Nonparametric Structural Analysis of Discrete Data : The Quantile-based Control Function Approach

Jinhyun Lee

Department of Economics

University College London

Supervisors:

Professor Andrew Chesher
Professor Sokbae Simon Lee

Thesis submitted for the degree of Doctor of Philosophy

at University College London

September 2010

Declaration of Authorship

I, *Jinhyun Lee*, confirm that the work presented in this thesis *"Nonparametric Structural Analysis of Discrete Data : The Quantile-based Control Function Approach"* is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:______________________

Date:_________________________

For Aslan

"...there should be no poor among you...
...there will always be poor people in the land... (*Deu.* 15:4, 11)..."

Acknowledgements

I am deeply indebted to my supervisor Professor Andrew Chesher. He was the only person who did not dismiss my naive hope to do both theoretical and empirical work for Ph D study[1]. My small observations have been enriched by his insightful comments meeting after meeting and anyone could easily imagine why if the person knew him. Many of the ideas in this thesis are also based on his lectures and the comments he gave in the seminars that I attended to. I am so lucky to have learned this unique architect's view on building a house and have a chance to scrutinize Chesher (2003), one of the breakthroughs in identification under endogeneity. Whether I can build anything myself or whether I can be an architect myself or remain as a mere builder would depend on me from now on. I just wish the depth and insights that he has would be something obtainable by efforts.

I thank Yoon-Jae Whang and Hide Ichimura for their advice on my studies in the master programs. I am grateful to Hide Ichimura and Christian Dustman for their decision[2], and Richard Blundell and Sokbae Simon Lee for their support and encouragement.

I had opportunities to present chapter 5 in 2008 the Latin American Meeting of Econometric Society (LAMES) in Rio de Janeiro, Brazil, 2009 Royal Economic Society (RES) annual meeting, 2009 European Meeting of the Econometric Society (ESEM) in Barcelona, cemmap/ESRC workshop on Quantile Regression, SETA in Kyoto, Japan. Chapter 2 is scheduled to present in 2011 North American Winter Meeting of Econometric Society (NAWMES). A preliminary paper implementing the testable implication on endogeneity in chapter 3 was presented in 2009 Netherlands Econometric Study Group (NESG) and 2009 Econometric Study Group(ESG) annual conference in Bristol and is planned to be presented in Econometric Society World Congress in 2010.

I am grateful to many friends in London and Seoul who endured all my complaints and grumblings over the period[3]. I should mention John Stott[4] who I believe is one of the reasons why I was sent to London. I also thank my family, the greatest asset I have. Special thanks should go to my mother and my younger sister, who are my mentors as well as best friends. It was the hardest time for three of us - we had to be alone in Seoul, Lima, and London and had to go through a period over which nobody could help and we could complete this period because of each other. I thank them for their love and wisdom that have sustained me.

[1] I think that theoretical developments should be motivated by economic examples and interact with empirical findings and one of the criteria in evaluation of theoretical work should be its applicability, of course, not the most important criteria as some claim. Now, I may have to agree with the skeptical views on the quality of the work in such an attempt or I may have to show that this is possible.

[2] I wish I would not have to mention why, and I also wish they would not remember why.

[3] Special thanks to Mare Sarr, a man of (worldly) wisdom with listening ears and to YongSung Lee and BoYoung Won for their continuous encouragement and advice.

[4] "Paul" in the 20th century, a perfect example of a man with warm heart and cool head. I would divide my life into two - before and after I know John Stott. I have "absorbed" everything he has said and so many things in me have changed ever since. London and even the boring and bizzare English tradition have become so special because of him. It was rather shocking to watch the congregation sing the hymn by John Wesley at All Souls. (I am used to it now and sing it with them, though.) He said that they were like that because of the education system called "English public school". That makes sense - it cannot be natural, it must be cultivated and encouraged by education. However, there must be something extraspecial in it since John Stott was brought up by this system.

My final acknowledgements go to my wonderful Counselor(wC). I need to quote this :

> "....our desire is not that others might be relieved, while you are hard pressed, but there might be *equality*. At the present time your plenty will supply what they need, so that in turn their plenty will supply what you need. Then there will be *equality*, as it is written : "He who gathered much did not have too much, and he who gathered little did not have too little....". 2*Cor* 8:13-15, *emphasis added*"

When I first read this, I could not believe my eyes to find the word "equality" and the idea of "social insurance" in this old book. This passage is why I decided to start economics in 1997, the severe year of financial crisis in East Asia. I observed the necessity of social insurance and the heterogeneous impacts of the crisis on people resulting in growing inequality. Then I took up econometrics as my specialization, then I came to UCL[5]. It seemed that I would not do anything related to social insurance. I did not know where I was going until quite recently when I began to appreciate the value of the identification study[6]. The role I would play if I ever do would not be the same as what I expected or what is usually expected. I used to think more about what I "can" do, which is little, but now I try to think about what I "ought to" do. It is still vague, but now I am more convinced where I "should" go as an econometrician. Life is not a random mess as it seems to be. Everyone has a purpose to fulfil and is heading toward it. The last a few years have been the evidence of this and this is the biggest lesson I have learned over the period. All the way through my wC has been with me and told me to be still and to wait. I often fretted, but I now realize that I did not have to because their swords will pierce their own hearts[7] in the end. I have been learning how to be silent. It is so great to be under this good and perfect guidance all the time.

"Everything is possible, but not everything is beneficial[8]". The economy will reach an equilibrium eventually, but we see so many traps and vicious circles in reality. Without intervention, poor people now may be poor in the future and poor countries now may be poor in the future. It said, "there will always be the poor, but there should be no poor people among us[9]". This is not a contradiction - the former is the nature of the world,

[5]UCL has a strong tradition in applied economics. Having been trained at UCL to be an econometrician will turn out to be an extraordinary experience in the end. Not that I fully agree with all the views on econometrics here, and not that the econometrics courses were very helpful in developing the necessary skills in the field, but that I could learn what would be used in practice. I do not think the values of work in econometrics lie in the usefulness only, but if possible, I want to work on something to be used.

[6]It seems to me the true value and necessity of identification has been rediscovered only recently under the nonparametric setup, where the structural features of interest should be specified, not just "β". Chesher (2003,2005,2007), Imbens and Newey (2009), Matzkin (1992, 2003, 2008), Roehrig (1988) would be considered to be in line with the classical identification analysis in the 50s under the structural approach. Also recent attempts to understand the empirical contents - what can be identified - of specific economic models such as Athey and Haile (2002) etc are examples that identification study is appreciated. Chapter 2, which is the last outcome among the four major chapters, is my favorite regardless of how it would be evaluated - I summarize and extend some concepts under the general nonparametric setup following Koopmans and Reiersol (1950), Breusch(1986), and Chesher (2008b). I managed to understand Koopmans and Reiersol (1950) and Hurwicz (1950a) only after I had spent a long time in recent identification studies.

[7]*Ps.* 37:15.

[8]1*Cor.* 6:12.

