is a non-increasing function of* j.

   (e) *If for some j, $G(i, j) > \eta(i, j)$ then for all $j' \leq j$, $G(i, j') > \eta(i, j')$ and $\eta(i, j') = \max(0, R(i, j'))$.*

2. *The algorithm allocates a probability mass of at most $\Delta\gamma_{[i]}$ in row i, that is:*

$$\sum_{k=1}^{K} \eta_{[i]k} \leq \Delta\gamma_{[i]}.$$

[17] This occurs because the algorithm uses *ordered* active indexes.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

The proof is in Annex 1.

3.3.3 An alternative derivation of the set $\mathcal{E}_t(z)$

In order to relate the inequalities that define the set $\mathcal{E}_t(z)$ to the inequalities given in Chesher (2010) and Chesher and Smolinski (2009) it is useful to give an alternative derivation and expression for the set $\mathcal{E}_t(z)$.

Associated with the lists of active and last-discharged indexes there are arrays of cumulative probabilities which are useful in the subsequent analysis. They also provide an alternative characterisation of the set $\mathcal{E}_t(z)$.

Consider the ith largest element $\gamma_{[i]}$. If this lies in the identified set then for each $i \in \{1, \ldots, N\}$ there exist non-negative values of the *cumulative* probabilities $\bar{\eta}_{[i]k}$ which sum to $\gamma_{[i]}$ across $k \in \{1, \ldots, K\}$. The set $\mathcal{E}_t(z)$ is derived by finding lower and upper bounds for each term $\bar{\eta}_{[i]k}$ in the sum, producing bounds on differences of elements of γ by combining the bounds.

Each cumulative probability $\bar{\eta}_{[i]k}$ is bounded *below* by the *maximum* of the terms $\bar{\rho}_{[j]}$ that appear in rows 1 through i of the column associated with x_k. That bound is $\lambda_{ik} \equiv \bar{\rho}_{[d_{ik}]}$ where d_{ik} is the last-discharged index in column k of row i.

Each term $\bar{\eta}_{[i]k}$ is bounded *above* by the *minimum* of the terms $\bar{\rho}_{[j]}$ that appear in rows i through $N+1$ of column k. That bound is $\pi_{ik} \equiv \bar{\rho}_{[a_{ik}]}$ where a_{ik} is the active index in column k of row i.

Combining results there are the following bounds for all i and k:

$$\lambda_{ik} \leq \bar{\eta}_{[i]k} \leq \pi_{ik} \tag{3.14}$$

and on summing and noting that for γ in the identified set the independence condition holds so that $\gamma_{[i]} = \sum_{k=1}^{K} \bar{\eta}_{[i]k}$ there are the following lower and upper bounds.

$$\lambda_i \equiv \sum_{k=1}^{K} \lambda_{ik} \leq \gamma_{[i]} \leq \sum_{k=1}^{K} \pi_{ik} \equiv \pi_i \tag{3.15}$$

Making explicit dependence on the arrangement under consideration, t, and the instru-

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

mental value, z, defining

$$\lambda_{ik}^t(z) \equiv \bar{\rho}^t_{[d^t_{ik}(z)]}(z) \qquad \pi_{ik}^t(z) \equiv \bar{\rho}_{[a^t_{ik}(z)]}(z)$$

$$\lambda_i^t(z) \equiv \sum_{k=1}^{K} \lambda_{ik}^t(z) \qquad \pi_i^t(z) \equiv \sum_{k=1}^{K} \pi_{ik}^t(z)$$

and intersecting the bounds (3.20) across $z \in \mathcal{Z}$ gives the following inequalities which hold for each arrangement t and for all $i \in \{1, \dots, N\}$.

$$\max_{z \in \mathcal{Z}} \left(\lambda_i^t(z)\right) \leq \gamma_{[i]} \leq \min_{z \in \mathcal{Z}} \left(\pi_i^t(z)\right) \tag{3.16}$$

The inequalities (3.14) can also be used to place bounds on *differences*, $\gamma_{[s]} - \gamma_{[r]}$, as follows. For all s and r in $\{0, \dots, N+1\}$ and for all k there are bounds on $\bar{\eta}_{[s]k}$ and on $-\bar{\eta}_{[r]k}$ as follows:

$$\lambda_{sk} \leq \bar{\eta}_{[s]k} \leq \pi_{sk}$$

$$-\pi_{rk} \leq -\bar{\eta}_{[r]k} \leq -\lambda_{rk}$$

and on adding there are the following bounds.

$$\lambda_{sk} - \pi_{rk} \leq \bar{\eta}_{[s]k} - \bar{\eta}_{[r]k} \leq \pi_{sk} - \lambda_{rk} \tag{3.17}$$

Summing across k there are the following inequalities.

$$\lambda_s - \pi_r = \sum_{k=1}^{K} (\lambda_{sk} - \pi_{rk}) \leq \gamma_{[s]} - \gamma_{[r]} \leq \sum_{k=1}^{K} (\pi_{sk} - \lambda_{rk}) = \pi_s - \lambda_r \tag{3.18}$$

This is nothing more than a direct implication of (3.15) but the lower bound here *can be improved upon* by exploiting the properness condition [2]. Thus, consider values s and r such that $s > r$. If γ is in the identified set then for all $s > r$ the inequality $\bar{\eta}_{[s]k} - \bar{\eta}_{[r]k} \geq 0$ holds. The lower bound in (3.17) can therefore be tightened as follows.

$$\max(0, \lambda_{sk} - \pi_{rk}) \leq \bar{\eta}_{[s]k} - \bar{\eta}_{[r]k} \leq \pi_{sk} - \lambda_{rk} \tag{3.19}$$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Summing across k gives the following bounds which hold for all $s > r \in \{0, \ldots, N+1\}$.

$$\sum_{k=1}^{K} \max(0, \lambda_{sk} - \pi_{rk}) \leq \gamma_{[s]} - \gamma_{[r]} \leq \pi_s - \lambda_r \tag{3.20}$$

The set defined by these bounds is precisely the set $\mathcal{E}_t(z)$.

Making explicit the dependence of the terms in these bounds on the arrangement under consideration, t, and the instrumental value z, and intersecting the bounds (3.20) across $z \in \mathcal{Z}$ gives the following inequalities which hold for each arrangement t and for all $N+1 \geq s \geq r \geq 0$.

$$\begin{aligned} \underline{\phi}^t_{sr}(\mathcal{Z}) \equiv \max_{z \in \mathcal{Z}} \left(\sum_{k=1}^{K} \max(0, \lambda^t_{sk}(z) - \pi^t_{rk}(z)) \right) & \\ \leq \gamma^t_{[s]} - \gamma^t_{[r]} \leq & \\ \min_{z \in \mathcal{Z}} \left(\pi^t_s(z) - \lambda^t_r(z) \right) \equiv \bar{\phi}^t_{sr}(\mathcal{Z}) & \end{aligned} \tag{3.21}$$

These bounds define $\mathcal{E}_t(\mathcal{Z})$, the component of the identified set in which γ is in arrangement t. The union of these sets, $\bigcup_{t=1}^{T} \mathcal{E}_t(\mathcal{Z})$, is the set $\mathcal{E}(\mathcal{Z})$, previously defined, which is equal to the identified set $\mathcal{H}(\mathcal{Z})$.

3.3.4 Binary outcomes

When Y is binary there is just one threshold function and the parameters of interest are $\gamma_{11}, \gamma_{12}, \ldots, \gamma_{1K}$. We now show that in this case the lower bound in (**??**) is zero when $s > 0$ so these bounds place no restrictions on γ additional to those defined by (3.16).

Without loss of generality we consider an arrangement t in which the elements of γ are arranged in the order of the index k. The situation for a case in which $K = 6$ is as pictured in Table 3.4. Notice that with the given arrangement of γ for every value of n, $k(n) = n$, so the values $\bar{\rho}_{[n]}$ lie on the diagonal in Table 3.4. Because Y is binary there is only one such entry in each column.

We now show that for all indices $s > r > 0$ the terms $\lambda_{sk} - \pi_{rk}$ are zero or negative for all k from which it follows that the lower bound in (**??**) is zero.

Consider some value k and the difference $\lambda_{sk} - \pi_{rk}$ with $s \geq r$. Referring to Table 3.4 it

	Value of X						Ordered		
n	x_1	x_2	x_3	x_4	x_5	x_6	values of γ	γ	$k(n)$
0	0	0	0	0	0	0	$\gamma_{[0]}$	0	
1	$\bar{\rho}_{[1]}$	$\eta_{[1]2}$	$\eta_{[1]3}$	$\eta_{[1]4}$	$\eta_{[1]5}$	$\eta_{[1]6}$	$\gamma_{[1]}$	γ_{11}	1
2	$\eta_{[2]1}$	$\bar{\rho}_{[2]}$	$\eta_{[2]3}$	$\eta_{[2]4}$	$\eta_{[2]5}$	$\eta_{[2]6}$	$\gamma_{[2]}$	γ_{12}	2
3	$\eta_{[3]1}$	$\eta_{[3]2}$	$\bar{\rho}_{[3]}$	$\eta_{[3]4}$	$\eta_{[3]5}$	$\eta_{[3]6}$	$\gamma_{[3]}$	γ_{13}	3
4	$\eta_{[4]1}$	$\eta_{[4]2}$	$\eta_{[4]3}$	$\bar{\rho}_{[4]}$	$\eta_{[4]5}$	$\eta_{[4]6}$	$\gamma_{[4]}$	γ_{14}	4
5	$\eta_{[5]1}$	$\eta_{[5]2}$	$\eta_{[5]3}$	$\eta_{[5]4}$	$\bar{\rho}_{[5]}$	$\eta_{[5]6}$	$\gamma_{[5]}$	γ_{15}	5
6	$\eta_{[6]1}$	$\eta_{[6]2}$	$\eta_{[6]3}$	$\eta_{[6]4}$	$\eta_{[6]5}$	$\bar{\rho}_{[6]}$	$\gamma_{[6]}$	γ_{16}	6
7	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	$\gamma_{[7]}$	1	

Table 3.4: Conditional distribution-mass function values for a binary outcome example with observational equivalence restrictions imposed.

can be seen that values taken by λ_{sk} and π_{rk} are as follows.

$$\lambda_{sk} = \begin{cases} 0 & , \quad s < k \\ \bar{\rho}_{[k]} & , \quad s \geq k \end{cases} \qquad \pi_{rk} = \begin{cases} \bar{\rho}_{[k]} & , \quad r \leq k \\ \delta_k & , \quad r > k \end{cases}$$

The resulting values of $\lambda_{sk} - \pi_{rk}$ are therefore as shown below.

Values of $\lambda_{sk} - \pi_{rk}$

	$s < k$	$s = k$	$s > k$
$r < k$	$-\bar{\rho}_{[k]}$	$-\bar{\rho}_{[k]}$	0
$r = k$	$*$	0	0
$r > k$	$*$	$*$	$\bar{\rho}_{[k]} - \delta_k$

All the values are zero or negative and the result is that the lower bounds $\underline{\phi}^t_{sr}(\mathcal{Z})$ are zero. Therefore in the binary Y case the restrictions imposed by the bounds (**??**) for $r \neq 0$ have no force. It is shown in the next Section that the bounds obtained from (**??**) setting $r = 0$, equivalently the bounds (3.16), are identical to the bounds given in Chesher (2009, 2010) which are shown in those papers to define the identified set $\mathcal{H}(\mathcal{Z})$.

3.3.5 Relationship to earlier results

It is shown in Chesher (2010) that all structural functions h that lie in the set identified by a SEIV model given a particular probability distribution $F_{YX|Z}$ with $Z = z \in \mathcal{Z}$ satisfy the

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

following inequalities for all $\tau \in (0,1)$.

$$\max_{z\in\mathcal{Z}} \Pr[Y < h(X,\tau)|Z=z] < \tau \leq \min_{z\in\mathcal{Z}} \Pr[Y \leq h(X,\tau)|Z=z] \tag{3.22}$$

Here probabilities are calculated using the distribution $F_{YX|Z}$.

The inequalities generated by (3.22) as τ varies over $(0,1)$ define a set of structural functions referred to as $\mathcal{C}(\mathcal{Z})$ in Chesher and Smolinski (2009). When X is discrete and characterized by a vector γ as in the previous discussion the set $\mathcal{C}(\mathcal{Z})$ is a union of convex sets, $\mathcal{C}_t(\mathcal{Z})$, one associated with each arrangement, t, of γ.

$$\mathcal{C}(\mathcal{Z}) = \bigcup_{t=1}^{T} \mathcal{C}_t(\mathcal{Z})$$

Each set $\mathcal{C}_t(\mathcal{Z})$ is an intersection of sets obtained as z varies within $\mathcal{Z}$.

$$\mathcal{C}_t(\mathcal{Z}) = \bigcup_{z\in\mathcal{Z}} \mathcal{C}_t(z)$$

We now show that the bounds (3.22) are identical to those generated by (3.16) as n varies over $\{1,\ldots,N\}$. Chesher and Smolinski (2009) show that in the discrete endogenous variable case considered here the bounds (3.22) hold for all $\tau \in (0,1)$ if and only if the following inequalities hold for all $l \in \{1,\ldots,M-1\}$ and $s \in \{1,\ldots,K\}$.

$$\max_{z\in\mathcal{Z}} \sum_{k=1}^{K}\sum_{m=1}^{M-1} \rho_{mk}(z)1(\gamma_{mk} \leq \gamma_{ls}) \leq \gamma_{ls} \leq \min_{z\in\mathcal{Z}} \sum_{k=1}^{K}\sum_{m=1}^{M} \rho_{mk}(z)1(\gamma_{m-1,k} < \gamma_{ls})$$

Consider a particular arrangement of γ and its nth largest element, $\gamma_{[n]}$. Substituting $\gamma_{[n]}$ for γ_{ls} above gives the following.

$$\max_{z\in\mathcal{Z}} \sum_{k=1}^{K}\sum_{m=1}^{M-1} \rho_{mk}(z)1(\gamma_{mk} \leq \gamma_{[n]}) \leq \gamma_{[n]} \leq \min_{z\in\mathcal{Z}} \sum_{k=1}^{K}\sum_{m=1}^{M} \rho_{mk}(z)1(\gamma_{m-1,k} < \gamma_{[n]}) \tag{3.23}$$

Comparing this with (3.16) it can be concluded that the bounds are identical because both

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

of the following equations are satisfied:

$$\bar{\rho}_{[d_{nk}]}(z) \equiv \lambda_{nk}(z) = \sum_{m=1}^{M-1} \rho_{mk}(z)1(\gamma_{mk} \leq \gamma_{[n]}) \tag{3.24}$$

$$\bar{\rho}_{[a_{nk}]}(z) \equiv \pi_{nk}(z) = \sum_{m=1}^{M} \rho_{mk}(z)1(\gamma_{m-1,k} < \gamma_{[n]}) \tag{3.25}$$

and on the right hand side of (3.24) and (3.25) are the expressions summed over k to produce the bounds in (3.23).

3.3.6 Alternative expressions for the bounds

The objects $\lambda_{nk}(z)$ and $\pi_{nk}(z)$ can be expressed in terms of probabilities involving the structural function for $n \in \{1, \ldots, N\}$ as follows.

$$\lambda_{nk}(z) = \begin{cases} \Pr[Y < h(x_k, \gamma_{[n+1]}) \wedge X = x_k | Z = z] & , \quad n \in \{1, \ldots, N\} \\ \Pr[Y \leq h(x_k, \gamma_{[n]}) \wedge X = x_k | Z = z] & , \quad n = N+1 \end{cases} \tag{3.26}$$

$$\pi_{nk}(z) = \Pr[Y \leq h(x_k, \gamma_{[n]}) \wedge X = x_k | Z = z] \quad , \quad n \in \{1, \ldots, N+1\} \tag{3.27}$$

With these expressions in hand the bounds (3.16) can be written:

$$\lambda_n(\mathcal{Z}) \equiv \max_{z \in \mathcal{Z}} \left(\Pr[Y < h(X, \gamma_{[n+1]}) | Z = z] \right) \leq \gamma_{[n]} \leq \min_{z \in \mathcal{Z}} \left(\Pr[Y \leq h(X, \gamma_{[n]}) | Z = z] \right) \equiv \pi_n(\mathcal{Z}) \tag{3.28}$$

and the bounds (**??**) take the following form.

$$\underline{\phi}^t_{sr}(\mathcal{Z}) \leq \gamma_{[s]} - \gamma_{[r]} \leq \bar{\phi}^t_{sr}(\mathcal{Z}) \tag{3.29}$$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

$$\underline{\phi}^t_{sr}(\mathcal{Z}) = \max_{z\in\mathcal{Z}}(\sum_k \max\{0, \Pr[Y < h(x_k, \gamma_{[s+1]}) \wedge X = x_k)|Z = z]$$

$$- \Pr[Y \leq h(x_k, \gamma_{[r]}) \wedge X = x_k|Z = z]\}) \quad (3.30)$$

$$\bar{\phi}^t_{sr}(\mathcal{Z}) = \min_{z\in\mathcal{Z}} \left(\pi^t_s(z) - \lambda^t_r(z)\right) \quad (3.31)$$

These expressions are elucidated using the example considered earlier. Table 3.5 shows the values of λ_{nk} and π_{nk} in the arrangement used in the example. Dependence on the value z is no longer made explicit in the notation.

Table 3.6 shows the value of the structural function $h(x, u)$ for all the combinations of x and u that arise in (3.26) and (3.27) in this example. For example the entry for $n = 3$ and $k = 2$ under the heading π_{n2} is $h(x_2, \gamma_{[3]}) = h(x_2, \gamma_{21}) = 2$. The entries in this Table are easily verified by referring to Figure 1.

Consider for example λ_{42}. From (3.26) we have

$$\lambda_{42} = \Pr[Y \leq h(x_2, \gamma_{[5]}) \wedge X = x_2|Z = z]$$

and since $\gamma_{[5]} = \gamma_{23}$ there is, from Table 3.6, $h(x_2, \gamma_{23}) = 2$. Accordingly

$$\lambda_{42} = \Pr[Y < 1 \wedge X = x_2|Z = z]$$

which is equal to $\bar{\rho}_{12}$ as shown in Table 3.5 in the entry for $n = 4$ and $k = 2$.

Consider for example π_{33}. From (3.27) we have

$$\pi_{33} = \Pr[Y \leq h(x_3, \gamma_{[3]}) \wedge X = x_3|Z = z]$$

and since $\gamma_{[3]} = \gamma_{21}$ there is, from Table 3.6, $h(x_3, \gamma_{21}) = 1$. Accordingly

$$\pi_{33} = \Pr[Y \leq 1 \wedge X = x_3|Z = z]$$

which is equal to $\bar{\rho}_{13}$ as shown in Table 3.5 in the entry for $n = 3$ and $k = 3$.