[9]*Deu.* 15:4,11.

while the latter is what we should aim at. Any policy, or "mechanism" would be required to break the vicious circle and they should be designed by taking the *nature* of individual decision makers into account. We know that *equality* should be desired, but we also know that it would not be *sustainable* due to the human nature[10].

Economists should be as "shrewd as snakes and as innocent as doves[11]" in this sense. We need to understand the nature of individual's behavior, but at the same time we need to judge whether it leads to beneficial outcomes or not and to suggest any possible remedies to any undesirable status quo[12]. Econometricians should understand the goals of economists who search for the social good, and should provide *evidence* from the observed individuals' choices for or against any suggested remedies so that policy makers make right decisions based on the evidence. The issue is to understand under what conditions the observed patterns of individual choices can be interpreted as evidence for or against the suggested solutions. Identification study comes in to play a role for this. Identification result is required to interpret any statistical facts in the economists' language. This thesis is an attempt to provide such an interface between economics and statistics without relying on hard to be justified assumptions to clarify the limitations in interpreting the observed facts as *evidence*.

February, 2010

Jinhyun Lee

[10] Provision of social insurance may be a solution to inequality, however, full insurance (equal treatments) would never achieve the optimal solution found ex ante because individuals' choice will be different ex post once full insurance is guaranteed. The fact that we can never achieve full insurance(the first best) and the fact that we need to provide incentives (unequal treatments) to achieve the second best are the most fascinating results to me in economics - economists admit the grim reality but nonetheless struggle to do something, which is better than nothing. I adore this struggling of those smart people.

[11] *Mat.* 10:16.

[12] This statement may be controversial. But I believe policy or intervention is necessary in which sense the exact information on the impacts of any policy ex ante and ex post is more and more important.

Contents

[13] I adopt this title just to indicate that I follow the spirit of the book entitled, *Structural Analysis of Discrete Data*, not that this thesis is comparable with the great classic.

[14] This chapter is motivated by A. Chesher's unpublished lecture note entitled "Evidence in Economics" in which *falsifiability* of a model is discussed using point identification.

Chapter 1

Introduction

> " ... One might regard problems of identifiability as a necessary part of a specification problem. We would consider such a classification acceptable, provided the temptation to specify models in such a way as to produce identifiability of relevant characteristics is resisted. *Scientific honesty demands that the specification of a model be based on prior knowledge of the phenomenon studied and possibly on criteria of simplicity, but not on the desire for identifiability of characteristics in which the researcher happens to be interested.*
>
> Identification problems are not problems of statistical inference in a strict sense, since the study of identifiability proceeds from a hypothetical exact knowledge of the probability distribution of observed variables rather than from a finite sample of observations. However, it is clear that the study of identifiability is undertaken in order to explore the *limitations of statistical inference.* ...(Koopmans and Reiersol (1950), pp169-170, emphasis added)"

This thesis focuses on identifiability and testability of *structural* features. As was mentioned in the above quotation, the study of identifiability is to clarify what can be inferred about underlying economic decision mechanisms from data analysis and what should be believed about them for such inference, which is to explore "*the limitations of statistical inference*". It is demonstrated in this thesis that identification results can also guide statistical inference - estimation and testing - of structural features of interest.

In this thesis the main interest is econometric modeling of individuals' economic decisions (choice) and the outcome of their decisions which may be modeled by a *triangular* system. Such individual decision mechanisms may be modeled based on economic models. Identification is an essential step in associating economic models with data, by which certain *economic interpretations* of the results from data analysis can be justified. Difficulties arise by the fact that the unobserved elements should be incorporated in econometric modeling. Economic models may specify individuals' decision mechanisms as deterministic relationships between relevant variables, data analysis should be conducted under the stochastic framework since we cannot observe all the relevant variables in reality. While economic models are constructed in a parsimonious way to address specific economic issues, when data analysis is attempted based on a parsimonious economic model, we need to be concerned about how we treat unspecified elements in the economic model that affect the outcome of concern. Under uncertainty as in decisions under asymmetric information the economic models specify stochastic relationships, where the source of randomness is well-defined in the specific economic context[1]. However, the sources of stochastic elements in econometric modeling for data analysis are more likely to be multi-dimensional and they are hard to define in economic terms. Hurwicz's (1950a) structure is adopted to incorporate both economic arguments from economic models and the stochastic aspects of data analysis.

1.1 Modeling Economic Processes - the Hurwicz (1950a) Structure

Suppose that the outcome of interest W is *generated* by a structural relation of the following

$$W = h(X, U), \tag{1}$$

where X are relevant observed variables and U is a vector of relevant unobservable heterogeneity with the conditional distribution $F_{U|X}$. Throughout the thesis the term "unobserved heterogeneity" is often used rather than "error" when the unobserved elements are considered to be determinants of the outcome. When we assume that individual's choice can be represented by structural relations, every argument of the structural relation should play distict roles in determining the value of the outcome. The terminology "error" would be more appropriate if we actually attempted to analyze the structural relation of interest with data. Then the issues regarding the measurement of the arguments of the structural function - whether they are observable or not, whether we can use proxy variables for the unobserved arguments, or whether some of the arguments are measured with error - should be considered.

The distribution of the unobservables ($F_{U|X}$) together with the structural relations (h) will determine the distribution of the observables as follows[2].

$$\begin{aligned}
\overbrace{F_{W|X}(w|x)}^{\textbf{Data/Reduced Form}} &= \Pr[W \le w | X = x] \\
&= \Pr[h(X,U) \le w | X = x] \\
&= \underbrace{\int_{\{u|h(x,u)\le w\}} dF_{U|X}(u|x)}_{\textbf{Structure}}
\end{aligned} \tag{HR}$$

This relation is called the *Hurwicz Relation* and will be referred to **(HR)** throughout the thesis. Identification of the structure (or structural feature), the elements of the right hand

[1] such as ability in education signalling models, effort in moral hazard models, productivity in adverse selection models, and agent's valuation in auction models. See Matzkin (2007).

[2] This is the general nonidentification result described in Chesher (2009) since identifying h and $F_{U|X}$ separately is not possible without imposing further restrictions.

side, $\{h, F_{U|X}\}$, is achieved from information from the observed distribution, $F_{W|X}(w|x)$, the left hand side object by imposing restrictions on either $\{h, F_{U|X}\}$ or $F_{W|X}(w|x)$, or on both. We call the structure, the tuple of $\{h, F_{U|X}\}$, an underlying *economic* data generating process, which is in contrast with the usage in statistics where the left hand side object, $F_{W|X}(w|x)$, is called the "data generating process"

The focus on the structure, rather than the distribution of the observables is required if we attempt to interpret the results from data analysis in economic contexts. The objects of interest in economic examples are usually defined in terms of structural relations. Note that without further restrictions we cannot recover the structure from data. For example, with linear structural function and standard normality assumption on the scalar U, we have

$$F_{W|X}(w|x) = \Phi(w - x\beta)$$

then we can identify the structural function by identifying β. Or with additively separable nonparametric structural function, $w = h(x) + u$ with normalization $E(U|X) = 0$,[3] we have

$$F_{W|X}(w|x) = F_{U|X}(w - h(x)|x)$$

then we can identify the structural function $h(x)$ by $E(W|X)$. When we assume that the structural function is additively nonseparable, $w = h(x, u)$, we need to assume that U is a scalar, normalized to Uniform (0,1) (single index unobservables (SIU) restriction, which will be discussed in Chapter 3), and h is strictly increasing in u to identify the structural "quantile" function $h(x, u)$.