3.4 Bounding inequalities for continuous endogenous variables

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

			$k=1$		$k=2$		$k=3$	
n	$\gamma_{(n)}$	$\gamma_{(n+1)}$	λ_{n1}	π_{n1}	λ_{n2}	π_{n2}	λ_{n3}	π_{n3}
0	0	γ_{11}	0	$\bar{\rho}_{11}$	0	$\bar{\rho}_{12}$	0	$\bar{\rho}_{13}$
1	γ_{11}	γ_{12}	$\bar{\rho}_{11}$	$\bar{\rho}_{11}$	0	$\bar{\rho}_{12}$	0	$\bar{\rho}_{13}$
2	γ_{12}	γ_{21}	$\bar{\rho}_{11}$	$\bar{\rho}_{21}$	$\bar{\rho}_{12}$	$\bar{\rho}_{12}$	0	$\bar{\rho}_{13}$
3	γ_{21}	γ_{13}	$\bar{\rho}_{21}$	$\bar{\rho}_{21}$	$\bar{\rho}_{12}$	$\bar{\rho}_{22}$	0	$\bar{\rho}_{13}$
4	γ_{13}	γ_{23}	$\bar{\rho}_{21}$	$\bar{\rho}_{31}$	$\bar{\rho}_{12}$	$\bar{\rho}_{22}$	$\bar{\rho}_{13}$	$\bar{\rho}_{13}$
5	γ_{23}	γ_{22}	$\bar{\rho}_{21}$	$\bar{\rho}_{31}$	$\bar{\rho}_{12}$	$\bar{\rho}_{22}$	$\bar{\rho}_{23}$	$\bar{\rho}_{23}$
6	γ_{22}	1	$\bar{\rho}_{21}$	$\bar{\rho}_{31}$	$\bar{\rho}_{22}$	$\bar{\rho}_{22}$	$\bar{\rho}_{23}$	$\bar{\rho}_{33}$
7	1		$\bar{\rho}_{31}$	$\bar{\rho}_{31}$	$\bar{\rho}_{32}$	$\bar{\rho}_{32}$	$\bar{\rho}_{33}$	$\bar{\rho}_{33}$

Table 3.5: Values of λ_{nk} and π_{nk} in the arrangement used in the example.

			$k=1$		$k=2$		$k=3$	
n	$\gamma_{(n)}$	$\gamma_{(n+1)}$	A: λ_{n1}	B: π_{n1}	A: λ_{n2}	B: π_{n2}	A: λ_{n3}	B: π_{n3}
0	0	γ_{11}	1	1	1	1	1	1
1	γ_{11}	γ_{12}	2	1	1	1	1	1
2	γ_{12}	γ_{21}	2	2	2	1	1	1
3	γ_{21}	γ_{13}	3	2	2	2	1	1
4	γ_{13}	γ_{23}	3	3	2	2	2	1
5	γ_{23}	γ_{22}	3	3	2	2	3	2
6	γ_{22}	1	3	3	3	2	3	3
7	1		3	3	3	3	3	3

Table 3.6: For the arrangement used in the example these are the values of A: $h(x_k, \gamma_{[n+1]})$ appearing in the definition of λ_{nk} and of B: $h(x_k, \gamma_{[n]})$ appearing in the definition of π_{nk}.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute
DOI: 10.2870/35896

So far the endogenous explanatory variable X has been required to be discrete. The argument used to develop the inequalities defining the identified set in that case can also be used in the continuous X case to define inequalities which, when constructed using probability distributions $F_{YX|Z}$ for $z \in \mathcal{Z}$, must be satisfied by all structural functions in the observationally equivalent structures that generate that distribution. At this time we do not have a proof of sharpness when $M > 2$ in the continuous case.[18] The development given here also applies in the discrete X case.

Consider probability distributions $F_{YX|Z}$ for $z \in \mathcal{Z}$ and an admissible structural function $h(\cdot, \cdot)$. Consider values $x \in \mathcal{X}$ and $z \in \mathcal{Z}$ and consider restrictions on admissible distributions $F_{UX|Z}$ which must hold if the structural function h and a distribution $F_{UX|Z}$ are to define a structure which generates the probability distribution $F_{YX|Z}$ with $Z = z$.

We derive expressions for the minimal and maximal probability mass that $F_{U|XZ}$ with $X = x$ and $Z = z$ can assign at or below a value $u \in (0, 1]$ given the requirement that $F_{U|XZ}$ and h must deliver the probability distribution $F_{Y|XZ}$ with $X = x$ and $Z = z$. Let these minimal and maximal probability masses be denoted by respectively $\underline{B}(u, x, z)$ and $\overline{B}(u, x, z)$ so that there is the following inequality.

$$\underline{B}(u, x, z) \leq F_{U|XZ}(u|x, z) \leq \overline{B}(u, x, z) \tag{3.32}$$

These bounds depend on h and on the probability distribution $F_{Y|XZ}(y|x, z)$.

Integrating with respect to the conditional distribution of X given $Z = z$ and, noting that the independence restriction requires

$$\int F_{U|XZ}(u|x, z) dF_{X|Z}(x|z) = u$$

holds for all $u \in (0, 1]$ and $z \in \mathcal{Z}$, delivers the following inequality which holds for all $u \in (0, 1]$ and $z \in \mathcal{Z}$.

$$\int \underline{B}(u, x, z) dF_{X|Z}(x|z) \leq u \leq \int \overline{B}(u, x, z) dF_{X|Z}(x|z) \tag{3.33}$$

The inequality places restrictions on h and involves $F_{YX|Z}$ with $Z = z$.

Now consider two values of U, u_1 and u_2 with $u_1 > u_2$, and the *maximal* probability

[18] The set derived in this Section is shown to be the identified set in the binary Y continuous X case in Chesher (2010).

mass that an admissible distribution function $F_{U|XZ}$ with $X = x$ and $Z = z$ can assign to the interval $(u_2, u_1]$ given the requirement that $F_{U|XZ}$ and h must deliver the probability distribution $F_{Y|XZ}$ with $X = x$ and $Z = z$. This is equal to the maximal mass that can lie at or below u_1 less the minimal mass that can lie below u_2, as follows.

$$F_{U|XZ}(u_1|x,z) - F_{U|XZ}(u_2|x,z) \leq \overline{B}(u_1,x,z) - \underline{B}(u_2,x,z) \tag{3.34}$$

The expression on the right hand side here is always non-negative because

$$\overline{B}(u_1,x,z) \geq \underline{B}(u_1,x,z) \geq \underline{B}(u_2,x,z)$$

the second inequality following because $u_1 > u_2$. The inequality (3.34) does not improve on the inequality (3.32) which delivers the inequality (3.34) on adding the two inequalities:

$$\begin{aligned} F_{U|XZ}(u_1|x,z) &\leq \overline{B}(u_1,x,z) \\ -F_{U|XZ}(u_2|x,z) &\leq -\underline{B}(u_2,x,z) \end{aligned}$$

which come directly from (3.32).

Now consider two values of U, u_1 and u_2 with $u_1 > u_2$ and the *minimal* probability mass that an admissible distribution function $F_{U|XZ}$ with $X = x$ and $Z = z$ can assign to the interval $(u_2, u_1]$ given the requirement that $F_{UX|Z}$ and h must deliver the probability distribution $F_{Y|XZ}$ with $X = x$ and $Z = z$.

This is at least equal to the minimal mass required to lie at or below u_1 minus the maximal mass that can lie below u_2, so there is the following inequality for all $(u_1, u_2) \in (0, 1]$ with $u_1 > u_2$ and for all $z \in \mathcal{Z}$.

$$F_{U|XZ}(u_1|x,z) - F_{U|XZ}(u_2|x,z) \geq \underline{B}(u_1,x,z) - \overline{B}(u_2,x,z)$$

This is a direct implication of the inequality (3.32). However the right hand side of this inequality can be negative, and since $u_1 > u_2$, an admissible distribution must satisfy

$$F_{U|XZ}(u_1|x,z) \geq F_{U|XZ}(u_2|x,z)$$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

and so the bound can be improved as follows.

$$F_{U|XZ}(u_1|x,z) - F_{U|XZ}(u_2|x,z) \geq \max(0, \underline{B}(u_1,x,z) - \overline{B}(u_2,x,z)) \tag{3.35}$$

Integrating with respect to the conditional distribution of X given $Z = z$ and exploiting the independence restriction gives the following inequality which must hold for all $(u_1, u_2) \in (0,1]$ with $u_1 > u_2$ and for all $z \in \mathcal{Z}$.

$$u_1 - u_2 \geq \int \max(0, \underline{B}(u_1,x,z) - \overline{B}(u_2,x,z)) dF_{X|Z}(x|z) \tag{3.36}$$

The inequalities (3.33) and (3.36) depend only on the structural function $h(\cdot,\cdot)$ and the distribution $F_{YX|Z}(y,x|z)$ at the value z under consideration. They must hold for all $z \in \mathcal{Z}$ leading to the following intersection bounds.

$$\max_{z\in\mathcal{Z}} \int \underline{B}(u,x,z) dF_{X|Z}(x|z) \leq u \leq \min_{z\in\mathcal{Z}} \int \overline{B}(u,x,z) dF_{X|Z}(x|z) \tag{3.37}$$

$$u_1 - u_2 \geq \max_{z\in\mathcal{Z}} \int \max(0, \underline{B}(u_1,x,z) - \overline{B}(u_2,x,z)) dF_{X|Z}(x|z) \tag{3.38}$$

It remains to determine expressions for $\underline{B}(u,x,z)$ and $\overline{B}(u,x,z)$.

First consider the **upper bound** $\overline{B}(u,x,z)$ and consider values $u \in (h_{m-1}(x), h_m(x)]$. In this case $Y = m$, the distribution $F_{U|XZ}$ cannot allocate a probability mass larger than $\Pr[Y \leq m|X = x, Z = z]$ at or below the value u, and so there is the following.

$$\overline{B}(u,x,z) = \Pr[Y \leq m|X = x, Z = z] = \Pr[Y \leq h(x,u)|X = x, Z = z] \tag{3.39}$$

Now consider the **lower bound**. Suppose $u \in (h_{m-1}(x), h_m(x))$. In this case $h(x,u) = m$. The distribution $F_{U|XZ}$ could allocate all the probability mass associated with $Y = m$ to the interval $(u, h_m(x)]$ so the probability mass *required* to fall at or below u is as follows.

$$\underline{B}(u,x,z) = \Pr[Y < m|X = x, Z = z] = \Pr[Y < h(x,u)|X = x, Z = z]$$

If $u = h_m(x)$ then $F_{U|XZ}$ must allocate the probability mass associated with $Y = m$ at or

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

below u as well and the total probability mass required to fall at or below u is:

$$\underline{B}(u,x,z) = \Pr[Y \leq m|X=x,Z=z] = \Pr[Y \leq h(x,u)|X=x,Z=z].$$

On considering all intervals in which u may lie, there is the following expression for the lower bound.

$$\underline{B}(u,x,z) = \Pr[Y < h(x,u)|X=x,Z=z] + \sum_{m=1}^{M} 1[u = h_m(x)] \Pr[Y = m|X=x,Z=z] \tag{3.40}$$

Substituting for $\underline{B}(u,x,z)$ and $\overline{B}(u,x,z)$ in (3.37) gives the following inequalities, which are satisfied by all structural functions in the identified set for all $u \in (0,1]$.

$$\max_{z\in\mathcal{Z}}\left\{\Pr[Y < h(X,u)|Z=z] + \sum_{m=1}^{M}\int 1[u = h_m(x)] \Pr[Y=m|X=x,Z=z]dF_{X|Z}(x|z)\right\}$$
$$\leq u \leq$$
$$\min_{z\in\mathcal{Z}} \Pr[Y \leq h(X,u)|Z=z] \tag{3.41}$$

The bounds for binary Y and discrete X given in Chesher (2009), which are shown there to define the identified set, arise as a special case.

When X is continuous events $\{u_1 = h_m(X)\}$ have measure zero so the second term on the left hand side of (3.41) is vanishingly small and the inequalities are as follows.

$$\max_{z\in\mathcal{Z}} \Pr[Y < h(X,u)|Z=z] < u \leq \min_{z\in\mathcal{Z}} \Pr[Y \leq h(X,u)|Z=z]$$

These inequalities were shown to define the identified set for binary Y and continuous X in Chesher (2010).

Substituting for $\underline{B}(u,x,z)$ and $\overline{B}(u,x,z)$ in (3.38) gives the following inequalities, which are satisfied by all structural functions in the identified set for all $(u_1,u_2) \in (0,1]$ with

$u_1 > u_2$.

$$u_1 - u_2 \geq \max_{z \in \mathcal{Z}} \int \max \{0, \Pr[Y < h(x, u_1)|X = x, Z = z] + \sum_{m=1}^{M} 1[u_1 = h_m(x)] \Pr[Y = m|X = x, Z = z] - \Pr[Y \leq h(x, u_2)|X = x, Z = z]\} \, dF_{X|Z}(x|z) \quad (3.42)$$

When X is continuously distributed the second term makes an infinitessimal contribution and the inequality simplifies as follows.

$$u_1 - u_2 > \max_{z \in \mathcal{Z}} \int \max \{0, \Pr[Y < h(x, u_1)|X = x, Z = z] - \Pr[Y \leq h(x, u_2)|X = x, Z = z]\} \, dF_{X|Z}(x|z)$$

The bounds (3.41) and (3.42) are precisely those given in (3.28) - (3.31) for the discrete X case. This is so because setting $u = \gamma_{[n]}$ in (3.39) and (3.40) for $n \in \{1, \ldots, N\}$ gives the following expressions

$$\overline{B}(\gamma_{[n]}, x, z) = \Pr[Y \leq h(x, \gamma_{[n]})|X = x, Z = z]$$

$$\underline{B}(\gamma_{[n]}, x, z) = \Pr[Y < h(x, \gamma_{[n+1]})|X = x, Z = z]$$

which on evaluating at $x = x_k$ and multiplying by $\Pr[X = x_k|Z = z]$ lead to the expressions for $\lambda_{nk}(z)$ and $\pi_{nk}(z)$ given in equations (3.26) and (3.27) in Section 3.12.

The set of structural functions which satisfy the inequalities (3.41) for all $u \in (0.1]$ and the inequalities (3.42) for all $(u_1, u_2) \in (0, 1]$ define a set of structural functions denoted by $\mathcal{E}^A(\mathcal{Z})$. We have shown that when X is discrete this is the set $\mathcal{E}^A(\mathcal{Z}) = \mathcal{E}(\mathcal{Z})$ which has been shown to equal the identified set, $\mathcal{H}(\mathcal{Z})$. The argument given in this Section shows that when X is continuous (or mixed discrete-continuous) $\mathcal{H}(\mathcal{Z}) \subseteq \mathcal{E}^A(\mathcal{Z})$. We conjecture that $\mathcal{E}^A(\mathcal{Z})$ is the identified set in the non-discrete X case as well.

3.5 Illustrative calculations, computation and estimation

3.5.1 Examples of bounds

We enumerate the bounds for a case with $M = 3$ and $K = 3$ and the arrangement of γ shown in equation (3.7) that has been considered throughout the paper.

Table 3.7 shows the values of lower bounds on $\gamma_{[s]} - \gamma_{[r]}$ for s (columns) and r (rows) varying in $\{1, \ldots, 7\}$. For example the entry in the row for $\gamma_{[1]}$ and the column for $\gamma_{[3]}$ gives the bound

$$\gamma_{[3]} - \gamma_{[1]} \geq \rho_{[2]}$$

that is

$$\gamma_{21} - \gamma_{11} \geq \rho_{21}$$

and note that this must hold for all $z \in \mathcal{Z}$. As z varies ρ_{21} varies and making this dependence explicit and dependence on the arrangement explicit too there is the bound

$$\gamma_{21} - \gamma_{11} \geq \max_{z \in \mathcal{Z}} \rho^t_{21}(z)$$

which contributes to the bounds defining $\mathcal{E}_t(\mathcal{Z})$.

The model places no restrictions on some of the differences other than those arising because of the ordering in the arrangement under consideration. An example is $\gamma_{[4]} - \gamma_{[2]}$ which is only required to be non-negative. Some of the restrictions that define a set $\mathcal{E}_t(\mathcal{Z})$ render others redundant. For example in Table 3.7 there is the restriction

$$\gamma_{[5]} - \gamma_{[4]} \geq \rho_{[5]} \tag{3.43}$$

which when satisfied ensures that two other restrictions are satisfied as follows.

$$\begin{aligned} \gamma_{[5]} - \gamma_{[3]} &\geq \rho_{[5]} \\ \gamma_{[5]} - \gamma_{[2]} &\geq \rho_{[5]} \end{aligned}$$

The restrictions $\gamma_{[6]} - \gamma_{[3]} \geq \rho_{[5]}$ and $\gamma_{[6]} - \gamma_{[4]} \geq \rho_{[5]}$ are also redundant, both being implied by the restriction (3.43).

In the final column lies $\gamma_{[7]}$ which is equal to 1. The entries in this column give lower

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

bounds on $1 - \gamma_{[r]}$ where r varies from 1 to 6 down the rows of the Table. Subtracting these entries from 1 (i.e. eliminating the leading unit terms and changing the signs of what remains) delivers *upper* bounds on $\gamma_{[r]}$ for $r \in \{1, \ldots, 6\}$.

Lower bounds on the $\gamma_{[r]}$'s are simply, for each $r \in \{1, \ldots, 6\}$

$$\sum_{r'=1}^{r} \rho_{[r']} \leq \gamma_{[r]}$$

as shown in Section 3.3.1.

Adding the negative of the upper bound for $\gamma_{[r]}$ to the lower bound for $\gamma_{[s]}$ delivers a lower bound on $\gamma_{[s]} - \gamma_{[r]}$ which we can compare with the bounds shown in Table 3.7. Doing this we find that the lower bounds on $\gamma_{[4]} - \gamma_{[1]}$, $\gamma_{[5]} - \gamma_{[1]}$, $\gamma_{[6]} - \gamma_{[1]}$ and $\gamma_{[6]} - \gamma_{[2]}$ in Table 3.7 are exactly the bounds obtained by comparing lower and upper bounds on individual elements of γ.

The only inequality in Table 3.7 that survives these various eliminations is $\gamma_{[5]} - \gamma_{[4]} \geq \rho_{[5]}$. So for this arrangement the set $\mathcal{E}_t(\mathcal{Z})$ is defined by this inequality and the lower and upper bounds on the individual elements of γ and the inequalities that express the ordering of the elements of γ in this arrangement.

In the $M = 3$, $K = 3$ example considered in detail in this paper there are 90 admissible arrangements of γ of which 15 are fundamental in the sense that each of these 15 generates $3! = 6$ arrangements by permuting the index identifying the three values of the conditioning variable. Annex 2 shows the bounds on $\gamma_{[s]} - \gamma_{[r]}$ just as in Table 3.7 for each of these 15 fundamental arrangements. In the sequence presented there the arrangement considered in this Section is number 8.

Comparisons amongst the inequalities on differences of elements of γ and comparing those inequalities with the implications of the lower and upper bounds on elements of γ leads to elimination of large numbers of the entries in the tables that refer to differences $\gamma_{[s]} - \gamma_{[r]}$ for s and r in $\{1, \ldots, 6\}$. In Arrangement 1 all such inequalities on differences disappear. In Arrangement 2 only the inequality $\gamma_{[5]} - \gamma_{[3]} \geq \rho_{[5]}$ remains. In Arrangement 3 only the inequality $\gamma_{[4]} - \gamma_{[2]} \geq \rho_{[4]}$ remains. In other cases there are more survivors. For example in

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

		γ_{11}	γ_{12}	γ_{21}	γ_{13}	γ_{23}	γ_{22}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	$\rho_{[2]}$	$\rho_{[3]}$	$\rho_{[3]} + \rho_{[5]}$	$\rho_{[3]} + \rho_{[5]} + \rho_{[6]}$	$1 - \rho_{[1]} - \rho_{[2]} - \rho_{[4]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	0	$\rho_{[5]}$	$\rho_{[5]} + \rho_{[6]}$	$1 - \rho_{[1]} - \rho_{[2]} - \rho_{[3]} - \rho_{[4]}$
γ_{21}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[5]}$	$\rho_{[5]}$	$1 - \rho_{[1]} - \rho_{[2]} - \rho_{[3]} - \rho_{[4]} - \rho_{[6]}$
γ_{13}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[5]}$	$\rho_{[5]}$	$1 - \delta_1 - \rho_{[2]} - \rho_{[4]} - \rho_{[6]}$
γ_{23}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1 - \delta_1 - \rho_{[2]} - \rho_{[4]} - \rho_{[5]} - \rho_{[6]}$
γ_{22}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1 - \delta_1 - \delta_3 - \rho_{[2]} - \rho_{[6]}$

Table 3.7: Values of lower bounds on $\gamma_{[s]} - \gamma_{[r]}$ for s (in columns) and r (in rows) for the example arrangement.

Arrangement 11 the following three inequalities on differences of elements of γ survive.

$$\begin{aligned} \gamma_{[3]} - \gamma_{[2]} &\geq \rho_{[3]} \\ \gamma_{[5]} - \gamma_{[4]} &\geq \rho_{[5]} \\ \gamma_{[6]} - \gamma_{[2]} &\geq \rho_{[5]} + \rho_{[6]} \end{aligned}$$

In the example considered here the number of discrete outcomes is $M = 3$ and the number of points of support of the endogenous variables is $K = 3$. When M or K are larger there are many more contributions to the definitions of sets $\mathcal{E}_t(\mathcal{Z})$ coming from inequalities involving differences of elements of γ.

3.5.2 A `Mathematica` notebook

This paper is accompanied by a `Mathematica` notebook which is viewable in the freeware `Math Player 7`.[19] The notebook does symbolic calculation of bounds as set out in Table 3.7. The user provides values for M, the number of discrete outcomes and K the number of points of support of the endogenous variables. A stylised graphical display of the $M - 1$ threshold functions appears with the values associated with the K points of support of an endogenous variable X highlighted. The user can manipulate these thereby generating particular arrangements of γ. For each arrangement t selected, the notebook produces a table like Table 3.7 showing in symbolic form the inequalities defining a set $\mathcal{E}_t(z)$.

[19]The notebook can be downloaded from www.cemmap.ac.uk/wps/sisdvivm.nbp. `Math Reader 7` is available at: http://www.wolfram.com/products/player/download.cgi .

3.5.3 Computation and estimation

When M and K are both large computation of the set $\mathcal{E}(\mathcal{Z})$ is challenging because of the large number of potential arrangements of γ, that is of the K values of the $M-1$ threshold functions, that may arise. For example when $M = K = 4$ there are $369,600$ admissible arrangements rising to over 300 billion when $M = K = 5$. *Shape restrictions* are helpful in reducing the scale of the problem.

In the binary outcome SEIV model a monotonicity restriction coupled with a single index restriction, requiring that the threshold function is a monotone function of a scalar valued function of endogenous and exogenous variables, brings great simplification as shown in Chesher (2009). The use and benefit of restrictions on threshold functions such as monotonicity, concavity, convexity and single-peakedness coupled with index restrictions is the subject of current research.

Shape restrictions can also be introduced by employing constrained sieve approximations. Parametric restrictions cut down the scale of the problem and provide a link to classical likelihood based analysis of discrete outcome data. This is illustrated in Chesher and Smolinski (2009) where ordered probit structural functions are employed with a coefficient on a scalar endogenous variable X that is common across threshold functions whose "intercept terms" differ. This model embodies strong shape restrictions, requiring threshold functions to be monotone in X and parallel after applying the inverse normal distribution function transformation.

As a prelude to consideration of methods for estimating identified sets in parametric or otherwise shape constrained models, first consider a *theoretical* analysis in which one has to hand probability distributions $\Pr[Y = m \wedge X = x_k | Z = z]$ for each value $z \in \mathcal{Z}$. Suppose there is a parametric model or sieve approximating model for the structural function with parameter vector Θ. For any value θ there is an associated value of γ denoted by $\gamma(\theta)$ which is in some arrangement denoted by $t(\theta)$. The values of $\gamma(\theta)$ and $t(\theta)$ are easy to compute. The value θ is in the identified set of parameter values, denoted by $\mathcal{H}^{\Theta}(\mathcal{Z})$, if and only if $\gamma(\theta) \in \mathcal{E}_{t(\theta)}(\mathcal{Z})$.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Define the non-negative valued distance measure $D(\theta)$ as follows.

$$D(\theta) \equiv \min_{w \in \mathcal{E}_{t(\theta)}(\mathcal{Z})} \left((\gamma(\theta) - w)' (\gamma(\theta) - w) \right) \tag{3.44}$$

This is the squared Euclidean distance from $\gamma(\theta)$ to the point in the set $\mathcal{E}_{t(\theta)}(\mathcal{Z})$ closest to $\gamma(\theta)$ as the crow flies. The measure is zero if and only if $\gamma(\theta) \in \mathcal{E}_{t(\theta)}(\mathcal{Z})$ and so zero if and only if $\theta \in \mathcal{H}^{\Theta}(\mathcal{Z})$. The value of $D(\theta)$ is easily found using a quadratic programming algorithm and the expressions for the linear half spaces defining the sets $\mathcal{E}_t(\mathcal{Z})$ that we have given in this paper.[20] The set of values of θ that minimise the function $D(\cdot)$ is the identified set $\mathcal{H}^{\Theta}(\mathcal{Z})$.

$$\mathcal{H}^{\Theta}(\mathcal{Z}) = \{\theta : \theta = \arg\min_{s} D(s)\} = \{\theta : D(\theta) = 0\}.$$

In applied econometric work there will be *estimates* of the probability distributions, $\Pr[Y = m \wedge X = x_k | Z = z]$ for $z \in \mathcal{Z}$, and so estimates of the sets $\mathcal{E}_t(\mathcal{Z})$. Let $\hat{D}(\theta)$ be the distance measure arising when $\mathcal{E}_t(\mathcal{Z})$ in (3.44) is replaced by an estimate $\hat{\mathcal{E}}_t(\mathcal{Z})$. The distance measure $\hat{D}(\theta)$ has the properties required of Chernozhukov, Hong and Tamer's (2007) "econometric criterion function" $Q(\theta)$ and their methods can be employed to estimate, and develop confidence regions for, the set $\mathcal{H}^{\Theta}(\mathcal{Z})$.

It will be prudent to use bias corrected estimates of the sets $\mathcal{E}_t(\mathcal{Z})$. Bias arises because the sets $\mathcal{E}_t(\mathcal{Z})$ arise as *intersections* of sets $\mathcal{E}_t(z)$ across values $z \in \mathcal{Z}$. The issue is explained in Chernozhukov, Lee and Rosen (2009) (CLR) where a solution is proposed. This is directly applicable in the case that arises here.

Define $\rho^t(z) \equiv \{\rho^t_{[1]}(z), \ldots, \rho^t_{[N]}(z)\}$. With $\gamma_{[0]} \equiv 0$ and $\gamma_{[N+1]} \equiv 1$ all the constraints defining a set $\mathcal{E}_t(\mathcal{Z})$ have the form

$$\gamma_{[s]} - \gamma_{[r]} \geq \max_{z \in \mathcal{Z}} \left(\alpha_{sr} \cdot \rho^t(z) \right) \tag{3.45}$$

for certain pairs of indices $s > r$ selected from $\{0, 1, \ldots, N+1\}$. Here α_{sr} is a vector of integers specific to the $s - r$ comparison.

The proposal in CLR is to calculate an estimate of $l^t(\alpha_{sr}, \mathcal{Z}) \equiv \max_{z \in \mathcal{Z}} \left(\alpha_{sr} \cdot \rho^t(z) \right)$ by

[20] An L1 norm and a linear programming calculation could be employed instead.

calculating the maximum over $z \in \mathcal{Z}$ of *precision corrected* estimates as follows.

$$\hat{l}^t(\alpha_{sr}, \mathcal{Z}) = \max_{z \in \hat{\mathcal{Z}}} \left(\alpha_{sr} \cdot \hat{\rho}^t(z) + \kappa \sigma^t(\alpha_{sr}, z) \right)$$

Here $\sigma^t(\alpha_{sr}, z)$ is the standard error of $\alpha_{sr} \cdot \hat{\rho}^t(z)$, $\hat{\mathcal{Z}}$ is a data dependent set of values of z that converges in probability to a non-stochastic set which contains $\arg\max_{z \in \mathcal{Z}} \left(\alpha_{sr} \cdot \rho^t(z) \right)$ and κ is an estimate of

$$median \left(\inf_{z \in \hat{\mathcal{Z}}} \frac{\alpha_{sr} \cdot \rho^t(z) - \alpha_{sr} \cdot \hat{\rho}^t(z)}{\sigma^t(\alpha_{sr}, z)} \right)$$

proposals for which are given in CLR.

The result is an asymptotically upward median unbiased estimate of the bound in (3.45). Proceeding in this way gives bias corrected estimates of all bounds and thus bias corrected estimated sets $\hat{\mathcal{E}}_t(\mathcal{Z})$ which will be used in the calculation of the distance measure $\hat{D}(\theta)$. An example of inference using the CLR method in a binary outcome case is given in Chesher (2009).

3.6 Concluding remarks

We have studied identification of a nonparametrically specified structural function for a discrete outcome, focussing attention mainly on the discrete endogenous variable case. The single equation instrumental variable (SEIV) model we have considered is attractive because it places no restrictions on the process generating values of endogenous variables. Commonly used control function alternatives based on triangular models do not deliver point identification when, as here, endogenous variables are discrete unless there are strong parametric restrictions.

The SEIV model *set* identifies the structural function. In the M outcome case the structural function is characterised by $M - 1$ threshold functions. When endogenous variables are discrete the identified set is a union of many convex sets. In principle there is one such set associated with each admissible ordering of the K values taken by $M - 1$ threshold functions as endogenous variables pass across their K points of support.

Each convex component of the identified set is the intersection of collections of linear half spaces, each value of the instrumental variables generating one such collection. The number and extent of the convex components of the identified set depends on the strength and support of the instrumental variables. When these are good predictors of the values of endogenous variables the identified set may comprise just a small number of convex components, perhaps just one.

We have developed expressions for a set $\mathcal{E}(\mathcal{Z})$ which can be calculated for any probability distribution of the outcome Y and endogenous variables X given instruments Z taking values in a set $\mathcal{Z}$. We have shown that when endogenous variables are discrete the set identified by the SEIV model, $\mathcal{H}(\mathcal{Z})$, is equal to $\mathcal{E}(\mathcal{Z})$. We provide a `Mathematica` notebook which conducts symbolic calculation of convex components of the identified set.

Unrestricted nonparametric estimation and inference pose challenging problems once M or K are at all large. Parametric restrictions or shape restrictions reduce the scale of the estimation problem. We have defined an easy-to-compute criterion function which can be employed in estimation using the methods proposed in Chernozhukov, Hong and Tamer (2007) with bias corrected estimates of bounds as proposed in Chernozhukov, Lee and Rosen (2009).

A set of structural functions, $\mathcal{E}^A(\mathcal{Z})$, has been derived for the general case in which endogenous variables are continuous, mixed discrete-continuous or discrete. When X is discrete the set $\mathcal{E}^A(\mathcal{Z})$ is the set $\mathcal{E}(\mathcal{Z})$ which is the identified set of structural functions when endogenous variables are discrete. We have shown that in the non-discrete case all structural functions in the identified set lie in the set $\mathcal{E}^A(\mathcal{Z})$ and conjecture that in this case too it is the identified set.

References

Blundell Richard W., and James L. Powell (2003): “Endogeneity in Nonparametric and Semiparametric Regression Models,” in Dewatripont, M., L.P. Hansen, and S.J. Turnovsky, eds., *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Vol. II.* Cambridge: Cambridge University Press.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Blundell, Richard W., and James .L. Powell (2004): "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, 71, 655-679.

Chernozhukov, V., and C. Hansen (2001): "An IV Model of Quantile Treatment Effects," Working Paper 02-06, Department of Economics, MIT.

Chernozhukov, V., And C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245-261.

Chernozhukov, V., H. Hong and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243-1284.

Chernozhukov, V., S. Lee, and A. Rosen (2009): "Intersection Bounds: Estimation and Inference," CeMMAP Working Paper 19/09.

Chesher, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405-1441.

Chesher, A. (2005): "Nonparametric Identification under Discrete Variation," *Econometrica*, 73, 1525-1550.

Chesher, A. (2007): "Endogeneity and Discrete Outcomes," CeMMAP Working Paper 05/07.

Chesher, A. (2009): "Single Equation Endogenous Binary Response Models," CeMMAP Working Paper 23/09.

Chesher, A. (2010): "Instrumental Variable Models for Discrete Outcomes," *Econometrica*, 78, 575-602.

Chesher, A., and K. Smolinski (2009): "IV Models of Ordered Choice," CeMMAP Working Paper 37/09.

Das, M. (2005): "Instrumental Variables Estimators of Nonparametric Models with Discrete Endogenous Regressors," *Journal of Econometrics*, 124, 335-361.

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

Imbens, G., and W. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77, 1481–1512.

Jun, S. J., J. Pinkse and H. Xu (2009): "Tighter Bounds in Triangular Systems," unpublished paper.

Lewbel, A. (2007). "Coherency And Completeness Of Structural Models Containing A Dummy Endogenous Variable," *International Economic Review*, 48, 1379-1392.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

TAMER, E. (2003). "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70, 147–165.

ZIEGLER, G., (2007): *Lectures on Polytopes.* (Updated seventh printing of the first edition). New York: Springer Science.

ANNEX 1: PROOF OF PROPOSITION 1

1. (a) This follows directly from (3.12) which states that for all i and j, $\eta(i,j)$ is either equal to $G(i,j)$ or, equal to $\max(0, R(i,j))$ if this is less than $G(i,j)$.

   (b) For any i and j, if $G(i,j) \geq 0$ then $\eta(i,j) \geq 0$ because (3.12) states that $\eta(i,j)$ is at least equal to $G(i,j)$ or a non-negative quantity, namely $\max(0, R(i,j))$. The recursion (3.13) taken together with (a) and $G(i,j) \geq 0 \implies \eta(i,j) \geq 0$ implies that, for any i and j, if $G(i,j-1) \geq 0$ then $G(i,j) \geq 0$. Since for all i, $G(i,1) = \Delta\gamma_{[i]} \geq 0$, the result follows by induction letting j pass from 2 to K.

   (c) As noted in the proof of (b), for all i and j, $G(i,j) \geq 0 \implies \eta(i,j) \geq 0$ and the result follows because the result (b) states that for all i and j, indeed, $G(i,j) \geq 0$.

   (d) This follows directly from (3.13) and (a) and (c).

   (e) If $G(i,j) > \eta(i,j)$ then $G(i,j) > 0$ and since $G(i,j)$ is a non-increasing function of j, for all $j' \leq j$, $G(i,j') > 0$. Therefore, for all $j' < j$, from (3.13), $G(i,j') > \eta(i,j')$ which by assumption also holds for $j' = j$. From (3.12), if $G(i,j') > \eta(i,j')$ then $\eta(i,j') = \max(0, R(i,j'))$.

2. Suppose that for some $j \leq K$, $G(i,j) \leq \max(0, R(i,j))$. Then $\eta(i,j) = G(i,j)$ and from (3.13) $G(i,j+1) = 0$ and by repeated application of (3.13), for all $j' > j$, $\eta(i,j') = G(i,j') = 0$ and so

$$G(i, K+1) = \Delta\gamma_{[i]} - \sum_{k=1}^{K} \eta_{[i]k} = 0.$$

Suppose that there is no $j \leq K$ such that $G(i,j) \leq \max(0, R(i,j))$. Then, considering $j = K$,

$$G(i,K) > \max(0, R(i,K))$$

so

$$\eta(i,K) = \max(0, R(i,K))$$

and so from (3.13)

$$G(i,K+1) = \Delta\gamma_{[i]} - \sum_{k=1}^{K} \eta_{[i]k} = G(i,K) - \eta(i,K) > 0$$

and so

$$\sum_{k=1}^{K} \eta_{[i]k} < \Delta\gamma_{[i]}.$$

Annex 2

This Annex provides tables like Table 3.7 giving lower bounds on differences $\gamma_{[s]} - \gamma_{[r]}$ for the 15 fundamental arrangements of γ in the $M = 3$, $K = 3$ case. In each case the final column gives lower bounds on $1 - \gamma_{[r]}$ for $r \in \{1, \ldots, 6\}$. Subtracting 1 from each of these expressions and changing sign gives upper bounds on $\gamma_{[r]}$. Lower bounds are simply $\gamma_{[r]} \geq \sum_{r'=1}^{r} \rho_{[r']}$.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

$A.1$		γ_{11}	γ_{12}	γ_{13}	γ_{21}	γ_{22}	γ_{23}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	0	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[5]}$	$\rho_{[4]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	0	$\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}$
γ_{13}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	0	$\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}$
γ_{21}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	0	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{22}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[5]}-\rho_{[6]}$
γ_{23}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[3]}-\rho_{[6]}$

$A.2$		γ_{11}	γ_{12}	γ_{13}	γ_{21}	γ_{23}	γ_{22}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	0	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[5]}$	$\rho_{[4]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	0	$\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}$
γ_{13}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[5]}$	$\rho_{[5]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[6]}$
γ_{21}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	0	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{23}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[5]}-\rho_{[6]}$
γ_{22}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[3]}-\rho_{[2]}-\rho_{[6]}$

$A.3$		γ_{11}	γ_{12}	γ_{13}	γ_{22}	γ_{21}	γ_{23}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	0	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[5]}$	$\rho_{[4]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	$\rho_{[4]}$	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[5]}$
γ_{13}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	0	$\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}$
γ_{22}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	0	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{21}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[2]}-\rho_{[1]}-\rho_{[3]}-\rho_{[5]}-\rho_{[6]}$
γ_{23}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[3]}-\rho_{[6]}$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