$$\begin{aligned} F_{W|X}(w|x) &= F_{U|X}(h^{-1}(x, w)|x) \qquad \text{(HR-SIU-C)} \\ &= h^{-1}(x, w) \end{aligned}$$

where the second equality follows by uniform normalization[4]. This relation will be referred to (HR-SIU-C) indicating "Hurwicz Relation with Single Index Unobservables for continuous variables". Under the uniform(0,1) normalization, the unobserved heterogeneity can be called "*unobserved type*". Throughout the thesis single index unobservables (SIU) assumption is maintained except for section 4.2 and 4.3 in Chapter 4 where "excess heterogeneity" is allowed for.

Econometric modeling involves finding out the restrictions by which identification and inference of (some features of) the structure can be achieved. Several issues should be considered in choosing the restrictions in micro-econometric modeling. Firstly, the nature[5] of the data encoding individual decisions that affects $F_{W|X}$, and the presence of endogeneity that affects $F_{U|X}$ need to be specified. Both issues require special attention in identification. Secondly, econometricians should also consider how to incorporate restrictions on the structure $\{h, F_{U|X}\}$ imposed by economic models.

[3] With mean zero restriction a vector U is admitted.

[4] Note that without uniform normalization we cannot separately identify the structural function from the distribution of the unobservable.

[5] By this we mean censoring, discreteness, or aggregation.

Econometric models would be constructed by restricting either the structure, the right hand side of (HR) or the distribution of observables, the left hand side of (HR), or both such that identification can be achieved. The rest of this chapter introduces the issues considered in the thesis with regards to identifiability and testability and describes the contents of each chapter.

1.2 Microdata - the Impacts of Observational Processes on Identification

In the data analysis of individual choices how they are observed, called observational processes in this thesis, has a crucial impact on econometric modeling. Studying the impacts of discrete variation on identification is required because we have to deal with qualitative data in economic applications where the data analysis uses micro-surveys. How a variable is observed - especially, whether it can be considered to be continuous or not - has crucial impacts on identification. This focus on the impacts of the nature of the data and the role of unobserved heterogeneity on the econometric methods under weak restrictions has the roots in the parametric analysis of simultaneous limited dependent variable (LDV) models (Heckman (1978), Smith and Blundell (1986), Blundell and Smith (1989,1994) and Rivers and Vuong (1988), etc), mixture models for duration (Lancaster (1990), van den Berg (2002) for survey) and count data (Cameron and Trivedi (1998) for survey).

Discreteness of the dependent variable would restrict both the left hand side and the right hand side of (HR) in the sense that $F_{W|X}$ or quantiles of it are not continuous nor differentiable and $h(X, U)$ should be additively non-separable. This thesis investigates how the observational processes influence the identifying power of the model.

1.3 Nonparametric Identification

Berry and Tamer (2007) categorizes identification practices into two different approaches : top-down and bottom-up. The top-down approach starts from a specific functional form, and sees if identification of the parameters of interest is achieved under that specific specification. This approach often is adopted for the development of estimators, where identification is required to ascertain the objects to be estimated are well-defined[6]. On the other hand, the bottom-up approach starts from the restrictions that are from the economic arguments and sees if identification is achieved, and if not, then more restrictions are sequentially imposed. Since I am concerned with the limitations of data analysis with the weakest possible restrictions - in the sense of just-identification, I advocate the bottom-up approach. I also advocate nonparametric restrictions, for example, shape restrictions such as monotonicity or convexity, rather than parametric restrictions such as linearity in the

[6] For example, see Newey and McFadden (1994). Their discussion on identification is based on this approach - identification is a necessary step, rather than the goal of the analysis.

structural function since it is hard to believe that they come from economic arguments[7].

As such I follow Roehrig (1988)'s argument for "nonparametric" identification and consider a non-additive triangular structure. That is, I do not impose any functional or distributional assumptions on the structure, and I assume the structural equations are additively nonseparable with the unobserved variable, which allow for the nonparametric analysis of limited dependent variables[8] as well as heterogeneity in response among the observationally same individuals. Identification should be distinguished from the "specification" problem, let alone the estimation and inference problems categorized by R.A. Fisher (1922). Rather, we follow the new definition of "specification" in Koopmans and Reiersol (1950), the specification of a model that is characterized by a set of nonparametric restrictions that guarantee identification.

The goal is to define the set of "justifiable" nonparametric restrictions[9] that achieve identification - either point or set - without relying on any parameterization of functional forms nor distribution of the unobservables. Structural interpretations of any objects obtained by data analysis are only possible when we believe these restrictions.

Many attempts have been made recently to link specific economic models with data under weak restrictions, for example, nonparametric identification of auction models[10], regulation models (Perrigne and Vuong (2007)), adverse selection models (D'Haultfaeuille and Fevrier (2007)) among others. I have the same goal as these studies, that is, defining the limits of data analysis with no restrictions unsupported by economic models imposed, and clarifying what should be believed to conclude anything based on the information from the data analysis, but we abstract from the specific economic models and focus on the issues that occur with the nature of observation - specifically, discreteness of data.

The first step in nonparametric structural analysis is to determine whether the objects of interest are identifiable. Under a specific economic model various structural features as functionals of the structure can be defined and convey specific economic meanings by identification results. When it is abstracted from any specific economic contexts, the usual structural features of interest would be partial derivatives or partial differences. In many cases such nonparametric identification analyses have been done under the assumption that structural relations are continuous and differentiable. However, when a structural relation includes observed and unobserved variables that may covary, discreteness of data limits the identifying power of models as I show in this thesis under triangularity.

[7] In many cases the usually assumed linear structural relations are often implausible, or have uninteresting economic implications - consider, for example, under which conditions we can have linear demand or supply functions.

[8] By this we mean binary outcomes, censored outcomes, count data, interval data, categorical data etc. Note that with these outcomes it is natural to assume the nonparametric structural function is additively nonseparable - with additive error we would have the support of the error always dependent on the observable arguments. Consider the drawbacks of the linear probability models as a well known example.

[9] "justifiable" in the economic context, such as monotonicity, or convexity, which are derived from the economic models.

[10] See Athey and Haile (2003,2005) for survey and for the compelling arguments for the study of nonparametric identification.

1.4 Examples : economic models, econometric modelling and identification

Example 1 : Auction models

The interpretation of $F_{U|X}$: The data available are individuals bids, and the underlying structure is $S = \{h, F_{U|X}\}$, where $F_{U|X}$ is the joint distribution of latent individual valuation and private information (types) satisfying certain statistical properties (e.g. independence and symmetry etc.) and h is a true mapping from the true distribution of types (U) to a distribution of observable bids (Y) implied by the assumption of Baysian Nash equilibrium. The structural relation h between bids and valuation is implied by the Bayesian Nash equilibrium, which is represented by $H(Y, X, U) = 0$. The different auction models assume different statistical properties on the bidder's private information, which opens room for testing the implications from different models using the observed data.