$A.4$		γ_{11}	γ_{12}	γ_{13}	γ_{22}	γ_{23}	γ_{21}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	0	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[5]}$	$\rho_{[4]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[5]}$	$\rho_{[4]}+\rho_{[5]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[6]}$
γ_{13}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[5]}$	$\rho_{[5]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[6]}$
γ_{22}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	0	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{23}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[2]}-\rho_{[1]}-\rho_{[3]}-\rho_{[5]}-\rho_{[6]}$
γ_{21}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[2]}-\delta_{[3]}-\rho_{[1]}-\rho_{[6]}$

$A.5$		γ_{11}	γ_{12}	γ_{13}	γ_{23}	γ_{21}	γ_{22}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	0	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[5]}$	$\rho_{[4]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	$\rho_{[4]}$	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[5]}$
γ_{13}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[4]}$	$\rho_{[4]}$	$\rho_{[4]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[5]}-\rho_{[6]}$
γ_{23}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	0	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{21}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[3]}-\rho_{[1]}-\rho_{[2]}-\rho_{[5]}-\rho_{[6]}$
γ_{22}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[3]}-\rho_{[2]}-\rho_{[6]}$

$A.6$		γ_{11}	γ_{12}	γ_{13}	γ_{23}	γ_{22}	γ_{21}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	0	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[5]}$	$\rho_{[4]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[5]}$	$\rho_{[4]}+\rho_{[5]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[6]}$
γ_{13}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[4]}$	$\rho_{[4]}$	$\rho_{[4]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[5]}-\rho_{[6]}$
γ_{23}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	0	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{22}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[3]}-\rho_{[1]}-\rho_{[2]}-\rho_{[5]}-\rho_{[6]}$
γ_{21}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[2]}-\delta_{[3]}-\rho_{[1]}-\rho_{[6]}$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

A.7		γ_{11}	γ_{12}	γ_{21}	γ_{13}	γ_{22}	γ_{23}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	$\rho_{[3]}$	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[5]}$	$\rho_{[3]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[4]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	0	$\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}$
γ_{21}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	0	$\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}$
γ_{13}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[6]}$	$1-\delta_{[1]}-\rho_{[2]}-\rho_{[4]}-\rho_{[5]}$
γ_{22}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[1]}-\rho_{[2]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{23}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[4]}-\rho_{[6]}$

A.8		γ_{11}	γ_{12}	γ_{21}	γ_{13}	γ_{23}	γ_{22}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	0	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[5]}$	$\rho_{[3]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[4]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	0	$\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}$
γ_{21}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[5]}$	$\rho_{[5]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[6]}$
γ_{13}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[5]}$	$\rho_{[5]}$	$1-\delta_{[1]}-\rho_{[2]}-\rho_{[4]}-\rho_{[6]}$
γ_{23}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[5]}-\rho_{[6]}$
γ_{22}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[3]}-\rho_{[2]}-\rho_{[6]}$

A.9		γ_{11}	γ_{12}	γ_{21}	γ_{22}	γ_{13}	γ_{23}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[4]}$	$\rho_{[3]}+\rho_{[4]}$	$\rho_{[3]}+\rho_{[4]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[5]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	$\rho_{[4]}$	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[5]}$
γ_{21}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	0	$\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}$
γ_{22}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[6]}$	$1-\delta_{[1]}-\rho_{[2]}-\rho_{[4]}-\rho_{[5]}$
γ_{13}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[6]}$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[5]}$
γ_{23}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[5]}-\rho_{[6]}$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute
DOI: 10.2870/35896

$A.10$		γ_{11}	γ_{12}	γ_{22}	γ_{13}	γ_{21}	γ_{23}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	$\rho_{[3]}$	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[5]}$	$\rho_{[3]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[4]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	$\rho_{[3]}$	$\rho_{[3]}$	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[4]}-\rho_{[5]}$
γ_{22}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	0	$\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}$
γ_{13}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[6]}$	$1-\delta_{[2]}-\rho_{[1]}-\rho_{[4]}-\rho_{[5]}$
γ_{21}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[2]}-\rho_{[1]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{23}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[4]}-\rho_{[6]}$

$A.11$		γ_{11}	γ_{12}	γ_{22}	γ_{13}	γ_{23}	γ_{21}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	$\rho_{[3]}$	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[5]}$	$\rho_{[3]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[4]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	$\rho_{[3]}$	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[4]}-\rho_{[6]}$
γ_{22}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[5]}$	$\rho_{[5]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[6]}$
γ_{13}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[5]}$	$\rho_{[5]}$	$1-\delta_{[2]}-\rho_{[1]}-\rho_{[4]}-\rho_{[6]}$
γ_{23}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[2]}-\rho_{[1]}-\rho_{[3]}-\rho_{[5]}-\rho_{[6]}$
γ_{21}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[2]}-\delta_{[3]}-\rho_{[1]}-\rho_{[6]}$

$A.12$		γ_{11}	γ_{12}	γ_{22}	γ_{21}	γ_{13}	γ_{23}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[4]}$	$\rho_{[3]}+\rho_{[4]}$	$\rho_{[3]}+\rho_{[4]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[5]}$
γ_{12}	$\gamma_{[2]}$	$\cdot$	$\cdot$	$\rho_{[3]}$	$\rho_{[3]}$	$\rho_{[3]}$	$\rho_{[3]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[4]}-\rho_{[5]}$
γ_{22}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	0	$\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}$
γ_{21}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[6]}$	$1-\delta_{[2]}-\rho_{[1]}-\rho_{[4]}-\rho_{[5]}$
γ_{13}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[6]}$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[5]}$
γ_{23}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[5]}-\rho_{[6]}$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

$A.13$		γ_{11}	γ_{21}	γ_{12}	γ_{13}	γ_{22}	γ_{23}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	0	$\rho_{[2]}$	$\rho_{[2]}$	$\rho_{[2]}+\rho_{[5]}$	$\rho_{[2]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[3]}-\rho_{[4]}$
γ_{21}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	0	$\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}$
γ_{12}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\delta_{[1]}-\rho_{[3]}-\rho_{[4]}$
γ_{13}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[6]}$	$1-\delta_{[1]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}$
γ_{22}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[1]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{23}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[4]}-\rho_{[6]}$

$A.14$		γ_{11}	γ_{21}	γ_{12}	γ_{13}	γ_{23}	γ_{22}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	$\rho_{[2]}$	$\rho_{[2]}$	$\rho_{[2]}$	$\rho_{[2]}+\rho_{[5]}$	$\rho_{[2]}+\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[3]}-\rho_{[4]}$
γ_{21}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	0	$\rho_{[3]}+\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[4]}$
γ_{12}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[5]}$	$\rho_{[5]}+\rho_{[6]}$	$1-\delta_{[1]}-\rho_{[3]}-\rho_{[4]}$
γ_{13}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[5]}$	$\rho_{[5]}$	$1-\delta_{[1]}-\rho_{[3]}-\rho_{[4]}-\rho_{[6]}$
γ_{23}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$1-\delta_{[1]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}-\rho_{[6]}$
γ_{22}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[3]}-\rho_{[3]}-\rho_{[6]}$

$A.15$		γ_{11}	γ_{21}	γ_{12}	γ_{22}	γ_{13}	γ_{23}	1
		$\gamma_{[1]}$	$\gamma_{[2]}$	$\gamma_{[3]}$	$\gamma_{[4]}$	$\gamma_{[5]}$	$\gamma_{[6]}$	$\gamma_{[7]}$
γ_{11}	$\gamma_{[1]}$	$\cdot$	$\rho_{[2]}$	$\rho_{[2]}$	$\rho_{[2]}+\rho_{[4]}$	$\rho_{[2]}+\rho_{[4]}$	$\rho_{[2]}+\rho_{[4]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[3]}-\rho_{[5]}$
γ_{21}	$\gamma_{[2]}$	$\cdot$	$\cdot$	0	$\rho_{[4]}$	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[6]}$	$1-\rho_{[1]}-\rho_{[2]}-\rho_{[3]}-\rho_{[5]}$
γ_{12}	$\gamma_{[3]}$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[4]}$	$\rho_{[4]}$	$\rho_{[4]}+\rho_{[6]}$	$1-\delta_{[1]}-\rho_{[3]}-\rho_{[5]}$
γ_{22}	$\gamma_{[4]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	0	$\rho_{[6]}$	$1-\delta_{[1]}-\rho_{[3]}-\rho_{[4]}-\rho_{[5]}$
γ_{13}	$\gamma_{[5]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\rho_{[6]}$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[5]}$
γ_{23}	$\gamma_{[6]}$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$\cdot$	$1-\delta_{[1]}-\delta_{[2]}-\rho_{[5]}-\rho_{[6]}$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

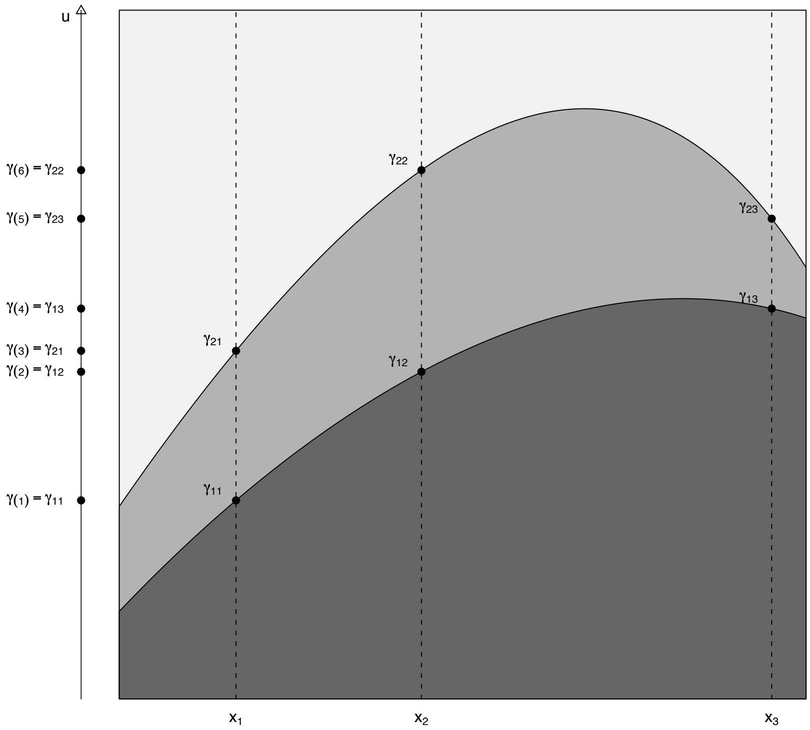


Figure 3.1: Examples of two threshold functions for the case $M = 3$ and $K = 3$ that are consonant with the arrangement of elements of γ shown on the vertical axis. The outcome Y takes the value 1 below the lowest threshold in the dark shaded region and the value 3 above the highest threshold in the light shaded region. The vertical scale is the unit interval $[0, 1]$.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Chapter 4

Core determining indexes for set identified models with discrete observables

This paper introduces a constructive and practical algorithm for obtaining *core determining indexes* in a class of partially identified models. Core determining indexes give rise to core determining sets and inequalities that are necessary and sufficient for the analysis of identification.

We focus on designs with discrete observables and continuous latent heterogeneity. Examples comprise ordered outcome, multinomial choice or random coefficient instrumental variable models. We elucidate the method with an elementary example where a binary endogenous covariate drives a three valued outcome. In the illustration we examine a non-parametric ordered outcome model with endogeneity. We invoke an instrumental variable restriction for identification.

Introduction

We propose a practical method to establish a finite number of moment inequalities characterizing the identified sets in the class of partially identified models when the observable variables are discrete. The method comprises of a large class of complete models prevalent

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

in econometric practice. Instrumental variable models with discrete covariates, like ordered outcome, multinomial choice, binary panel data or random coefficients models, account for particular instances.

Without lost of generality we ground our discussion on generalized instrumental variable models. We demonstrate the method by focusing on structural models where the discrete outcome variable, Y, is a function of a vector of discrete covariates, X, and a vector of latent, continuously distributed U. These models are silent about the source of endogeneity, functional relationship between X and Z. For identification we call the instrumental variable restriction that which invokes stochastic independence between latent U and the instrument Z.

$$Y = h(X, U) \qquad U \perp Z, \quad U \sim F_U$$

Unobservable U follows the probability law F_U. Chesher, Rosen, Smolinski (2011) demonstrates that this class of models set, rather than point, identify structural functions h and distributions of unobservables, F_U.

The identified set, that is a set of structural functions and distributions of unobservables, (h, F_U), obeys the following system of inequalities. For any subset S on the support of unobservables, $\mathcal{U}$,

$$Q_S(h, F_U): \qquad \int_S dF_U \geq \sum_{y \in \mathcal{Y}, x \in \mathcal{X}} \mathbf{1}(\tau_{y,x}(h) \subseteq \mathcal{S}) Pr_0(Y = y, X = x | Z = z)$$

must hold for all z in the support of the instrument, $\mathcal{Z}$. The left hand side denotes the probability mass allocated to the set S computed with respect to the probability law F_U. The right hand side denotes a probability mass taken with respect to the distribution of the data and aggregated over all level sets, $\tau_{y,x}(h)$, that are subsets of the proposed set S. A detailed definition of level sets and discussion of this result are presented in the next section.

For now it is important to point out that this theoretical formulation may fall short in practice. The inequality must hold for any subset S in the support of unobservables, $\mathcal{U}$, which in general is an uncomfortably vast space of sets and possibly infeasible to deal with in applications. However, when there is discrete variation in observables it becomes possible to characterize the identified set by a finite number of moment inequalities. This

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

characterization constitutes the essence of the discussion coming up in this paper.

There have been a few attempts to address the question of determining a finite number of moment inequalities characterizing the identified set in partially identified problems. Chesher, Rosen, Smolinski (2011) is the closest to the attempt of this paper and demonstrates a construction of a collection of a finite number of inequalities in the context of multinomial choice instrumental variable models. Here we consider any model with discrete variation in observables and a functional relationship between covariates and outcomes.

Galichon and Henry (2009) discusses incomplete models with discrete variation in unobservables. They propose to consider a power set of the support of unobservables to establish core determining sets. Bersteanu, Molchanov and Molinari (2009, 2010) use techniques similar to the results in Galichon and Henry (2009).

Our contribution to the existing literature is as follows. We extend previous results to the general class of set identified discrete choice problems with discrete covariates but potentially continuous latent heterogeneity. We provide a practical, computational method for establishing a finite number of questions that are necessary and sufficient to address identification.

In our analysis only a finite number of inequalities, say P indexed by p, characterize the identified set, (h, F_U). They constitute a small subset of all inequalities presented above. For all values of the instrument Z and all p the identified set is characterized as follows.

$$Q_p(h, F_U): \qquad \int_{u \in \mathcal{C}_p(h)} dF_U \geq \sum_{(m,k) \in \mathcal{J}_p(h)} Pr_0(Y = y_m, X = x_k | Z = z) \tag{4.1}$$

In this paper we focus on $\mathcal{C}_p(h)$ and $\mathcal{J}_p(h)$, namely *core determining sets* and *core determining indexes*. They give rise to *core determining inequalities*, $Q_p(h, F_u)$, representing the set of ultimate identification questions. Core determining indexes, $\mathcal{J}_p(h)$, gather indexes of discrete data points corresponding to the core determining sets, $\mathcal{C}_P(h)$, in a way that only some subsets of $\mathcal{U}$ need to be examined. This significantly simplifies the practical implementation of the identification analysis. The model and data dictate both the number and the form of $\mathcal{J}_p(h)$ and $\mathcal{C}_p(h)$ through sparsity of the data and shape restrictions imposed on the structural function.

Having $\mathcal{J}_p(h)$ and $\mathcal{C}_p(h)$ reduces the enormous practical challenge hidden in the general

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

characterization of the identified set. So much so that possibly an infinite number of inequalities melts away and we need to cope with a potentially large, but limited, number of identification queries.

We propose a practically feasible and easy to implement algorithm for obtaining core determining sets $C_p(h)$ and core determining indexes $J_p(h)$. We elucidate the method in the context of the nonparametric identification problem studied previously by Chesher and Smolinski (2009, 2010). The ordered outcome model with the instrumental variable restriction serves as an example.

We proceed as follows. Section 1 lays out a generic setup for the partially identified complete models represented by generalized instrumental variable structures. Section 2 discusses identification and develops core determining partitions, indexes and sets for a class of models with discrete observables and continuous latent heterogeneity. Section 3 presents a constructive algorithm to attain core determining sets. Sections 4 elucidates the results in the context of ordered outcome instrumental variable models. Finally section 5 concludes.

4.1 Set up

In this section we briefly introduce the model and present identification results. We define elemental level sets that constitute the basis of the model and identification. We present inequalities characterizing the identified set and show their equivalent representation for the discrete outcome models. We also define core determining partitions, indexes and sets.

4.1.1 The model

We illustrate identification problem in a class of instrumental variable models where unknown function h of the observable, discrete X and a vector of latent, continuously distributed U determines discrete, scalar outcome Y. We assume that latent U is distributed according to some distribution function F_U. In general this distribution remains unspecified. The model follows.

$$Y = h(X, U) \qquad U \perp Z, \qquad U \sim F_U$$

Instrumental variable restriction, $U \perp Z$, excludes dependence between U and the instrumental variable, Z. It must hold for all values of Z in the support of the instrument. We

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

restrict random vector Z to have discrete support denoted by $\mathcal{Z}$. We denote supports of Y, X and U by $\mathcal{Y}$, $\mathcal{X}$ and $\mathcal{U}$ respectively. Sets of discrete values $\mathcal{Y}$, $\mathcal{X}$ and $\mathcal{Z}$ have cardinality M, K and R. We define them as follows.

$$\mathcal{Y} \equiv \{y_m\}_{m=1}^{M} \qquad \mathcal{X} \equiv \{x_k\}_{k=1}^{K} \qquad \mathcal{Z} \equiv \{z_r\}_{r=1}^{R} \qquad \mathcal{U} \equiv \mathbb{R}^L$$

Data follows a probability law, Pr_0, represented by the conditional distribution $F^0_{YX|Z}$.

Our model falls into the class of *generalized instrumental variable models* (GIV) studied by Chesher, Rosen, Smolinski (2011). GIV models incorporate models where Y and X can be both continuous or discrete. GIV models also allow for the structural h to be a correspondence and therefore incorporates a possibility of delivering incomplete designs. Chesher, Rosen, Smolinski (2011) shows that these models are set, rather than point, identifying in general and discusses identification of the structural h together with a distribution of the unobservables, F_U, that is a duple (h, F_U). They derive a set of moment inequalities characterizing the identified sets in the class of GIV models. We build on these results.

4.1.2 Elemental level sets

Elemental level set represents a set of values of unobservable U such that for given value of $x_k \in \mathcal{X}$, function h delivers the outcome $y_m \in \mathcal{Y}$ as follows.

$$\tau_{m,k}(h) = \{u : \quad y_m = h(x_k, u)\} \tag{4.2}$$

Elemental level sets establish a relationship between sets of values of latent heterogeneity and observable outcomes conditional on the value of the covariate. Notice that in general they need not be closed or connected. The structural function, h, determines their shape. Elemental level sets pivot our formulation of identification and constitute the basis of coming developments of the core determining indexes, sets and inequalities.