Example 2 : Demand for medical utilization and health insurance

Justification of triangularity : Individuals in the model make decisions on health insurance and demand for health care sequentially. In the first stage, given a menu of insurance options, individuals make a choice about health insurance based on the expectation on the future health status or their degree of risk aversion, and in the second stage after the realization of their own unobserved health status (U), they make a decision on medical utilization. In this example the structural relations (h) for health demand (W) and health insurance (Y) choice are assumed to satisfy the first-order conditions ($H(W, Y, X, U) = 0$) of agent' utility maximization problem, where X would be individual characteristics. Thus, the structure in this exmple is $S = \{h, F_{U|X}\}$, where h are the structural relations for health demand and health insurance as functions of unobservable health status as well as other variables and $F_{U|X}$ is the distribution of unobserved health status possibly dependent on the health shock. The two-step decision processes may justify the triangular structural relation.

Example 3 : Econometric modelling of contract theory

Testable implications from contract theory : When we want to analyze data which are generated by an economic model under asymmetric information, the testable implications of the model are usually expressed in terms of unobservables. The predictions from the contract theory can be expressed in terms of high/low "types" of some unobservable characteristics of agents such as agents' efforts, health status, or valuation. The unobservable variables and the distribution of them as an element of the Hurwicz (1950a) structure can be useful in modelling this situation and the identification of the underlying data generating process would be the issue we have to deal with in this case.

The use of additively nonseparable models can be useful in modelling contract theory, where interesting comparative statics with respect to unobservable variables, that is, the responses of the agents predicted by the model are heterogenous in the unobserved "type". Such responses can be measured by the partial derivatives/differences of the structural functions, and nonseparable structural relations allow for heterogeneous and random response. Thus, identification of partial derivatives or differences can be used to test the predictions of the contract theory.

Example 4 : Individual welfare analysis from aggregate household expenditure data

Econometric modelling of *observational* processes : Lack of individual data on consumption(or expenditure) has restricted the micro-object of empirical research to households. As long as all the members in a household share the same preference and the same needs, this practice of using households as the smallest micro-units of decision will be enough. Nevertheless, when we conduct welfare analysis, using households as a unit of study limits the information on individual welfare on which policy measures should be based. If we ignore inequality in intrahousehold allocation, inequality measures based on household level consumption will underestimate the real level of inequality. Furthermore, measuring poverty based on poverty line which is formed from household data may underestimate the severity of problem if there is serious inequality among members. Econometric modelling can be constructed based on this in data collection to derive individual level of consumption as in Chesher (1997).

Chesher (1997) incorporates household characteristics multiplicatively by imposing linear index restriction, thus, the average age and gender - specific individual expenditure for the whole population can be obtained by controlling the household characteristics. Lee (2009) extends Chesher (1997) to incorporate household characteristics nonparametrically in the study of unobserved individual health expenditures using household health expenditure data, which allows age and gender- specific individual demand to vary with household characteristics. This may be informative by providing the information regarding how rich and poor household allocate resources differently in the households. Health expenditures require a special attention because zero expentitures are commonly observed. Thus, Lee (2009) includes nonnegativity restriction in the disaggregation process. This procedure can be applied to any study using household budget information. The identification issue in this context involves with clarifying under what conditions the age and gender specific disaggregation can be interpreted as individual demand as a function of price and income derived from utility maximaziation.

Example 5 : *Non*identification of a nonparametric model with risk aversion in auction model (Guerre, Perrigne and Vuong (2009))

Although the important role of risk aversion of individuals in the bidders' behavior is accepted, there is lack of consensus on how to measure risk aversion. Guerre, Perrigne and Vuong (2009) studies the nonparametric identification of the utility function under the first-price auction model with risk averse bidders within the private value paradigm. They show that the benchmark model is nonindentified in general from observed bids. They find that risk aversion does not impose testable restrictions on bids and impose more restriction to achieve identification - exclusion restriction or linear index in the specification. This study demonstrates how nonparametric identification results clarify the limitations of data analysis under the specific economic contexts. Some of economic arguments have no "testable" implications on the data without imposing further restrictions.

1.5 *Nonparametric* Structural Analysis of Discrete Data[11]

By *structural* analysis, I mean causal analysis of data. The objects of economic interest in this thesis are *heterogeneous causal effects* of a variable. There have been two approaches to measuring causal effects - the potential outcomes framework and the structural approach.

In the potential outcomes framework the causal effects are measured by the difference between the counterfactuals. When the "cause" variable is binary, the causal effect of the binary variable is measured by the difference between the counterfactual outcomes when the binary variable is 1, W_1, the counterfactual outcome when the binary variable is 0, W_0. Measuring $W_1 - W_0$ is an issue because only either W_1 or W_0 would be observed : there exists a missing data problem. Since individuals' (possibly) heterogeneous causal effects are not measured, usually average effects are considered. There exists another econometric issue in measuring average effects. A simple way of measuring the average effects would be by comparing the mean of the two groups - those with the value of the binary variable 1, those with 0. However, this way may not measure the true casual effects correctly since the difference of the average outcomes in the two groups may not be solely due to the value of the binary variable. This is called the selection problem - unless the value of the binary variable is randomly assigned, there may be systematic differences in the two groups other than the value of the binary variable. If the *systematic* differences disappear once conditioning on other observed characteristics, this is called *selection on observables.* If the systematic difference still exists even after conditioning on the observed characteristics, this is called *selection on unobservables.*

Alternatively, one could adopt the regression idea to measure causality, which could be understood as the structural approach. This is the setup I take in this thesis to measure causal effects. In the structural approach the econometric issue in measuring causal effects arises due to endogeneity. Endogeneity is a structural concept - without assuming the existence of the structure, endogeneity, defined as correlation between an *observed* explanatory variable and the *unobserved* explanatory variable, would not be defined. For example, the potential outcomes approach does not specify unobserved element as deter-

[11] I adopt this title just to indicate that I follow the spirit of the book entitled, *Structural Analysis of Discrete Data*, not that this thesis is comparable with the great classic.

minants. The selection on unobservables under the potential outcomes framework can be understood as the endogeneity problem in the structural approach.

I adopt the structural approach to deal with the two econometric issues, namely, controlling for endogeneity, and recovering heterogeneity in responses even after conditioning on observables.

1.5.1 The Structural Approach, Causality and Endogeneity

In the structural approach where the outcome of interest is assumed to be determined by a structural relation, endogeneity of an explanatory is defined as dependence between the explanatory variable and the unobserved variables included in a structural relation. When the variable is chosen by individuals, the exogeneity of a variable would not be guaranteed because the uncontrolled unobserved individual heterogeneity is very likely to affect the decision of other observable explanatory variables. The information regarding endogeneity is contained in the joint distribution of the unobserved variables and the endogenous variable, which is not observed.