We define collections of elemental level sets.

$$\tau_k(h) = \{\tau_{m,k}(h)\}_{y_m \in \mathcal{Y}} \qquad \tau_m(h) = \{\tau_{m,k}(h)\}_{x_k \in \mathcal{X}} \qquad \tau(h) = \{\tau_{m,k}(h)\}_{y_m \in \mathcal{Y}, x_k \in \mathcal{X}}$$

The list $\tau_k(h)$ assembles all elemental level sets corresponding to a particular value of the

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

covariate, x_k. Similarly, $\tau_m(h)$ collects elemental level sets for a given value of the outcome y_m. Finally, the list $\tau(h)$ gathers all elemental level sets in the model.

The essential observation follows. For every value of $x_k \in \mathcal{X}$ pairwise exclusive sets arise in a collection $\tau_k(h)$. This collection partitions the domain of unobservable $\mathcal{U}$. This remark underlines and supports a constructive development of the algorithm in the paper. We state it formally in the following lemma.

Lemma. *For every value of the covariate, $x_k \in \mathcal{X}$, elemental level sets $\tau_{m,k}(h)$ in the list $\tau_k(h)$ partition the space of unobservables, $\mathcal{U}$, according to the support of the outcome, $\mathcal{Y}$.*

Proof. Consider some x_k in $\mathcal{X}$. Consider elemental level sets $\tau_{y,x}(h)$ and $\tau_{y',x}(h)$ of $\mathcal{U}$ that correspond to outcomes y_m and $y_{m'}$, both in $\mathcal{Y}$. Since h is a function then $\tau_{m,k}(h) \cap \tau_{m',k}(h) = \emptyset$ if and only if $y_m \neq y_{m'}$. $\square$

4.1.3 Identified set

Chesher, Rosen, Smolinski (2011) shows that the model set identifies the structural function, h, and the distribution of unobservables, F_U. Following these results, let $\mathcal{S}$ be a collection of all subsets of $\mathcal{U}$. We characterize the identified set $\mathcal{A}$ as a set of structural functions and distributions of unobservables, that is a set of duples (h, F_U), such that for any set S in $\mathcal{S}$ the following inequalities must hold.

$$Q_S(h, F_U): \qquad \int_S dF_U \geq \max_{z_r \in \mathcal{Z}} \sum_{y_m \in \mathcal{Y}, x_k \in \mathcal{X}} \mathbf{1}(\tau_{m,k}(h) \subseteq S) Pr_0(Y = y_m, X = x_k | Z = z_r) \tag{4.3}$$

The left hand side denotes the probability mass allocated to the set S computed with respect to the probability law F_U. The right hand side denotes a probability mass, taken with respect to the distribution of the data, $F^0_{YX|Z}$, aggregated over all elemental level sets, $\tau_{m,k}(h)$, that are subsets of the proposed set S. Inequalities must hold for all values z_r in the support of the instrument, $\mathcal{Z}$.[1]

By specifying the distribution of U in the model we restrict the left hand side of the inequality. Also by shaping the structural function, h, we form elemental level sets and

[1]Inequality (4.3) has its equivalent representation in the language of *random sets theory.* Let $\mathcal{T}(Y, X; h)$ be a random set defined on the probability space over Y, X, Z and U as follows.

$$\mathcal{T}(Y, X; h) \equiv \{u : \quad Y = h(X, u)\}$$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

therefore interfere with the right hand side of the inequality.

4.2 The Core

We follow the results presented in Chesher, Rosen, Smolinski (2011), which shows that the identified set in the multinomial choice model can be represented by a finite number of inequalities.

4.2.1 Tightening bounds

Recall that $\mathcal{S}$ is a collection of all subsets in $\mathcal{U}$. We introduce the following notation. For any $S \in \mathcal{S}$ we define $\tau^S(h)$ to be a collection of all elemental level sets that are subsets of S. Similarly, we define $\mathcal{J}^S(h)$ as a collection of indexes of those elemental level sets that are subsets of S.

$$
\begin{aligned}
\tau^S(h) &\equiv \{\tau_{m,k}(h) : \quad \text{such that } \tau_{m,k}(h) \subset S\} \\
\mathcal{J}^S(h) &\equiv \{(m,k) : \quad \text{such that } \tau_{m,k}(h) \subset S\}
\end{aligned}
$$

We define a *choice function*, μ, which transforms the collection of indexes into a set that is a union of all elemental level sets indexed by the input collection.

$$
\mu(\mathcal{J}^S(h)) \equiv \left\{ u : \quad \bigcup_{(m,k)\in\mathcal{J}^S(h)} \tau_{m,k}(h) \right\}
$$

The probability distribution of random sets can be charactarized by containment functionals. The containment functional of the random set $\mathcal{T}(Y,X;h)$ for every value of z in the support of Z follows.

$$
Pr_0(\mathcal{T}(Y,X;h) \subseteq S|Z=z) = \sum_{x\in\mathcal{X},y\in\mathcal{Y}} \mathbf{1}(\tau_{yx}(h) \subseteq S) Pr_0(Y=y, X=x|Z=z)
$$

This is precisely the right hand side of (4.3). Therefore the identified set $\mathcal{A}$ of admissible duples (h, F_U) can be characterized by the following set of inequalities.

$$
\forall S \in \mathcal{U} : \qquad \int_S dF_U \geq \max_{z\in\mathcal{Z}} Pr_0(\mathcal{T}(Y,X;h) \subseteq S|Z=z)
$$

As before, the left hand side takes a mass allocated to some set S in a support of unobservables. This mass is compared with a probability delivered by a containment functional of a random set $\mathcal{T}(Y,X;h)$ that is a subset of some set S and maximized over all values of the instrument. Chesher, Rosen, Smolinski (2011) or Bernstenau, Molchanov, Molinarii (2010) use the language of random set theory presented by Molchanov (2005).

Let $\mathcal{C}^S(h)$ be a set delivered by the choice function applied to the collection of indexes $\mathcal{J}^S(h)$. Embodied within these definitions we characterize the identified set in the following proposition.

Proposition 1. *For any set S in $\mathcal{S}$ the following inequalities characterize the identifed set of (h, F_U).*

$$Q_{\mathcal{C}^S(h)}(h, F_U): \qquad \int_{\mathcal{C}^S(h)} dF_U \geq \max_{z_r \in \mathcal{Z}} \sum_{(m,k) \in \mathcal{J}^S(h)} Pr_0(Y = y_m, X = x_k | Z = z_r) \tag{4.4}$$

Proof. We show that these inequalities are implied by the inequalities derived by Chesher, Rosen, Smolinski (2011). Let $\tilde{\mathcal{S}}$ be a subset of $\mathcal{S}$ that collects all sets satisfying the following restriction.

$$\tilde{\mathcal{S}}: \quad \forall_{S', S'' \in \tilde{\mathcal{S}}} \qquad \mathcal{J}^{S'}(h) = \mathcal{J}^{S''}(h) \tag{4.5}$$

All sets in $\tilde{\mathcal{S}}$ generate the same indexes of the elemental levels sets. By applying the definition of $\mathcal{J}^S(h)$ and using the choice function, μ, on the right hand side of the characterization of the identified set in (4.3), we get the right hand side of (4.4). The right hand side takes the same value for any set $S \in \tilde{\mathcal{S}}$. We denote this constant by $\rho(\tilde{\mathcal{S}})$.

The inequality (4.3) holds for all $S \in \tilde{\mathcal{S}}$ and in particular, it must hold for $\mathcal{C}^S(h)$. By construction, the set $\mathcal{C}^S(h)$ is a subset of all sets S in $\tilde{\mathcal{S}}$ and is the set with the smallest volume in $\tilde{\mathcal{S}}$. Since F_U is a distribution function, it follows that the integral on the left hand side is a monotone function of S, i.e. whenever $S' \subseteq S''$ then $\int_{S'} dF \leq \int_{S''} dF$. Hence the following holds.

$$\forall_{S \in \tilde{\mathcal{S}}} \quad \mathcal{C}^S(h) \subseteq S \ \text{ hence } \int_{\mathcal{C}^S(h)} dF = \min_{S \in \tilde{\mathcal{S}}} \int_S dF \geq \rho(\tilde{\mathcal{S}})$$

And it must hold for all subset $\tilde{\mathcal{S}}$ of $\mathcal{S}$ with property (4.5). □

Proposition 1 specifies inequalities that must hold for all subsets $S \in \mathcal{S}$, which in general is an uncomfortably vast space of sets and is possibly unfeasible to deal with in practice. They may lead to an infinite number of inequalities characterizing the identified set, as in (4.3).

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

When observables are discrete then the power set $\mathcal{T}(h)$ of all elemental level sets $\tau(h)$ has 2^{MK} elements. Therefore, all $S \in \mathcal{S}$ deliver at most 2^{MK} distinct collections $\mathcal{J}^S(h)$ and the same maximal number of distinct sets $\mathcal{C}^S(h)$. However, the maximal number of distinct $\mathcal{J}^S(h)$ or $\mathcal{C}^S(h)$ may often be much smaller then the size of the power set, $\mathcal{T}(h)$. This may happen when some elemental level sets are subsets of others, which is induced by the shape of the structural function and the support of unobservable heterogeneity.

A finite number of sets leads to a finite number of inequalities characterizing the identified set. However, the number of inequalities may still be large because it is induced by the power set of the number of all points of support of the observables in the model. Shrinkage can occur when the set $C^S(h)$ can be split into two disjoint sets, say $C^{S'}(h)$ and $C^{S''}(h)$. Then the inequality defined by the set S can be induced from two inequalities defined by sets S' and S''. The following corollary formalizes this observation.

Corollary 1. *Inequality $Q_{\mathcal{C}^S(h)}(h, F_U)$ is induced by inequalities $Q_{\mathcal{C}^{S'}(h)}(h, F_U)$ and $Q_{\mathcal{C}^{S''}(h)}(h, F_U)$ if following conditions hold.*

$$(i) \quad \mathcal{J}^S(h) = \mathcal{J}^{S'}(h) \cup \mathcal{J}^{S''}(h) \quad \textit{and} \quad (ii) \quad \mathcal{J}^{S'}(h) \cap \mathcal{J}^{S''}(h) = \emptyset$$

Proof. Suppose conditions (i) and (ii) hold. They imply the same conditions on $\mathcal{C}^S(h)$, $\mathcal{C}^{S\prime}(h)$ and $\mathcal{C}^{S''}(h)$. The integral on the left hand side of (4.4) is a linear operator. Therefore, it can be split into a sum of two integrals over disjoint sets induced by $S'$and $S''$of everywhere possitive measure F_U. For every value of z_r, the right hand side of (4.4) splits into two sums over all elements in $\mathcal{J}^{S'}(h)$ and $\mathcal{J}^{S''}(h)$ respectively. Hence for every value of the instrument, inequality (4.4) induces two inequalities for S' and S''.

$$\begin{array}{llll} S': & \displaystyle\int_{\mathcal{C}^{S'}(h)} dF_U & \geq & \displaystyle\sum_{(m,k)\in\mathcal{J}^{S'}(h)} Pr_0(Y = y_m, X = x_k | Z = z_r) \\ S'': & \displaystyle\int_{\mathcal{C}^{S''}(h)} dF_U & \geq & \displaystyle\sum_{(m,k)\in\mathcal{J}^{S''}(h)} Pr_0(Y = y_m, X = x_k | Z = z_r) \end{array}$$

If they both hold for all $z_r \in \mathcal{Z}$ then the inequality generated by S must hold by the property of max operator and the fact that F_U and Pr_0 and non-negative everywhere. $\square$

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

4.2.2 Determinig the core

We define the *core determining partition* of $\mathcal{S}$. Let $\mathcal{P}$ be index set $\{1, ..., P\}$. A collection of P subsets of $\mathcal{S}$ indexed by p, $\mathcal{S}_p$, we call the core determining partitions of $\mathcal{S}$ if the following conditions hold. For any subset $S_p \in \mathcal{S}_p(h)$ and any subset $S_q \in \mathcal{S}_q(h)$:

(a) $\mathcal{J}^{S_p}(h) = \mathcal{J}^{S_q}(h)$ if and only if $p = q$,

(b) For any $S \in \mathcal{S}_p(h)$ Corollary 2 applies only to a set S itself and the empty set $\emptyset$.

Condition (a) states that all sets in the p^{th} element of the core determining partition, $\mathcal{S}_p(h)$, generate the same collection of indexes. Condition (b) asserts that the inequality (4.4) produced by any subset S of the core determining partition $\mathcal{S}_p(h)$ cannot be derived from inequalities induced by two non-empty and distinct sets in $\mathcal{S}_p(h)$.

A list of indexes induced by the element of the core determining partitions, $\mathcal{S}_p(h)$, we call *core determining indexes*, $\mathcal{J}_p(h)$. The set generated by the choice function on $\mathcal{J}_p(h)$ core determining indexes we call the *core determining set*, $\mathcal{C}_p(h)$. For every $p \in \mathcal{P}$ element of the core determining partition, $\mathcal{S}_p(h)$, is defined as follows.

$$\forall_{S \in \mathcal{S}_p(h)} \qquad \mathcal{J}_p(h) \equiv \mathcal{J}^S(h) \quad \text{and} \quad \mathcal{C}_p(h) \equiv \mathcal{C}^S(h)$$

We formulate the characterization of the identified set.

Proposition 1. *For every p in the index set, $\mathcal{P}$, core determining indexes, $\mathcal{J}_p(h)$, and sets, $\mathcal{C}_p(h)$, induce core determining inequalities $Q_p(h, F_U)$ as follows.*

$$Q_p(h, F_U) : \qquad \int_{\mathcal{C}_p(h)} dF_U \geq \max_{z_r \in \mathcal{Z}} \sum_{(m,k) \in \mathcal{J}_p(h)} Pr_0(Y = y_m, X = x_k | Z = z_r) \tag{4.6}$$

A collection of all P core determining inequalities, $\mathcal{Q}(h, F_U)$, characterizes the set of structural functions and distribution of unobservables, the set of admissible duples (h, F_U). That is the identified set $\mathcal{A}$.

Proof. The identified set $\mathcal{A}$ is characterized by (4.4) that must hold for all $S \in \mathcal{S}$. The result follows directly by definition of core determining partitions and Corollary 2 applied to $\mathcal{C}_p(h)$. □

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

This result comes obviously from the construction of the core determining indexes. The core determining set is connected in the sense that it is impossible to partition it into two non-empty subsets composed of distinct elemental level sets. However, the core determining set can be a disconnected set because elemental level sets it contains need not be connected.

We derive the collection of core determining partitions of $\mathcal{S}$ that induce a finite collection of subsets of $\mathcal{S}$, which are necessary and sufficient for characterization of the identified set. This collection of core determining sets together with a collection of corresponding core determining indexes delivers a collection of the core determining inequalities, i.e. a set of ultimate identification questions.

$$\begin{aligned} \mathcal{S}(h) &= \{\mathcal{S}_p(h)\}_{p=1}^{P} & \mathcal{J}(h) &= \{\mathcal{J}_p(h)\}_{p=1}^{P} \\ \mathcal{C}(h) &= \{C_p(h)\}_{p=1}^{P} & \mathcal{Q}(h, F_U) &= \{Q_p(h, F_U)\}_{p=1}^{P} \end{aligned}$$

All collections have the same cardinality P. Cardinality depends on the complexity of the problem driven by a structural function, h, dimensionally of latent heterogeneity U or sparsity of the data. Apart from very specific designs we can not determine P analytically. However, in what follows, we propose a numerical construction delivering collections $\mathcal{J}(h)$ and $\mathcal{C}(h)$.

4.3 Algorithm

This section presents a constructive algorithm, which generates core determining indexes for the single equation instrumental variable model when observables possess discrete supports. The algorithm involves operations on subsets of the support of unobservables. In particular, it requires predefined routines that perform elemental algebraic operations on sets, i.e. takes unions, intersections or verifies if one set is a subset of the other. We assume that these operations are available and can be incorporated as functions into the algorithm.

We begin by establishing notation. We outline the set up in fine detaill. We define essential objects and functions and we briefly describe the evolution of the algorithm. Finally, we discuss complexity and potential improvements of the construction proposed.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

4.3.1 Notation

At the outset, recall that the structural function, h, links a K valued covariate X and latent vector U with the M valued outcome Y. The model delivers a list of elemental level sets, $\tau(h)$, for the specified function h. There are $P = M \times K$ distinct elemental level sets $\tau_{m,k}(h)$ in that list. For every value x_k we aggregate elemental level sets in rank lists, $\tau_k(h)$, of length M that partition the support of unobservables. We organize the list of elemental level sets, $\tau(h)$, as a collection of K consecutive rank lists. In what follows, we utilize this fact in the information matrix to simplify the algorithm and enhance construction of the core determining indexes.

Throughout, we use curly parentheses interchangeably in two ways. On the one hand as a collection or list of elements. On the other, as a function that uniquely concatenates elements from different lists. Therefore whenever applied, curly parenthesis stand for a list of distinct components.

Depth and height. The algorithm evolves in two dimensions dependently, which we call the *step* and *height* of the algorithm. Step appears outer with respect to height in a sense that the number of height levels, represented by d, changes in every step of the algorithm, indexed by q. We consider height to be superior with respect to step because climbing the highest level terminates the algorithm. We express it in an alternating list of height level paths, $\mathcal{D}_q$. We define the height level paths after an introduction of an information matrix.

Level Sets Indicators and Development Sets. A pair (m, k) uniquely indexes every elemental level set $\tau_{m,k}(h)$. The number m indicates the value of the outcome and k indicates the value of the covariate. Given the structural function, h, there are $L = M \cdot K$ distinct pairs of this type since there are L distinct elemental level sets in the model. In the initial step of the algorithm, denoted by zero, all the (m, k) pairs compose the initial list of length L of elemental *level set indicators,* $\mathcal{J}^0(h)$. The list delivers a corresponding initial list of *development sets,* $\mathcal{C}^0(h)$, i.e. a list equivalent to $\tau(h)$ at step zero. As the algorithm evolves, the list of level set indicators, $\mathcal{J}^q(h)$, gets updated and expands by one component in every step q of the construction. At the same instant the list of development sets, $\mathcal{C}^q(h)$, evolves accordingly. The choice function, μ, establishes correspondence between level set indicators

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

and development sets.

$$\mathcal{C}^q(h) = \{C_p^q(h)\}_{p=1}^{L+q} \qquad \mathcal{J}^q(h) = \{\mathcal{J}_p^q(h)\}_{p=1}^{L+q} \qquad \text{where} \quad C_p^q(h) = \mu\left(\mathcal{J}_p^{L+q}(h)\right)$$

Inheritance Indexes and Link Function. We employ a list of *inheritance indexes*, $\mathcal{N}^q$, to encode the content of the level set indicators, $\mathcal{J}_p^q(h)$, with respect to the remaining elements in the collection, $\mathcal{J}^q(h)$, while the steps of the algorithm alter. We define the initial list of inheritance indexes as follows.

$$\mathcal{N}^0 \equiv \{\{1\}, ..., \{P\}\} \qquad \text{and} \qquad \mathcal{N}^q = \left\{\mathcal{N}_p^q\right\}_{p=1}^{P+q}$$

Inheritance indexes turn out to be very useful in practice. They trace the evolution of the development sets in $\mathcal{C}^q(h)$ obtained from level set indicators in $\mathcal{J}^q(h)$ in all preceding steps of the algorithm. This knowledge allows for the shortening of the height level path, $\mathcal{D}_q$, at almost every step of the procedure when the construction expands and a new set arises.