Causality - Objects of Interest

The causal effects are measured by partial derivatives or partial differences of the structural relation, $\frac{\partial h(x,u)}{\partial x}$, or $h(x^a,u)-h(x^b,u)$. To control for the endogeneity, the control function methods are used by assuming auxiliary equations specifying how the endogenous variables are determined by observed and unobserved arguments.

1.5.2 Nonseparable Structural Relations and Heterogeneity in Responses

Recovering more and more heterogeneity from data is desirable in the sense that one can derive more information from data. However, the recovered heterogeneity should be interpretable[12]. Conditioning on the observables is one obvious way of recovering heterogeneity, and using quantiles rather than focusing on the mean would produce more heterogeneity from the data as Bitler, Gelbach, and Hoynes (2003) argue. As many authors indicated[13], the existence of unobserved heterogeneity causes difficulties in data analysis especially with micro-data. Typical methods of modeling unobserved heterogeneity have been specifying it explicitly, then finding out a legitimate way of eliminating it.

Using additively nonseparable structural functions leads to random sensitivity. Under the nonseparability the approach - the quantile-based control function approach - taken in this thesis suggests an alternative way of modeling unobserved heterogeneity. The key implication of the nonseparable functional form is that partial derivatives or partial differences are themselves stochastic objects that have distributions, since partial derivative, $\frac{\partial h(x,u)}{\partial x}$, or $h(x^a,u)-h(x^b,u)$ are stochastic objects.

[12] The most heterogeneous form of information would be data themselves, which are not be interpretable. To derive an interpretable information from data we need the process of "data reduction" - a process of producing statistics such as mean, or quantiles etc from the data.

[13] See, for example, Heckman (2000), Blundell and Stoker (2005), Browning and Carro (2007), Matzkin (2007b), and Lewbel (2007a).

Nonseparable structural functions can be used to test the predictions in models under asymmetric information, which are often expressed in terms of a certain *unobserved type*. Since the responses under the nonseparable structural relations allow for heterogeneity in both observable and unobservable variables, the identification results can be beneficially used to falsify the predictions of the economic models using data.

Focusing on identification of the structural relation evaluated at different values of observable and unobservable variables is the way to recover heterogeneity in responses in the thesis.

1.6 Identification and "Measurement"

This thesis considers the "*identification*" problem by proposing justifiable restrictions. Identification results would allow for economic *interpretation (causality in this case)* under such restrictions. Identification matters especially when measurement is imperfect in recovering an object of economic interest as in individual's treatment effect example. However, *measurement also matters* - I emphasize the importance of devising novel methods to "measure" otherwise hidden information in the data collection stage via for example, experiments, or asking hypothetical questions.

Identification results provide *model-based evidence* and they justify certain interpretation of objects obtained from data reduction processes. Thus, the *credibility* of the evidence depends on the credibility of the model - characterized by a set of restrictions. If one could measure an object of economic interest, the issue of credibility of the model could be avoided[14]. I think the two lines of research could complement each other in recovering *evidence* from the population of concern.

1.7 Falsifiability of Econometric Models and Testability of Restrictions

The identifying power of a model comes from the restrictions imposed by the model and identification results would be believed to the extent that the restrictions are considered to be true. If one could test the restrictions using data, credibility of restrictions can be confirmed. Testability of restrictions is also informative in determining which minimum set of restrictions should be believed when they are not directly testable and in clarifying the limits of data analysis. However, some of the restrictions imposed on the structure may not be "directly" testable. In such cases identification results on the structure can provide a way to test the restrictions. I develop some principles of testability using the identification results and adopt one of them to test exogeneity.

[14] The issue would be then whether the new measurement device, for example, using an experiment or survey questionnairs, is justified.

1.8 In This Thesis - the Quantile-based Control Function Approach

The rest of the thesis is composed of as follows. The first chapter is introduction and Chapter 2 proposes formal frameworks of identifiability and testability of structural features allowing for set identification. The results in Chapter 2 are used in other chapters. The second section of Chapter 3, Chapter 4 and Chapter 5 contain new results.

Chapter 3 has two sections. The first section introduces the quantile-based control function approach (QCFA) proposed by Chesher (2003) to compare and contrast other results in Chapter 4 and 5. The second section contains new findings on the local endogeneity bias and testability of endogeneity. Chapter 4 assumes that the structural relations are differentiable and applies the QCFA into several models for discrete outcomes. Chapter 4 reports point identification results of partial derivatives with respect to a continuously varying endogenous variable. Chapter 5 relaxes differentiability assumption and apply the QCFA into an ordered discrete endogeneous variable. The model in Chapter 5 set identifies partial differences of a nonseprable structural function.

Chapter 2

Identifiability and Testability of Structural Features under Set Identification[1]

> *"Most of economic intuition is expressed in terms of the structure, so the structure is often the object of interest for estimation and for testing. .... the reduced form is convenient theoretically, but to be most useful, facts about it have to be translated back into structural statements...* (Kadane and Anderson (1977), p1028, quoted by Breusch (1986)"

Following the spirit of partial identification of features of *probability distributions*, pioneered by C. Manski (see Manski (1995) for economic examples that motivate partial identification, and Manski (2003) for a survey of recent developments), in this section formal definitions regarding partial identification of features of a *Hurwicz (1950a) structure*[2] are provided. A structure, as a tuple of structural relations and the distribution of the unobservables, has been used in many nonparametric identification studies. In this chapter set identification and sharpness of an identified set are formally defined by using the nonparametric Hurwicz (1950a) structure that can be applied to models with multiple equilibria. The logic of testability of structural features is discussed by extending earlier results of Koopmans and Reiersol (1950) into a general nonparametric setup allowing for set identification. Jovanovic (1989) modified Koopmans and Reiersol (1950)'s framework so as to deal with models with multiple equilibria and offers a general framework for statis-

[1] This chapter is motivated by A. Chesher's unpublished lecture note entitled "Evidence in Economics" in which *falsifiability* of a model is discussed using point identification.

[2] Although in many cases this distinction is not essential, it is required in the discussion of this thesis. If treatment responses are homogeneous among the observationally identical indivisuals, this case can be modelled using an additively separable structural relations. When the structural relations are not additively separable, for example, when treatment responses are likely to be heterogenous, conditional moment restrictions do not identify *structural* parameters. (See Hahn and Ridder (2009)). Also, "endogeneity" is defined by specifying the unobserved heterogeneity as determinants of an outcome. In the setup of Manski (2003) the unobserved heterogeneity is implicit.

tical inference in such models. Jovanovic (1989) noticed the possibility of set identification, however, "identification" in his paper is restricted to mean *point* identification.

When *strong, often parametric* assumptions that are hard to be justified by economic arguments are avoided, loss of identifying power of a model may result, in the sense that the value of a feature of the underlying economic process is not determined *uniquely* by data using prior information. In such a case, one may choose to impose alternative restrictions, but instead of imposing strong arbitrary parametric restrictions which may lead to point identification, one could search for weaker *justifiable* nonparametric restrictions that define a set in which the value of the feature can lie.