Further we introduce the *link set function*, λ, that is essential to our developments. It is a four valued function that verifies the relationship between two test sets, S and S'. The definition of the link function follows and the table enumerates outcomes that the function λ delivers.

$$\forall_{S,S' \in \mathcal{S}} \qquad \lambda : (S, S') \to \{\emptyset, \subset, \supset, \sigma\}$$

$\lambda(S, S')$	Relationship
$\emptyset$	$S \cap S' = \emptyset$
$\subset$	$S \subset S'$
$\supset$	$S \supset S'$
σ	$S \cap S' \neq \emptyset$ and neither $S \subset S'$ nor $S \supset S'$

Information Matrix. Finally we introduce a binary *information matrix,* $\mathbf{A}^q$, which encodes the information content of the algorithm at every step q. The information matrix is a square matrix in which the p^{th} row(column) has assigned the p^{th} development set from the list of all development sets $\mathcal{C}^q(h)$. Therefore the size of the information matrix at step q of the algorithm is $P_q \times P_q$ and corresponds to the number of elements in $\mathcal{C}^q(h)$[2]. Entries in the information matrix depict the knowledge, or lack of, concerning the relationship between development sets in $\mathcal{C}^q(h)$. As long as the relationship between the n^{th} and the b^{th} development sets in $\mathcal{C}^q(h)$ remains unknown, the (n, b) entry of the informatin matrix, $A^q_{n,b}$, stays at value one. It gets updated to zero when the outcome of the function λ applied to $C^q_n(h)$ and $C^q_b(h)$ is revealed.

Entries of the matrix $\mathbf{A}^q$ get updated from one (uncertainty) to zero (certainty) while climbing up the hight levels, d, of the construction. In subsequent steps of the algorithm, the information matrix $\mathbf{A}^q$ expands with new columns and rows that match the new items in the augmented list of level sets indicators, $\mathcal{J}^q(h)$. This process continues until uncertainty encoded in the information matrix has been dispelled, i.e. is all entries has been set to zero. The state of information matrix $\mathbf{A}^q$ governs both the height level, d, and step, q, of the construction and determines when the algorithm terminates.

The rows and columns of the initial information matrix, $\mathbf{A}^0$, correspond to the elements in the initial list of development sets, $\mathcal{C}^0(h)$, that is complete list of L elemental level sets collected in $\tau(h)$. As set out, this collection binds rank lists, $\tau_k(h)$, that incorporate prior knowledge from the model. We recall that, by definition, rank lists comprise pairwise disjoint sets partitioning space of unobservables for every value of the covariates, x_k. Therefore, the outcome of the link function λ applied to any two distinct elemental level sets within every rank list $\tau_k(h)$ is known a priori and delivers an empty set, $\emptyset$. Blocks of zeros encode this prior knowledge into the initial information matrix.

Let $\mathbf{1}_M$ be a square $M \times M$ matrix of ones and let $\mathbf{0}_M$ be a square $M \times M$ matrix of zeros. We initialize information matrix, $\mathbf{A}^0$, as a square matrix with $K \times K$ block elements as follows.

$$\mathbf{A}^0_{k,l} = \begin{cases} \mathbf{1}_M & \text{if } k < l \\ \mathbf{0}_M & \text{otherwise} \end{cases}$$

The initial infirmation matrix is an upper triangular matrix composed of square blocks $\mathbf{1}_M$.

Height Levels Path. Let $\mathcal{D}^q$ be a *hight levels path* that collects positions of all ones in the information matrix $\mathbf{A}^q$. Let $A^q_{n,b}$ be a bit of information; the element of the information

[2] where $P_q = L + q$.

matrix $\mathbf{A}^q$ corresponding to the n^{th} row and the b^{th} column at the q^{th} step of the algorithm. We define a height levels path, $\mathcal{D}^q$, at the q^{th} step as follows.

$$\mathcal{D}^q \equiv \{(n,b): \qquad \text{for which } A^q_{n,b} = 1\}$$

The number of ones in the initial information matrix, $\mathbf{A}^0$, is equal to the sum of dimensions of all its $\mathbf{1}_M$ blocks, that is $D_0 = MK \times M(K-1)/2$. That being so, there are D_0 height levels to climb in the first step of the algorithm or rather, D_0 relations to be inquired from the link function λ. The number of inquires evolves progressively in succeeding steps indexed by q.

4.3.2 Construction

The algorithm has two major parts, (i) the initialization chunk and (ii) the learning loop. In the former we initialize basic objects according to the structure of the model and definitions outlined. In the latter, the algorithm finds core determining indexes by exploring the information matrix and climbing up the height levels path. The Algorithm frame presents these two parts schematically.

Initialization. For the specified model and proposed structural function, h, we begin by setting up basic objects as defined above. These are the information matrix $\mathbf{A}^0$, the list of level set indicators, $\mathcal{J}^0(h)$, inheritance indexes, $\mathcal{N}^0$, and initial height levels path, $\mathcal{D}^0$. We set out a list of development sets, $\mathcal{C}^0(h)$, by applying the choice function μ, to a list of indexes from $\mathcal{J}^0(h)$.

Henceforth, we ask for a predefined class of set objects with an empty and full sets as special elements of the class. This class must represent sets in the domain $\mathcal{U}$ of latent heterogeneity U. We also require predefined elemental methods conducting operations on subsets of $\mathcal{U}$. These are union and intersection. Lastly we define link function λ. Both step, q, and height level, d, are initialized with zeros. The learning loop follows.

Learning Loop. In the learning part, the algorithm progressively updates step, q, until it reaches the top of the height levels path, $\mathcal{D}^q$. A loop command `Until` in the algorithm frame expresses this process. The height index, d, controls the loop cycle and its value points on an *active loop*. In every round it increments by one, $d \leftarrow d+1$, until it reaches the top of the height levels, D_q. The height index, d, indicates element on the height levels path $\mathcal{D}^q$ and hence, it provides pointers n and b. These pointers indicate elements in the list of level sets indicators, $\mathcal{J}_n^q$ and $\mathcal{J}_b^q$. We call them *active level sets indicators*[3]. Next the algorithm applies the choice function, μ, to the active level sets indicators and finds related development sets, $\mathcal{C}_n^q$ and $\mathcal{C}_b^q$, called *active development sets* hereafter.

Active development sets evaluated with the link function λ deliver the link indicator $\mathcal{L}_d^q$. Depending on the value of this indicator one of four actions may happen.

If active development sets are mutually exclusive then link function returns an empty set, with link indicator set to $\emptyset$. It means there is no common area in the active development sets and the algorithm remains actionless on level set indicators.

When one of active development sets is a subset of the other then the link indicator, $\mathcal{L}_d^q$, takes one of two values, either $\subset$ or $\supset$. In both cases we consider the superset to be *parental* with respect to the subset. Suppose that $\mathcal{C}_n^q$ is a subset of $\mathcal{C}_b^q$, in which case $\mathcal{C}_b^q$ is parental for $\mathcal{C}_n^q$. Then level sets indicators, $\mathcal{J}_n^q$, of the subset development set, $\mathcal{C}_n^q$, are incorporated into the parental level set indicators, $\mathcal{J}_b^q$, i.e. they are merged together into an updated parental active level sets indicator, $\mathcal{J}_b^q \leftarrow \{\mathcal{J}_n^q, \mathcal{J}_b^q\}$.

Finally, if the link indicator, $\mathcal{L}_d^q$, delivers a value σ then the active development sets have a common area and can be possibly married. The algorithm concatenates active level sets indicators into the temporal level sets indicator, $\mathcal{I}_{tmp} \leftarrow \{\mathcal{J}_n^q, \mathcal{J}_b^q\}$, with the corresponding temporal development set, $\mathcal{C}_{tmp} \leftarrow \mu(\mathcal{I}_{tmp})$. Inheritance indexes get updated $\mathcal{N}_{tmp} \leftarrow \{\mathcal{N}_n^q, \mathcal{N}_b^q, n, b\}$ tracing active level sets.

If the temporal development set, $\mathcal{C}_{tmp}$, has its twin in the list of development sets, $\mathcal{C}^q(h)$, say the p^{th} element, then the algorithm merges their corresponding level sets indexes $\mathcal{J}_p^{q+1} \leftarrow \{\mathcal{J}_p^q, \mathcal{I}_{tmp}\}$ and $\mathcal{N}_p^{q+1} \leftarrow \{\mathcal{N}_p^q, \mathcal{N}_{tmp}\}$. However, if the temporal development set, $\mathcal{C}_{tmp}$, does not have a twin in the list of development sets $\mathcal{C}^q(h)$, then the link remains and the list of level set indicators expands, $\mathcal{J}^{q+1} \leftarrow \{\mathcal{J}^q, \mathcal{I}_{tmp}\}$. This implies an expansion of the information matrix $\mathbf{A}^q$ by additional row of zeros and column of ones with zeros at positions indicated by inheritance indexes $\mathcal{N}_{tmp}$. The algorithm defines new height levels path, $\mathcal{D}^{q+1}$, related to the

[3]We drop dependence on the structural function, h, in the notation for the active level sets indicators, $\mathcal{J}_n^q$, and the active development sets, $\mathcal{C}_n^q$.

 DOI: 10.2870/35896

updated information matrix $\mathbf{A}^{q+1}$. This way we move to the next step of the construction, $q \leftarrow q+1$, and start walking height levels path from the beginning again, $d \leftarrow 0$.

Algorithm 4.1 Development of the core determining indexes.

`INITIALIZE:`

`Define` information matrix $\mathbf{A}^0$, level sets indicators $\mathcal{J}^0$ and inheritance indexes $\mathcal{N}^0$.

`Create` height levels path $\mathcal{D}^0$ of length D_0 from unit elements of the information matrix, $\mathbf{A}^0$.

`Set` step $q = 0$ and height $d = 0$.

`LEARN:`

`Until` $d < D_q$ proceed:

`(1)` `Update` $d \leftarrow d + 1$

`(2)` `Set` indexes, n and b, corresponding to the d^{th} element of $\mathcal{D}^q$.

`(3)` `Evaluate` $\mathcal{L}_d^q \leftarrow \lambda(\mu(\mathcal{J}_n^q), \mu(\mathcal{J}_b^q))$ and `Set` $A_{n,b}^q \leftarrow 0$. Check `IF`:

$\mathcal{L}_d^q = \emptyset$ then do nothing
$\mathcal{L}_d^q = \subset$ then `Update` indexes, $\mathcal{J}_b^q \leftarrow \{\mathcal{J}_n^q, \mathcal{J}_b^q\}$
$\mathcal{L}_d^q = \supset$ then `Update` indexes, $\mathcal{J}_n^q \leftarrow \{\mathcal{J}_n^q, \mathcal{J}_b^q\}$
$\mathcal{L}_d^q = \sigma$ then
`Set` $\mathcal{I}_{tmp} \leftarrow \{\mathcal{J}_n^q, \mathcal{J}_b^q\}$ and $\mathcal{C}_{tmp} \leftarrow c(\mathcal{I}_{tmp})$ and $\mathcal{N}_{tmp} \leftarrow \{\mathcal{N}_n^q, \mathcal{N}_b^q, n, b\}$
`If:` $\mathcal{C}_{tmp}$ is equal to $c(\mathcal{J}_p^q)$, for any p, then `Update` indexes $\mathcal{J}_p^q \leftarrow \mathcal{I}_{tmp}$
`Else:` do the following:
`Update` the list of indexes $\mathcal{J}^{q+1} \leftarrow \{\mathcal{J}^q, \mathcal{I}_{tmp}\}$ and $\mathcal{N}^{q+1} \leftarrow \{\mathcal{N}^q, \mathcal{N}_{tmp}\}$
`Define` matrix $\mathbf{A}^{q+1}$ as $\mathbf{A}^q$ extended by a row of zeros and a column of ones
`Update` the last column of $\mathbf{A}^{q+1}$ with zeros at positions in $\mathcal{N}_{tmp}$.
`Create` height levels path $\mathcal{D}^{q+1}$ of length D_{q+1} from unit elements of the information matrix, $\mathbf{A}^{q+1}$.
`Set` step $q \leftarrow q + 1$ and `Set` heigth $d \leftarrow 0$

4.3.3 Concluding remarks on the algorithm

The algorithm develops a list of core determining indexes, $\mathcal{J}(h)$. It is constructive in a sense that it learns about the structure of the model progressively. It expands the initial list of core determining indexes and always delivers the core in finite number of steps.

We find it very helpful to use the information matrix in the construction. It simplifies the algorithm and sustains all the information necessary. Also inheritance indexes improve the efficient use of information by the procedure. By tracing the history and the content of level sets indicators the algorithm avoids double checks.

The information matrix brings some limitations to the storage of the data and perhaps, to search speed. We conjecture that by substituting the search over the information matrix by some form of evolution trees or other, more sophisticated, search methods could enhance speed of the procedure.

4.4 Illustration

Generalized instrumental variable models, to which this work refers, comprise of an immensely rich class of set identified models. When dealing with discrete data, the identified sets can be characterized by the inequalities (4.6). The algorithm presented in the paper delivers core determining indexes and sets.

Identification of complicated models with high dimensional heterogeneity requires analysis of sets placed in high dimensional spaces. Numerical representation of any subsets in these spaces leads to considerable practical twists. Therefore, dimensionality of the unobserved heterogeneity and richness of functional space of the structural h and distribution F_U is of significant practical concern and can substantially influence the complexity of the identification analysis.

However, to demonstrate methodological insights we find the ordered outcome instrumental variable model remarkably instructive. Its univariate nature of the error term brings enormous simplification and clarity to the picture. At the same time, the model comprises all the features of the method delivering core determining indexes, yet allows for simplicity of demonstration.

This model has been studied by a number of authors. Chernozhukov and Hansen (2005) provides identification results and propose estimation for the set up with continuous outcome. Chesher (2010) shows that the model set identifies the structural function h and derives an outer set that comprises the identified set. Chesher and Smolinski (2009) discusses inequalities necessary for the characterization of the identified set when the outcome is discrete and the explanatory variable is binary. Chesher and Smolinski (2010) characterizes the identified set under discrete variation of the outcome and the explanatory covariates.

This section presents an illustrative model in the notation proposed in the previous section. We apply the algorithm and discuss its steps. We present core determining sets and indexes.

4.4.1 Ordered Outcome IV Models

We consider a class of ordered outcome instrumental variable models. In this class, the unknown function h of the observable X and univariate, latent variable U determines the outcome variable Y as follows.

$$Y = h(X, U) \qquad U \perp Z \qquad U \sim Unif[0, 1]$$

An instrumental variable restriction, $U \perp Z$, excludes dependence between unobserved heterogeneity, aggregated in U, and the instrumental variable, Z. It must hold for all values in the support of the instrument, $Z \in \mathcal{Z}$. We normalize the support of the latent variable to a unit interval and restrict its distribution to uniform. The discrete outcome takes ordered values indexed by integeres, $m \in \mathcal{Y}$. The covariate X has K points of support indexed by k, $x_k \in \mathcal{X}$.

Subsequently, we pursue elementary illustration where a binary X explains three valued outcome Y with supports defined as follows.

$$\mathcal{Y} \equiv \{1, 2, 3\} \qquad \mathcal{X} \equiv \{x_1, x_2\}$$

The model imposes a weak monotonicity restriction on the structural h in its second argument, U. This restriction implies a threshold crossing representation. Let u take some value in $(0, 1]$. Then for all $m \in \{1, 2, 3\}$ and $k \in \{1, 2\}$ the threshold crossing is written as

$$m = h(x_k, u) \qquad \text{if} \qquad h_{m-1,k} < u \leq h_{m,k} \tag{4.7}$$

where the structural parameter $h_{m,k}$ is an abbreviation for the m^{th} threshold function evaluated at point x_k, that is $h_m(x_k)$. For convenience, we use fixed parameters $h_{0,1}$, $h_{0,2}$ and $h_{3,1}, h_{3,2}$ to denote respectively the lower and upper limits of the support $\mathcal{U}$.

There are four identifiable parameters in this model, two for each value of X. Weak monotonicity imposes inequality restrictions on these parameters. There is one inequality on

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

parameters $h_{1,1}, h_{2,1}$ corresponding to x_1 and analogues inequality imposed on parameters $h_{1,2}, h_{2,2}$ corresponding to x_2.

$$\{h_{1,1}, h_{2,1}, h_{1,2}, h_{2,2}\} \quad \text{with} \quad h_{1,1} < h_{2,1} \quad \text{and} \quad h_{1,2} < h_{2,2} \tag{4.8}$$

Elemental level sets

There are six elemental level sets. They divide the support of unobservables, a unit interval, into three disjoint subintervals for x_1 and x_2 and deliver corresponding rank lists, $\tau_1(h)$ and $\tau_2(h)$. Division into disjoint sets is a consequence of the weak monotonicity restriction in this model. Elemental level sets follow.

$$\begin{array}{lcllclcll}
\tau_{1,1}(h) & = & \{u: & u \in [h_{0,1}, h_{1,1}]\} & , & \tau_{1,2}(h) & = & \{u: & u \in [h_{0,2}, h_{1,2}]\} \\
\tau_{2,1}(h) & = & \{u: & u \in (h_{1,1}, h_{2,1}]\} & , & \tau_{2,2}(h) & = & \{u: & u \in (h_{1,2}, h_{2,2}]\} \\
\tau_{3,1}(h) & = & \{u: & u \in (h_{2,1}, h_{3,1}]\} & , & \tau_{3,2}(h) & = & \{u: & u \in (h_{2,2}, h_{3,2}]\}
\end{array}$$

The left and right columns list elemental level sets that correspond to rank lists $\tau_1(h)$ and $\tau_2(h)$ respectively. Let u take some value in $[0,1]$. Then threshold crossing representation in a language of elemental level sets follows.

$$h(x_1, u) = \begin{cases} 1, & \text{if} \quad u \in \tau_{1,1}(h) \\ 2, & \text{if} \quad u \in \tau_{2,1}(h) \\ 3, & \text{if} \quad u \in \tau_{3,1}(h) \end{cases} \qquad \text{and} \qquad h(x_2, u) = \begin{cases} 1, & \text{if} \quad u \in \tau_{1,2}(h) \\ 2, & \text{if} \quad u \in \tau_{2,2}(h) \\ 3, & \text{if} \quad u \in \tau_{3,2}(h) \end{cases}$$

The Identified Set

In the ordered outcome model, the distribution F_U is restricted to be uniform on a unit interval, $[0,1]$. Any subset S of this interval is a set of values of u bounded by u_1 from above and u_2 from below. Hence, the left hand side of inequalities in (4.3) simplifies to a difference between the boundaries of the set S as follows.

$$\forall_{u_1, u_2 \in [0,1]} \text{ let } S = \{u: \quad u_1 < u \leq u_2\} \qquad \text{then} \quad \int_S dF_U = u_1 - u_2$$

When observables in the model demonstrate discrete variation then the identified set

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

can be characterized by a list of finite number of core determining inequalities, as in (4.6), driven by the core determining indexes and sets. In the ordered outcome IV model, the core determining set, say $C_p(h)$, translates to a subset of the unit interval bounded by the values of structural parameters, say $h_{n,l}$ and $h_{m,k}$. This fact together with restricted uniform distribution of unobservables simplify the left hand side of the p^{th} core determining inequality. Eventually, it becomes a difference in the limiting values of the structural parameters.

$$C_p(h) = \{u : \ h_{n,l} \leq u \leq h_{m,k}\} \ \text{ then } \int_{u \in C_p(h)} dF_U = h_{m,k} - h_{n,l}$$

Notice that $C_p(h)$ is a subset of S. The left hand side of any p^{th} core determining inequality is straightforward to compute provided we know the corresponding core determining set, $C_p(h)$. Since every core determining set has its list of core determining indexes, the right hand side of the core determining inequality computes easily.