The identifying power of a model comes from the restrictions imposed by an econometric model. The credibility of restrictions should be discussed in each application of the model. However, not all restrictions have testable implications on the distribution of the observables (data). Even in such a case the restrictions can be tested when there exist two distinct models that identify the same structural feature with the one model *nested* by the other. The criteria proposed for testability or falsifiability involve comparison of the two identified sets - the identified set defined by the nested model should be smaller. Galichon and Henry (2009) develop a test of nonidentifying restrictions under the Jovanovic (1989) setup by explicitly allowing for partial identification. Galichon and Henry (2009)'s test can be used to falsify an econometric model. This chapter suggests an alternative framework of falsification of an econometric model, by which falsification of restrictions may be achieved.

2.1 Set Identification of Hurwicz (1950a) Structural Features

2.1.1 Elements of Identification

Distribution functions are denoted by F_A indicating the distribution function of A. $F_{A|B}$ indicates the conditional distribution of A given B. The corresponding $\tau-$quantiles are denoted by $Q_A(\tau)$ and $Q_{A|B}(\tau|b)$[3].

Economic processes are assumed to be generated by individual[4]'s decision mechanisms. The decision mechanisms are usually described as relationships between variables. I denote these underlying mechanisms as "*structures*" following Hurwicz (1950a). Hurwicz (1950a) assumed that the distribution of the observables is generated by a transformation $\mathcal{H}$ performed on the distribution of the unobservables, $F_{U|X}$ and defined the structure, S, as a tuple of the mapping ($\mathcal{H}$) and the distribution of the unobservables ($F_{U|X}$) where Y is a vector of endogenous[5] variables determined by the economic decision processes, X is a vector of covariates (exogenous variables) and U is a vector of unobserved elements to the

[3] The quantiles are defined by $Q_{A|B}(\tau|b) = \inf\{a|F_{A|B}(a|b) \geq \tau\}$

[4] By "individual" I mean any economic decision unit of interest.

[5] Following Koopmans (1949, p133), endogenous variables are "observed variables which are not known, or assumed to be statistically *not* independent of the latent variables, and whose occurrence in one or more equations of the set of equations is necessary on grounds of "theory"".

analyst. From a random sample of observations on Y and X the conditional distribution of Y given X is identified[6]. Denote $F^S_{Y|X}$ to be the distribution of the observables that is generated by a structure S.

Let $\mathcal{S}$ be a set of all structures. Let Ψ^U be a set of all distributions of unobservabels ($F_{U|X}$) and $\Psi^{\mathcal{S}}$ be the set of all distribution function of observables generated by elements in $\mathcal{S}$, that is, $\Psi^{\mathcal{S}} = \{F^S_{Y|X} | \ S \in \mathcal{S}\}$. Then $\mathcal{H}$ is a mapping from Ψ^U to $\Psi^{\mathcal{S}}$. More specifically, a Hurwicz (1950) structure, $S = \{\mathcal{H}, F_{U|X}\}$, and observed data ($F_{Y|X}$) have the following relation

$$\underbrace{F_{Y|X}}_{\text{Reduced form}} = \underbrace{\mathcal{H}(F_{U|X})}_{\text{Structure}}.$$

The mapping $\mathcal{H}(\cdot)$ is assumed to *uniquely*[7] determine the distribution of the observables. $\mathcal{H}(\cdot)$ can specify structural relations and can assume the existence of an equilibrium selection mechanism if there are multiple equilibria. However, this does not mean that the econometric model for any economic setup with multiple equilibria should specify the equilibrium selection mechanism. Only the existence of a selection mechanism is required to be assumed.

Let S_0 be the true structure that generates the distribution of observables available to us, $F^0_{Y|X}$. The two structures S and S' are called **observationally equivalent to each other** if $F^S_{Y|X} = F^{S'}_{Y|X}$. Define $\Omega_0 = \{S : F^S_{Y|X} = F^0_{Y|X}\}$, a set of structures that are observationally equivalent to the true structure, S_0 (note that $S_0 \in \Omega_0$ by definition of Ω_0 . See <Figure 2.1>.) The structural feature, $\theta(S)$ is defined as a functional of the structure. It can be an economic object that is important in policy design such as elasticities, risk attitudes, or time preferences etc. One of the main objectives of specifying an econometric model is to recover the true economic data generating structure S_0, or some features of it, $\theta(S_0)$.

The econometric model, $\mathcal{M}$, is characterized by a priori information (restrictions) applied to the structure and the distribution of the observables. The model, $\mathcal{M}$, is defined to be the set of the structures that satisfy the restrictions. Let $\Psi^{\mathcal{M}}$ be a set of all possible distribution functions generated by $S \in \mathcal{M}, \Psi^{\mathcal{M}} = \{F^S_{Y|X} | \ S \in \mathcal{M}\}$. S_0 is said to be point identifiable in $\mathcal{M}$ if there is no other member of $\mathcal{M}$ that is observationally equivalent to S_0. A structural feature $\theta(S_0)$ is said to be point identifiable if there is no variation in the values of the structural feature of the admitted structures. See Hurwicz (1950a), Koopmans and Reiersol (1950), Roehrig (1988), and Matzkin (1994, 2007) for general discussion of point identification. Matzkin (2007) reviews recent developments of nonparametric identification.

[6]The exact knowledge of $F_{Y|X}$ cannot be derived from any finite number of observations. Such knowledge is the limit approachable but not attainable by increasing the number of observations (Koopmans and Reiersol (1950)).

[7]Note that this uniqueness should be distinguished from the assumption that the *structural relations* uniquely determine values of endogenous variables given exogenous variables. When structural relations do not specify one-to-one mappings between the endogenous variables and unobserved variables given exogenous variables, this setup assums that there should be a mechanism that selects one point among many. This selection is possibly unknown, thus, unspecified by the analysist since there may not be a well defined and convincing way of doing it.

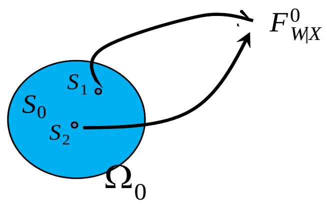


Figure 2.1: Note that $S_0 \in \Omega_0$, by definition of Ω_0. Suppose $S_1, S_2 \in \Omega_0$. Then $F^{S_1}_{Y|X} = F^{S_2}_{Y|X} = F^0_{Y|X}$. That is, S_1 and S_2 are observationally equivalent structures and they are indistinguishable from data - no amount of data can distinguish S_1 from S_2. Note that the true structure, S_0, that generates the distribution of the observables we have, is always in Ω_0.

Examples of Hurwicz (1950a) Structure

1. **Nonparametric Simultaneous Equations Models** of Roehrig (1988), Benkard and Berry (2006), or Matzkin (2008) : The mapping $\mathcal{H}$ is a system of structural functions, $U = G(Y, X)$, where Y, X, and U are defined as before. The dimension of Y needs to be the same as that of U. Their studies assume that the structural functions are unique valued, one-to-one mappings between Y and U, thereby excluding discrete endogeneous variables cases. The structural feature of interest, $\theta(S)$, can be the value of the structural function evaluated at a specific point, or partial derivatives of the structural functions.