Figure (4.1) depicts the intuition behind the characterization of the identified set in (4.3) and (4.6) applied to the ordered outcome IV model. The left pane presents the red set S bounded by values u_1 and u_2. This set contains two elemental level sets, $\tau_{2,1}(h)$ and $\tau_{2,2}(h)$, marked as light blue, vertical stripes. On the right pane, we illustrate different set S', painted in dark blue and bounded by $u_1 = h_{2,2}$ and $u_2 = h_{1,1}$, This set contains exactly the same elemental level sets as set S, namely $\tau_{2,1}(h)$ and $\tau_{2,2}(h)$. Therefore S' defines the core determining set, $C_p(h)$, corresponding to the core determining indexes $(\{2,1\},\{2,2\})$. The core determining indexes match both sets, S and $C_p(h)$.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

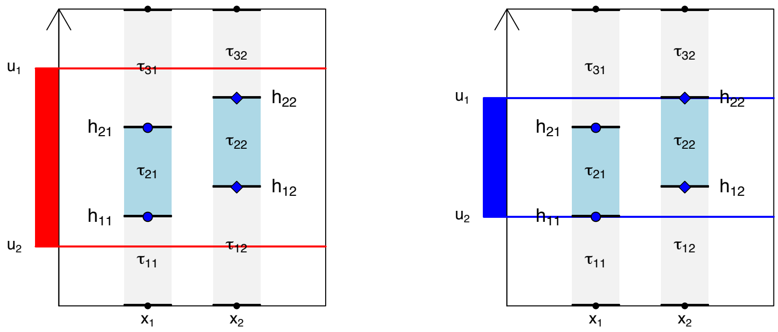


Figure 4.1: Both west and east panes present the same structural function h for the model with three valued outcome Y and a binary covariate X. Gray and blue vertical stripes display elemental level sets, $\tau_{m,k}(h)$. Black bars, with blue dots for x_1 and blue diamonds for x_2, denote the structural parameters, $h_{m,k}$. West pane shows the red set, $S = (u_2, u_1]$ covering two elemental level sets, $\tau_{2,1}(h)$ and $\tau_{2,2}(h)$. East pane displays related core determining set in blue, $C_p(h) = (h_{1,1}, h_{2,2}]$.

The next section demonstrates the construction of the core determining indexes and sets for the elementary example presented in Figure 4.1.

4.4.2 The algorithm in action

In the ordered outcome model with three outcomes and binary explanatory variable there are six arrangements of the structural parameters out of which three are substantively distinct. The remaining three are symmetric with respect to change of the second index of the structural parameter. We set them out as follows.

$$(a) \qquad h_{1,1} < h_{1,2} < h_{2,1} < h_{2,2}$$

$$(b) \qquad h_{1,2} < h_{1,1} < h_{2,1} < h_{2,2}$$

$$(c) \qquad h_{1,1} < h_{2,1} < h_{1,2} < h_{2,2}$$

We illustrate development of the core determining indexes for the arrangement (a). The core determining indexes and sets for arrangements (b) and (c) are graphed in the Appendix B.

Initialization part

The algorithm begins by initializing basic objects: a list of level sets indicators, inheritance indexes and information matrix. There are six elemental level sets in the model and they

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

compose the initial list of development sets, $\mathcal{C}^0(h)$. Related is the initial list of level sets indicators, $\mathcal{J}^0(h)$, consisting of (m,k) indexes corresponding to the initial list of development sets.

$$\begin{aligned} \mathcal{J}^0(h) \equiv \{\mathcal{J}_p^0(h)\}_{p=1}^6 \quad , \quad \mathcal{J}^0(h) &= \{(1,1),(2,1),(1,2),(2,2),(1,3),(2,3)\} \\ \mathcal{C}^0(h) \equiv \{C_p^0(h)\}_{p=1}^6 \quad , \quad \mathcal{C}^0(h) &= \{\tau_{1,1}(h),\tau_{2,1}(h),\tau_{3,1}(h),\tau_{1,2}(h),\tau_{2,2}(h),\tau_{3,2}(h)\} \end{aligned}$$

Notice that the first three development sets, $C_1^0(h)$, $C_2^0(h)$ and $C_3^0(h)$, partition the unit interval into mutually disjoint subsets. They form the rank list $\tau_1(h)$. The same applies to the rank list $\tau_2(h)$, that contains the last three development sets, $C_4^0(h)$, $C_5^0(h)$ and $C_6^0(h)$. The initial information matrix, $\mathbf{A}^0$, reflects this prior knowledge. It is initialized as a square 6×6 matrix that has six nonzero entries in the right-upper block of ones $\mathbf{1}_3$. Rows of this block corresponds to the rank list $\tau_1(h)$ and columns represent the rank list $\tau_2(h)$.

The initial height levels path, $\mathcal{D}^0$, picks indexes of the nonzero entries of the initial information matrix, $\mathbf{A}^0$. These are the following nine pairs representing height levels of the algorithm in the initial step.

$$\mathcal{D}^0 \equiv \{\mathcal{D}_d^0\}_{d=1}^9 \quad , \quad \mathcal{D}^0 = \{(1,4),(2,4),(3,4),(1,5),(2,5),(3,5),(1,6),(2,6),(3,6)\}$$

The list of level sets indicators is composed of single element lists indexed by p. The initial list of inheritance indexes reflects this fact. Every element of this list is a singleton as follows.

$$\mathcal{N}^0 = \{(1),(2),(3),(4),(5),(6)\}$$

Learning Loop

We present learning process in two complementary tables in Appendix A. Table 4.2 shows outcomes of the link function λ when the algorithm climbs the height levels path $\mathcal{D}^q$ (rows) for every steps q of the algorithm (columns). Table 4.3 presents progressive updates of the list of level sets indicators, $\mathcal{J}^q(h)$. Starting from the initial list, $\mathcal{J}^0(h)$, to the output list of the core determining indexes, $\mathcal{J}(h)$. Updates in Table 4.3 in six steps of the construction correspond to the values of the link indicator $\mathcal{L}_d^q$, different from the empty set, $\emptyset$.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Consider the first step of the construction, $q = 1$ and the first height level, $d = 1$. Height level sets path, $\mathcal{D}_1^1$ indicates on sets $(1, 4)$ for which the link function returns $\subset$. It implies that the elemental level set $\tau_{1,1}(h)$ corresponding to the first development set, $\mathcal{C}_1^0$, is a subset of the elemental level set $\tau_{1,2}(h)$ corresponding to the fourth development set, $\mathcal{C}_4^0$. Therefore the algorithm updates the fourth component on the list of level set indicators, $\mathcal{J}_4^0(h)$, with indexes corresponding to the first component of the list of level set indicators $\mathcal{J}_1^0(h)$ as follows.

$$
\begin{aligned}
\mathcal{J}_4^1(h) &\leftarrow \{\mathcal{J}_4^0(h), \mathcal{J}_1^0(h)\} = \{(1,1),(1,2)\} \\
C_4^1(h) &\leftarrow \mu\left(\mathcal{J}_4^1(h)\right) = \tau_{1,1}(h) \cup \tau_{1,2}(h)
\end{aligned}
$$

Indexes of the first and the fourth components of the list of initial level sets indicators get concatenated into a new list of level sets indicators, $\mathcal{J}_4^1(h)$. Also the fourth development set $C_4^0(h)$ gets updated as indicated in the second line.

The second and the third height levels, $d = 2, 3$, are silent about updates other then parts in the information matrix $\mathbf{A}^0$. However, at the fourth height level the link function returns a nonempty intersection, σ, between the second development set $C_2^0(h)$ and the newly updated fourth development set $C_4^1(h)$. This delivers new level sets indicator, $\mathcal{J}_7^1(h)$, extending the list of level sets indicators and moving the construction to the new step, $q = 2$. The seventh components of the level sets indexes and development sets follow.

$$
\begin{aligned}
\mathcal{J}_7^1(h) &\leftarrow \{\mathcal{J}_4^1(h), \mathcal{J}_2^0(h)\} = \{(1,1),(1,2),(2,2)\} \\
C_7^1(h) &\leftarrow \mu\left(\mathcal{J}_7^1(h)\right) = \tau_{1,1}(h) \cup \tau_{1,2}(h) \cup \tau_{2,2}(h)
\end{aligned}
$$

The inheritance indexes updates $\mathcal{N}_7^1 \leftarrow \{\mathcal{N}_2^0, \mathcal{N}_4^1\} = \{1, 4, 2\}$. The new set leads to extension of the information matrix by a row of zeros and a column of ones everywhere but positions indicated by the inheritance indexes. Lastly, the new path of height levels is defined as$\mathcal{D}^1$, with its first element $(2, 5)$. The step index is updated to $q = 2$ and d counter is reset.

In the second, third and fourth steps of the algorithm nonempty intersections appear at

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

the first, fourth and the second height levels respectively. New level sets indicators follow.

$$\begin{aligned}
\mathcal{J}_8^2(h) &= \left\{\mathcal{J}_2^1(h), \mathcal{J}_5^1(h)\right\} = \{(2,1),(2,2)\} \\
\mathcal{J}_9^3(h) &= \left\{\mathcal{J}_3^2(h), \mathcal{J}_5^2(h)\right\} = \{(3,1),(2,2)\} \\
\mathcal{J}_{10}^4(h) &= \left\{\mathcal{J}_2^3(h), \mathcal{J}_9^3(h)\right\} = \{(2,1),(3,1),(2,2)\}
\end{aligned}$$

Obviously the algorithm creates corresponding development sets, $C_8^2(h)$, $C_9^3(h)$ and $C_{10}^4(h)$. At the end of each of these steps, the information matrix extends by a row of zeros and column of ones reset to zeros at positions indicated by inheritance indexes.

$$\begin{aligned}
&q = 2 \text{ and } d = 1: \qquad &\mathcal{N}_8^2 &\equiv \left\{\mathcal{N}_2^1, \mathcal{J}_5^1, 8\right\} = \{2,5,8\} \\
&q = 3 \text{ and } d = 4: \qquad &\mathcal{N}_9^3 &\equiv \left\{\mathcal{N}_3^2, \mathcal{N}_5^2, 9\right\} = \{3,5,9\} \\
&q = 4 \text{ and } d = 2: \qquad &\mathcal{N}_{10}^4 &\equiv \left\{\mathcal{N}_2^3, \mathcal{N}_9^3, 10\right\} = \{2,3,5,9,10\}
\end{aligned}$$

The fifth step of the construction appears particularly interesting. The algorithm updates three times at height levels $d = 2, 4, 5$ before expansion takes place at the sixth level, $d = 6$. At the second level, the link indicator shows $\supset$ for indexes $(3,6)$. It means that the third development set is parental with respect to the sixth one. The level sets indicator $\mathcal{J}_3^5(h)$ is updated with the indexes in $\mathcal{J}_6^5(h)$. Also inheritance indexes updates, $\mathcal{N}_3^5 \leftarrow \{\mathcal{N}_3^5, \mathcal{N}_6^5\} = \{3,6\}$. At the fourth and fifth height level, the link indicator returns the value $\subset$ for indexes $(3,9)$ and $(3,10)$, meaning that the level sets indicators $\mathcal{J}_9^5(h)$ and $\mathcal{J}_{10}^5(h)$ get updated with the indexes in $\mathcal{J}_3^5(h)$. We emphasize that updates in steps four and five use level sets indicators $\mathcal{J}_3^5(h)$ updated at the second level.

The algorithm expands the information matrix at the sixth height level when the link indicator $\mathcal{L}_6^5$ delivers σ for indexes $(4,8)$. It leads to an eleventh component of level sets indicators $\mathcal{J}_{11}^5(h)$ composed of elements in $\mathcal{J}_4^4(h)$ and $\mathcal{J}_8^4(h)$. The process of updating inheritance

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

indexes at step $q = 5$ results in the following.

$$
\begin{aligned}
q = 5 \ \text{ and } \ d = 2: \qquad & \mathcal{N}_3^5 && \leftarrow \left\{\mathcal{N}_3^4, \mathcal{J}_6^4\right\} = \{3, 6\} \\
q = 5 \ \text{ and } \ d = 4: \qquad & \mathcal{N}_9^5 && \leftarrow \left\{\mathcal{N}_3^5, \mathcal{N}_9^4\right\} = \{3, 5, 6, 9\} \\
q = 5 \ \text{ and } \ d = 5: \qquad & \mathcal{N}_{10}^5 && \leftarrow \left\{\mathcal{N}_3^5, \mathcal{N}_{10}^4\right\} = \{2, 3, 5, 6, 9, 10\} \\
q = 5 \ \text{ and } \ d = 6: \qquad & \mathcal{N}_{11}^5 && \equiv \left\{\mathcal{N}_4^4, \mathcal{N}_8^4, 11\right\} = \{2, 4, 5, 8, 11\}
\end{aligned}
$$

In the seventh step, a new set of indicators, $\mathcal{J}_{12}^6(h)$, emerges at the first height level out of $\mathcal{J}_3^5(h)$ and $\mathcal{J}_{11}^5(h)$. The algorithm moves to the last step where the height levels path $\mathcal{D}^7$ has eighteen height levels. The algorithm explores them all with the link function delivering all possible values of the link indicator $\mathcal{L}_d^7$. If the value is different from the empty set, $\emptyset$, then new test development set, $C_{tmp}(h)$, is compared with all existing development sets. If such a set is already in the list of the development sets, $\mathcal{C}(h)$, then the algorithm updates appropriate level sets indicators and inheritance indexes of the existing development set and moves on. In this step all of the test development sets along the height level path have been previously added to the collection of development sets.

At the very last step, the seventh, the algorithm reaches the eighteenth level on the height levels path, $\mathcal{D}^7$, and terminates. It returns the final list of level sets indicators, $\mathcal{J}^7(h)$, which is precisely the list of core determining indexes $\mathcal{J}(h)$.

Table 4.1 presents a list of core determining indexes, $\mathcal{J}(h)$, together with a list of corresponding core determining sets, $\mathcal{C}(h)$.

The lists comprise eleven components. The first six core determining sets are equivalent to the elemental level sets. However, the third and the fourth core determining sets, $C_3(h)$ and $C_4(h)$ are composed of unions of two elemental level sets. In both cases one of the component sets is a parental set, $\tau_{3,1}(h)$ and $\tau_{1,2}(h)$ respectively, corresponding to the elemental level set at the initialization stage. These sets extend the trivial list of elemental levels delivering the outer set of the identified set.

Rows seven through eleven describe non-trivial components that, together with rows one through six, deliver core determining collection of sets. We skip the twelfth trivial set, the whole interval $[0, 1]$. Its corresponding core determining inequality always holds.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

p	$C_p(h)$		content of $C_p(h)$	$\mathcal{J}_p(h)$
1	$(0, h_{1,1})$	$=$	$\tau_{1,1}$	$\{(1,1)\}$
2	$(h_{1,1}, h_{2,1})$	$=$	$\tau_{2,1}$	$\{(1,1)\}$
3	$(h_{2,1}, h_{3,1})$	$=$	$\tau_{3,1} \cup \tau_{3,2}$	$\{(3,1),(3,2)\}$
4	$(0, h_{3,1})$	$=$	$\tau_{1,1} \cup \tau_{1,2}$	$\{(1,1),(1,2)\}$
5	$(h_{1,1}, h_{2,2})$	$=$	$\tau_{2,2}$	$\{(2,2)\}$
6	$(h_{2,2}, 1)$	$=$	$\tau_{3,2}$	$\{(3,2)\}$
7	$(0, h_{2,2})$	$=$	$\tau_{1,1} \cup \tau_{2,1} \cup \tau_{1,2}$	$\{(1,1),(2,1),(1,2)\}$
8	$(h_{2,1}, h_{2,2})$	$=$	$\tau_{2,1} \cup \tau_{2,2}$	$\{(2,1),(2,2)\}$
9	$(h_{2,2}, 1)$	$=$	$\tau_{2,2} \cup \tau_{3,1} \cup \tau_{3,2}$	$\{(2,2),(3,1),(3,2)\}$
10	$(0, h_{2,2})$	$=$	$\tau_{1,1} \cup \tau_{1,2} \cup \tau_{1,2} \cup \tau_{2,2}$	$\{(1,1),(1,2),(2,1),(2,2)\}$
11	$(h_{2,1}, 1)$	$=$	$\tau_{2,1} \cup \tau_{2,2} \cup \tau_{3,1} \cup \tau_{3,2}$	$\{(2,1),(2,2),(3,1),(3,2)\}$

Table 4.1: Core determining indexes, $\mathcal{J}(h)$, and corresponding core determining sets, $\mathcal{C}(h)$, with their components, elemental level sets. We suppress h from notation of elemental level sets. We write $\tau_{m,k}$ instead of $\tau_{m,k}(h)$.

4.5 Concluding remarks

We propose a practical method for obtaining core determining sets and core determining indexes in the class of partially identified models when the observable variables are discrete. They give rise to core determining inequalities, i.e. the set of ultimate identification questions.

We focus on the models where discrete outcome is a non-additive function of discrete covariates and continuous, latent heterogeneity. This setup comprises a large class of complete models prevalent in econometric practice of which instrumental variable models with discrete covariates, like ordered outcome, multinomial choice, binary panel data model or random coefficients, account for particular instances.

We elucidate our results in the context of ordered outcome model and demonstrate practical feasibility and usefulness of the approach proposed.