2. **Treatment Effects** using Hurwicz (1950a) Structure : $\mathcal{H}$ can be a mapping from a set of joint distributions of the scalar potential outcomes, Y_1 and Y_0, $F_{Y_1Y_0|X}$, to $\Psi^{\mathcal{S}}$, such that
$$F_{Y|X} = \mathcal{H}(F_{Y_1Y_0|X}).$$
Note that $F_{Y_1Y_0|X}$ is unobeserved. $\theta(S)$ can be average of quantiles of treatment effects, $\theta(S) = E(Y_1 - Y_0|X = x)$, or $\theta(S) = Q_{Y_1-Y_0|X}(\tau|x)$.

3. **Models for Oligopoly Entry Games** : The mapping $\mathcal{H}$ can be structural relations together with an equilibrium selection mechanism. Let the structural relations be specified by the threshold crossing structures as $Y_1 = 1(X_1\beta_1 + Y_2\Delta_1 + U_1 \geq 0)$ and $Y_2 = 1(X_2\beta_2 + Y_1\Delta_2 + U_2 \geq 0)$ as in Bresnahan and Reiss (1991). This structural relations do not predict unique outcomes. By assuming a specific equilibrium selection rule, π, point identification can be achieved.

4. **Binary Choice Models** without endogeneity of Manski (1988) and Matzkin (1992) : $\mathcal{H}$ is a latent structural relation together with a threshold crossing structure that

transfoms the latent structural relation into the observed distribution. Manski (1988)'s case is $Y = 1(X\beta + U \geq 0)$, and Matzkin (1992)'s case is $Y = 1(h(X)+U \geq 0)$. The structural feature of interest in these papers are : $\theta(S) = \{\beta,\ F_{U|X}\}$ and $\theta(S) = \{h(x),\ F_{U|X}\}$, where x is a realized value of the random variable X. Under a set of restrictions, point identification of the structural features is established. Note that neither of the restrictions of the two models are included by the other. One of the common restrictions in the two models is the existence of a continuous explanatory variable (can be called a special regressor as in Lewbel (2000)), which is relaxed in Magnac and Maurin (2007, 2008).

5. **Auction models - the interpretation of $F_{U|X}$** : The data available are individuals' bids, and the underlying structure is $S = \{h, F_{U|X}\}$, where $F_{U|X}$ is the joint distribution of latent individual valuation and private information (types) satisfying certain statistical properties (e.g. independence and symmetry etc.) and h is a true mapping from the true distribution of types (U) to a distribution of observable bids (Y) implied by the assumption of Baysian Nash equilibrium. The structural relation h between bids and valuation is implied by the Bayesian Nash equilibrium, which is represented by $H(Y, X, U) = 0$. The different auction models assume different statistical properties on the bidder's private information, which opens room for testing the implications from different models using the observed data.

Lemma 1 in Chesher (2007) and Constructive Identification

The above definition is regarding *identifiability.* Lemma 1 in Chesher (2007) can be used for constructive identification. Lemma 1 in Chesher (2007) states that if there exists a unique-valued functional $\mathcal{G}(\cdot)$ of distribution of the observables such that $\theta(S) = \mathcal{G}(F_{Y|X}), \forall S \in \mathcal{M} \cap \Omega_0$, then the structural feature $\theta(S)$ is identified by $\mathcal{G}(F_{Y|X})$. $\theta(S) = \mathcal{G}(F_{Y|X})$ indicates the identifying relation.

Identifiability does not necessarily imply how to find out the form of $\mathcal{G}(F_{Y|X})$. Finding out the expression for the structural feature in terms of a functional of the distribution of the observables, $\mathcal{G}(F_{Y|X})$, can be useful because once the exact form is known by the identification result, then the analogy principle can be used.

Chapter 4 in which point identification of ceteris paribus effects of a continuous variable are discussed, Lemma 1 in Chesher (2007) is used to find out the identifying relation, $(\theta(S) = \mathcal{G}(F_{Y|X}))$.

2.1.2 Set Identification and Sharpness

Sometimes point identification is not achievable unless we impose stronger restrictions on the structure. If such strong, often parametric restrictions are hard to be justified in the context of an economic application, then we may try to obtain partial identification by imposing weaker restrictions instead of imposing unreasonable restrictions that guarantee point identification.

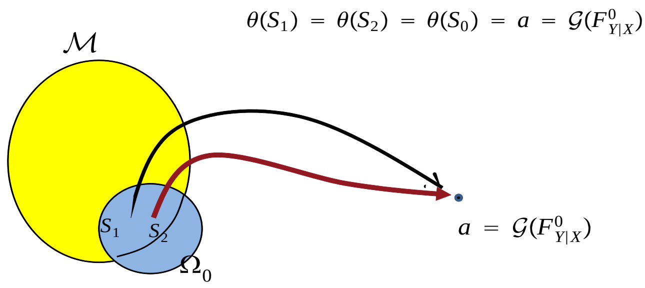


Figure 2.2: Point identification of a structural feature, $\theta(S)$ by Lemma 1 in Chesher (2007) : it is said to be point identified if there exists a unique valued functional $\mathcal{G}(\cdot)$ such that $\theta(S) = \mathcal{G}(F^0_{Y|X}), \forall S \in \mathcal{M} \cap \Omega_0$. (Lemma 1 in Chesher (2007)). In other words, if $S_1, S_2 \in \mathcal{M} \cap \Omega_0$, then $\theta(S_1) = \theta(S_2) = \theta(S_0) = a = \mathcal{G}(F^0_{Y|X})$. Thus, in this case, $\Theta^{\mathcal{M}}(F^0_{Y|X}) = \mathcal{G}(F^0_{Y|X}) = \{a\}$. The identification analysis will provides a way to find out the form of unique-valued functional $\mathcal{G}(F^0_{Y|X})$ to achieve point identification, or the form of the set, $\Theta^{\mathcal{M}}(F^0_{Y|X})$.

For a given econometric model, $\mathcal{M}$, defined in the previous section, let $\mathbf{\Theta}^{\mathcal{M}}(\cdot)$ be a mapping from $\Psi^{\mathcal{M}}$ to a class of sets in R^d, where d is the dimension of the structural feature, $\theta(S)$, specified in the economic example. The mapping is written as $\mathbf{\Theta}^{\mathcal{M}}(F^S_{Y|X})$. We say the model, $\mathcal{M}$, set identifies the structural feature, $\theta(S_0)$, if we can *determine* a set $\mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X})$, given data, $F^0_{Y|X}$, such that for **any** admitted structure S that is observationally equivalent to S_0, $\theta(S) \in \mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X})$. $\mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X})$ can be defined either explicitly with the boundary explicitly specified (Manski (1990,1997), Manski and Pepper (2000), Chesher (2005), and Lee (2010)), or implicitly by some moment inequalities (some of the entry models (see Berry and Tamer (2007) for survey), Honore and Tamer (2006), Magnac and Maurin (2008), and Chesher (2010), for example).