References

Beresteanu, A., I. Molchanov, and F. Molinari (2009): "Sharp Identification Regions in Models with Convex Predictions: Games, Individual Choice, and Incomplete Data," *In CEMMAP working paper CWP27/09.*

Beresteanu, A., I. Molchanov, and F. Molinari (2010): "Partial Identification Using Random Set Theory," *In CEMMAP working paper CWP40/10.*

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Chesher, A.: (2010): "Instrumental Variable Models for Discrete Outcomes," *In Econometrica 78(2), 575-601.*

Chesher A., and K. Smolinski (2009): "IV Models of Ordered Choice," *In CeMMAP Working Paper CWP 37/09.*

Chesher A., and K. Smolinski (2010): "Sharp Identified Sets for Discrete Variable IV Models," *In CeMMAP Working Paper CWP 11/10.*

Chesher, A., A. Rosen, and K. Smolinski (2011): "An instrumental variable model of multiple discrete choice," *In CeMMAP Working Paper CWP 06/11.*

Chesher, A., A. Rosen, and K. Smolinski (2012): "Generalized Instrumental Variable Models". *In preparation.*

Chernozhukov V., and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245-261.

Molchanov, I. S. (2005): *Theory of Random Sets.* Springer Verlang, London.

Galichon, A., and M. Henry (2009): "Set Identification in Models with Multiple Equilibria," *In working paper, Montreal and Ecoles Polytechnique.*

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Appendix A

	$q=1$		$q=2$		$q=3$		$q=4$		$q=5$		$q=6$		$q=7$	
$d\downarrow$	(n,b)	$\mathcal{L}^q_{n,b}$	(n,b)	$\mathcal{L}^q_{n,b}$	(n,b)	$\mathcal{L}^q_{n,b}$	(n,b)	$\mathcal{L}^q_{n,b}$	(n,b)	$\mathcal{L}^q_{n,b}$	(n,b)	$\mathcal{L}^q_{n,b}$	(n,b)	$\mathcal{L}^q_{n,b}$
1	$(1,4)$	$\subset$	$(2,5)$	σ	$(1,8)$	$\emptyset$	$(1,9)$	$\emptyset$	$(1,10)$	$\emptyset$	$(3,11)$	σ	$(4,9)$	$\emptyset$
2	$(1,5)$	$\emptyset$			$(2,6)$	$\emptyset$	$(2,9)$	σ	$(3,6)$	$\supset$			$(4,10)$	$\emptyset$
3	$(1,6)$	$\emptyset$			$(3,4)$	$\emptyset$			$(3,7)$	$\emptyset$			$(5,7)$	σ
4	$(2,4)$	σ			$(3,5)$	σ			$(3,9)$	$\subset$			$(6,7)$	$\emptyset$
5									$(3,10)$	$\subset$			$(6,8)$	$\emptyset$
6									$(4,8)$	σ			$(6,11)$	$\emptyset$
7													$(7,8)$	σ
8													$(7,9)$	σ
9													$(7,10)$	σ
10													$(7,11)$	σ
11													$(7,12)$	σ
12													$(8,9)$	σ
13													$(8,10)$	$\subset$
14													$(8,12)$	σ
15													$(9,10)$	$\subset$
16													$(9,11)$	σ
17													$(9,12)$	σ
18													$(10,11)$	σ

Notation
$\mathcal{L}^q_{n,b} \equiv \mu(\mathcal{C}^q_n, \mathcal{C}^q_b)$

Table 4.2: This table illustrates seven steps of the algorithm with their corresponding height levels paths for three (ordered) outcomes binary covariate instrumental variable model. The analysis is conditional on the following arrangement of structural parameters, $h_{1,1} < h_{1,2} < h_{2,1} < h_{2,2}$. Columns describe seven steps of the construction. Rows relate to the height levels of the algorithm. In each step q and for every height d we present link indicators $\mathcal{L}^q_d$, that are results of the link function λ evaluated on the n^{th} and b^{th} development sets, $\mathcal{C}^q_n$ and $\mathcal{C}^q_b$.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

Progressive update of the lists of level sets indicators $\mathcal{J}^q(h)$							
	$q=1$	$q=2$	$q=3$	$q=4$	$q=5$	$q=6$	
$\mathcal{J}^0(h)$ coding	$\mathcal{J}^1(h)$ coding	$\mathcal{J}^2(h)$ coding	$\mathcal{J}^3(h)$ coding	$\mathcal{J}^4(h)$ coding	$\mathcal{J}^5(h)$ coding	$\mathcal{J}^6(h)$ coding	$\mathcal{J}(h)$ coding
$\mathcal{J}^0_1(h)$ ●○○○○○							$\mathcal{J}_1(h)$ ●○○○○○
$\mathcal{J}^0_2(h)$ ○●○○○○							$\mathcal{J}_2(h)$ ○●○○○○
$\mathcal{J}^0_3(h)$ ○○●○○○					$\{\mathcal{J}^4_3, \mathcal{J}^4_6\} \to \mathcal{J}^5_3$ ○○●○○●		$\mathcal{J}_3(h)$ ○○●○○●
$\mathcal{J}^0_4(h)$ ○○○●○○	$\{\mathcal{J}^0_1, \mathcal{J}^0_4\} \to \mathcal{J}^1_4$ ●○○●○○						$\mathcal{J}_4(h)$ ●○○●○○
$\mathcal{J}^0_5(h)$ ○○○○●○							$\mathcal{J}_5(h)$ ○○○○●○
$\mathcal{J}^0_6(h)$ ○○○○○●							$\mathcal{J}_6(h)$ ○○○○○●
	$\{\mathcal{J}^0_2, \mathcal{J}^1_4\} \Longrightarrow \mathcal{J}^1_7$ ●●○●○○						$\mathcal{J}_7(h)$ ●●○●○○
		$\{\mathcal{J}^1_1, \mathcal{J}^1_5\} \Longrightarrow \mathcal{J}^2_8$ ○●○○●○					$\mathcal{J}_8(h)$ ○●○○●○
$\Longrightarrow$ define			$\{\mathcal{J}^2_3, \mathcal{J}^2_5\} \Longrightarrow \mathcal{J}^3_9$ ○○●○●○		$\{\mathcal{J}^4_6, \mathcal{J}^4_9\} \to \mathcal{J}^5_9$ ○○●○●●		$\mathcal{J}_9(h)$ ○○●○●●
$\to$ update				$\{\mathcal{J}^3_2, \mathcal{J}^3_9\} \Longrightarrow \mathcal{J}^4_{10}$ ○●●○●○	$\{\mathcal{J}^4_6, \mathcal{J}^4_{10}\} \to \mathcal{J}^5_{10}$ ○●●○●●		$\mathcal{J}_{10}(h)$ ○●●○●●
●/○ true/false					$\{\mathcal{J}^4_4, \mathcal{J}^4_8\} \Longrightarrow \mathcal{J}^5_{11}$ ●●○●●○		$\mathcal{J}_{11}(h)$ ●●○●●○
$\{\cdot,\cdot\}$ concatenate						$\{\mathcal{J}^5_3, \mathcal{J}^5_{11}\} \Longrightarrow \mathcal{J}^6_{12}$ ●●●●●●	$\mathcal{J}_{12}(h)$ ●●●●●●
●●●●●● $(1,1), (2,1), (3,1), (1,2), (2,2), (3,2)$							

Table 4.3: Evolution of the level sets indicators, $\mathcal{J}^q(h)$, in the progressive steps of the algorithm. In columns we illustrate updates and expansion of the list of level sets indicators in the consecutive steps of the algorithm. Further right column presents core determining indexes, $\mathcal{J}(h)$, for the three (ordered) outcomes and binary covariate instrumental variable model where the structural parameters are in the arrangement $h_{1,1} < h_{1,2} < h_{2,1} < h_{2,2}$. Throughout we trace inclusion of elemental level sets with empty and filled circles. Circles marked in blue correspond to elemental level sets in the rank list $\tau_1(h)$ and marked in black correspond to the sets in rank list $\tau_2(h)$. At the same time filled dots stand for indexes that are included into level sets indicators. In the core of the table we drop h simplifying notation for the list of level sets indicators. We write $\mathcal{J}^q_p$ instead of $\mathcal{J}^q_p(h)$.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute
DOI: 10.2870/35896

Appendix B

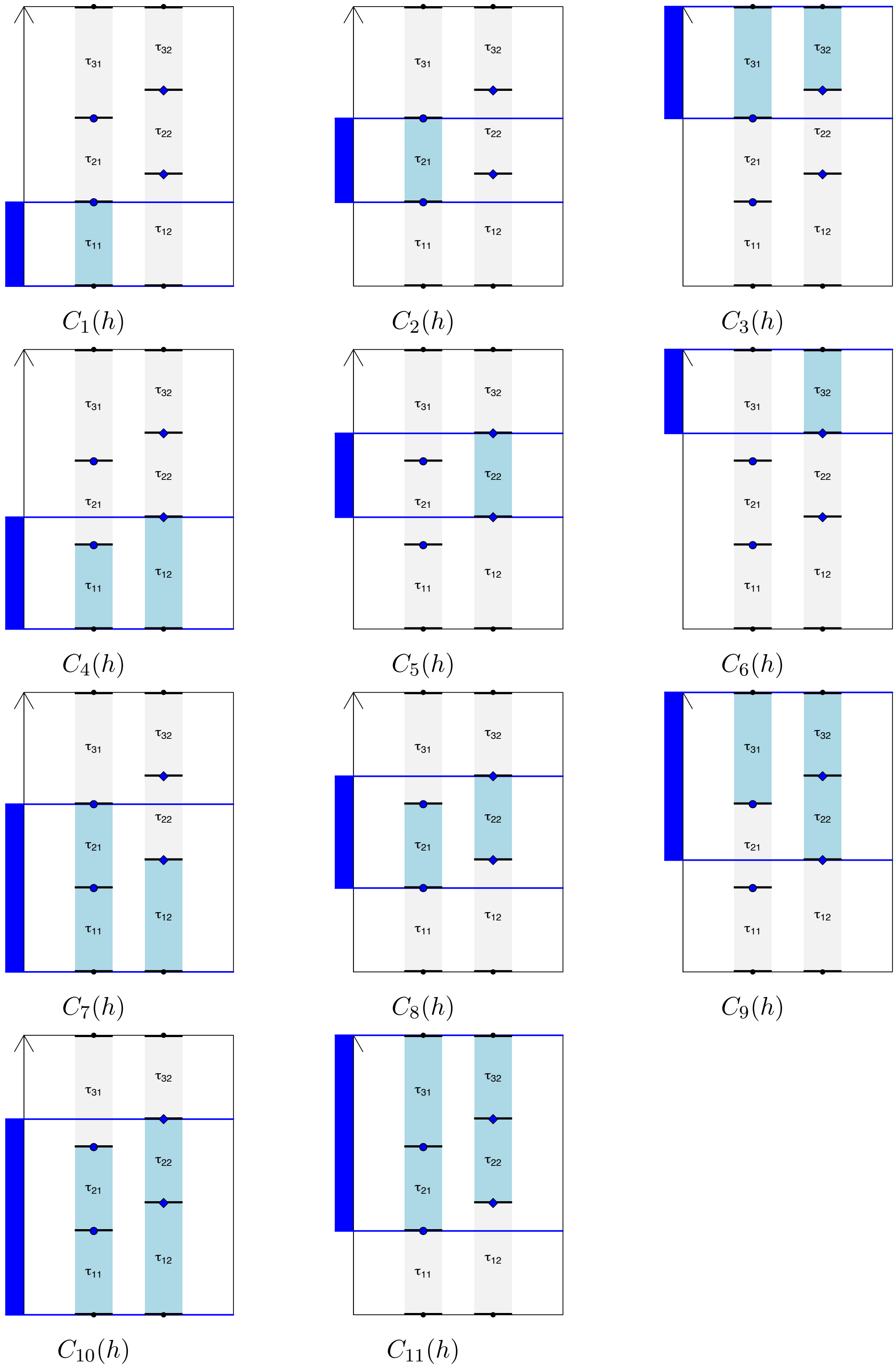


Figure 4.2: Case A: The model with three outcomes and binary explanatory variable. Structural parameters in in the arrangement (b) $h_{1,1} < h_{1,2} < h_{2,1} < h_{2,2}$. Graphs present core determining sets, $C_p(h)$ in blue, and their corresponding elemental level sets.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute
DOI: 10.2870/35896

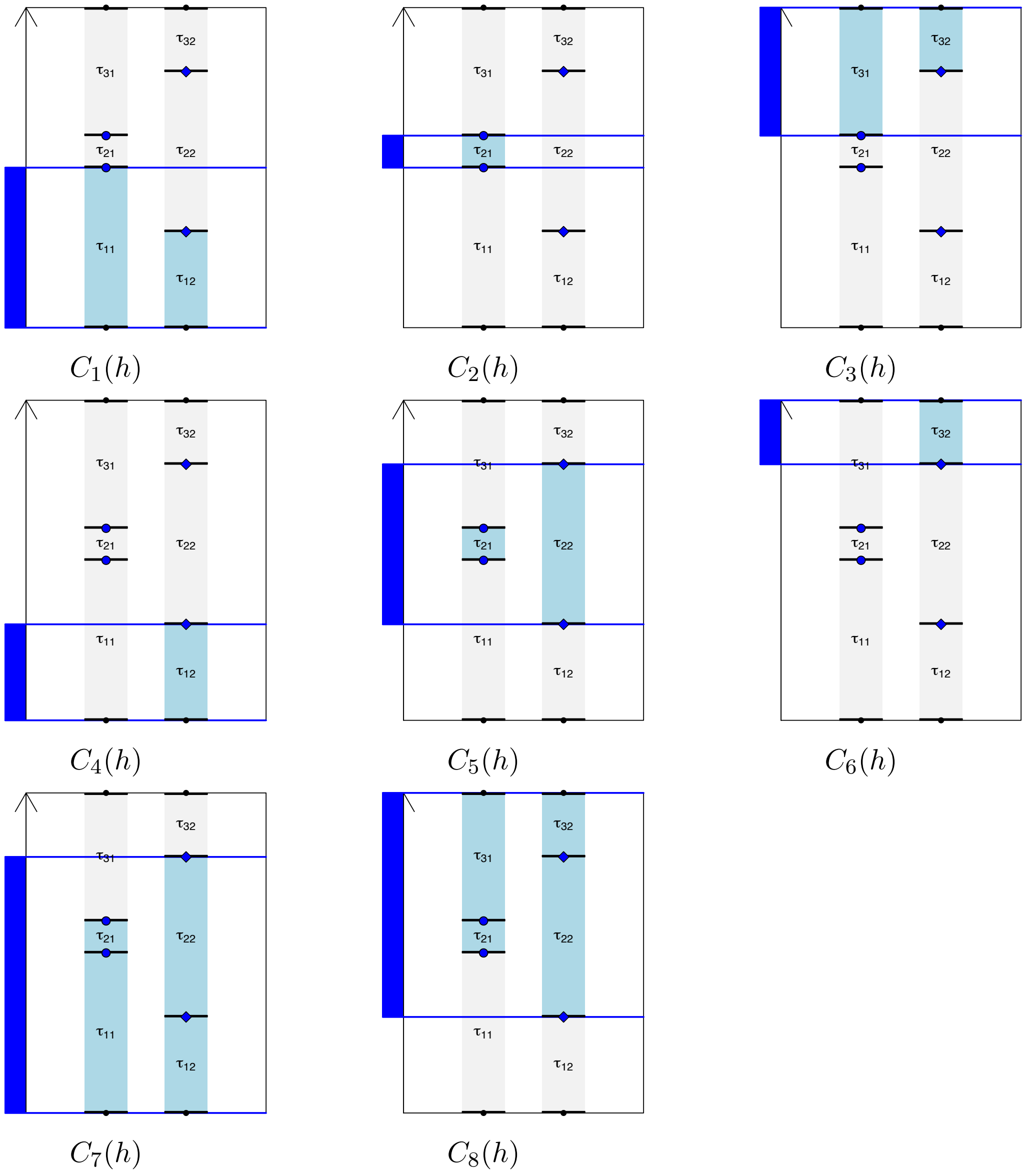


Figure 4.3: Case B: The model with three outcomes and binary explanatory variable. Structural parameters in in the arrangement (b) $h_{1,2} < h_{1,1} < h_{1,2} < h_{2,2}$. Graphs present core determining sets, $C_p(h)$ in blue, and their corresponding elemental level sets.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896

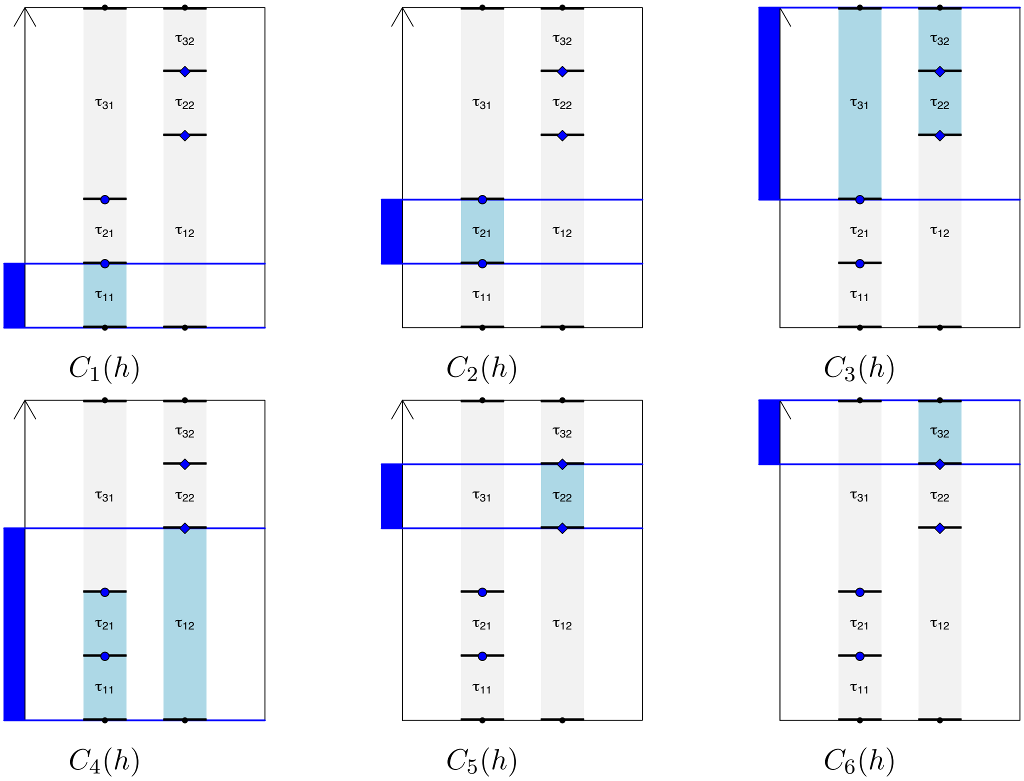


Figure 4.4: Case C: The model with three outcomes and binary explanatory variable. Structural parameters in in the arrangement (c) $h_{1,1} < h_{2,1} < h_{1,2} < h_{2,2}$. Graphs present core determining sets, $C_p(h)$ in blue, and their corresponding elemental level sets.

Smolinski, Konrad (2012), Single Equation Instrumental Variable Models: Identification under Discrete Variation
European University Institute DOI: 10.2870/35896