Since several studies define their identified sets as those that may contain outer regions[8], an identified set is defined as a bigger set that contains a sharp identified set, which will be defined later. An identified set is defined as the following :

Definition 1 Set Identification : the model $\mathcal{M}$ **set identifies** the structural feature, $\theta(S_0)$ if $\exists\, \mathbf{\Theta}^{\mathcal{M}}(\cdot)$ s.t. $\forall S \in \mathcal{M} \cap \Omega_0$, $\theta(S) \in \mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X})$, where Ω_0 is defined as before.

Definition 2.1 Set Identification : the model $\mathcal{M}$ **set identifies** the structural feature, $\theta(S_0)$ if $\exists\, \mathbf{\Theta}^{\mathcal{M}}(\cdot)$ s.t. $\forall S \in \mathcal{M} \cap \Omega_0$, $\theta(S) \in \mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X})$, where Ω_0 is defined as before.

[8] An identified set may contain some *outer regions* as discussed in Beresteanu, Molchanov, and Molinari (2008).

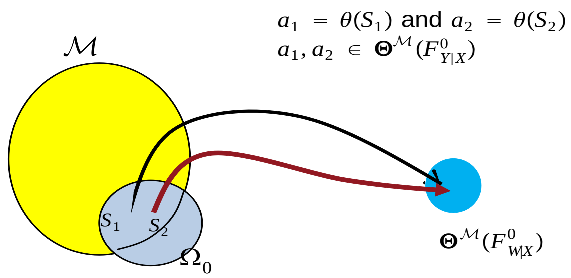


Figure 2.3: When the structural feature is set identified, and not point identified, then for any two admitted and observationally equivalent structures, $S_1, S_2 \in \mathcal{M} \cap \Omega_0$, with $a_1 = \theta(S_1)$ and $a_2 = \theta(S_2)$, all we can say is that $a_1 \in \Theta^{\mathcal{M}}(F^0_{Y|X})$ and $a_2 \in \Theta^{\mathcal{M}}(F^0_{Y|X})$, a_1 and a_2 can be distinct values.

If $\mathbf{\Theta}^{\mathcal{M}}(\cdot)$ is singleton, in other words, if every admitted and observationally equivalent structure generates the same value of the structural feature, then we say the structural feature is point identified by the model. See <Figure 2.2>.

Failure of point identification and set identification (examples continued)

1. **Nonparametric and Nonseparable Simultaneous Equations Models** : Identifiability results in Matzkin (2008) cannot be applied when any of endogenous variables is discrete since differentiability and the one-to-one mapping assumption between the unobservables and the discrete endogenous variable do not hold. Other (point) identification strategies using a triangular system or single equation IV models have been proposed. For a triangular system, see Chesher (2003) and Imbens and Newey (2009), and for single equation IV models see Chernozhukov and Hansen (2005). However, when the regressor is discrete under triangular systems (see Chesher (2005), Jun, Pinkse, and Xu (2010), and Lee (2010) for discrete endogenous regressor) and when the outcome is discrete in single equation IV models (see Chesher (2010)), point identification fails.

2. **Treatment Effects** : When parametric assumptions on the distribution function are relaxed, strong restrictions such as identification at infinity (see Heckman (1990)) are required for point identification of average treatment effects. Several studies report partial identification results under weaker restrictions : see Manski (1990,1997) and Heckman and Vytlacil (2001), Manski and Pepper (2000), Shaikh and Vytlacil (2005), and Bhattacharya, Shaikh, Vytalcil (2008).

3. **Models for Oligopoly Entry Games :** without specifying an equilibrium selection mechanism point identification in the entry models is not achievable. See Tamer (2003) and Berry and Tamer (2007) for recent survey.

4. **Binary Choice Models with Endogeneity** : When the large support condition in Lewbel (2000)'s model with a special regressor is relaxed, Magnac and Maurin (2008) show partial identification results using the moment conditions derived from their restrictions.

If a set, $\mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X})$, includes **all** the values of a feature of structures that are admissible and observationally indistinguishable and if it contains **only** such values, then $\mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X})$ is called a sharp identified set.

Definition 2.2 A sharp identified set, $\mathbf{\Theta}^{\mathcal{M}}_{Sharp}(F^0_{Y|X})$ is defined as $\mathbf{\Theta}^{\mathcal{M}}_{Sharp}(F^0_{Y|X}) \equiv \{a : \theta(S) = a, \forall S \in \mathcal{M} \cap \Omega_0\}$.

To show set identification, it needs to be shown that an identified set, $\mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X})$, contains *all* the values of a feature of structures that are observationally equivalent and admitted by $\dot{\mathcal{M}}$. However, not every point in $\mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X})$ is necessarily generated by an admitted structure that is observationally equivalent. (e.g. Andrews, Berry, and Jia (2004), Ciliberto and Tamer (2009)). Alternatively, a set defined by an identification strategy may not include *all* the points that are generated by admitted and observationally equivalent structures.

Beresteanu, Molchanov, and Molinari (2008) define a sharp identified region as

> "... the region in the parameter space which includes *all* possible parameter values that (i) could generate the same distribution of observables for some data generation process (ii) *consistent* with the maintained modeling assumptions and no other parameter value, is called the sharp identified region. .." .

This is a descriptive definition of sharpness. This descriptive definition can be mapped into Definition 2.2 because $\mathbf{\Theta}^{\mathcal{M}}_{Sharp}(F^0_{Y|X})$ is the set of all values of the structural feature, $\theta(S)$, that is generated by an element in $S \in \mathcal{M} \cap \Omega_0$. "Consistent with the model ($S \in \mathcal{M}$)" and "generate the same the distribution of the observables ($S \in \Omega_0$)" can be guaranteed by the fact that $S \in \mathcal{M} \cap \Omega_0$.

If one can show that $\mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X}) = \mathbf{\Theta}^{\mathcal{M}}_{Sharp}(F^0_{Y|X})$, then sharpness of an identified set is shown. Another way of showing sharpness is to use the following lemma. Suppose that for every value in an identified set, there exists an admitted and observationally equivalent structure whose feature is that value, then the identified set is sharp.

Lemma 2.1 Suppose that $\forall\ a \in \mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X}), \exists S \in \mathcal{M} \cap \Omega_0$ with $\theta(S) = a$. Then $\mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X}) = \mathbf{\Theta}^{\mathcal{M}}_{Sharp}(F^0_{Y|X})$.

Proof. Suppose that $\forall\ a \in \mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X}), \exists S \in \mathcal{M} \cap \Omega_0$ with $\theta(S) = a$. First, note that $\mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X}) \subseteq \mathbf{\Theta}^{\mathcal{M}}_{Sharp}(F^0_{Y|X})$, since for any $a \in \mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X}), \exists S \in \mathcal{M} \cap \Omega_0$ with $\theta(S) = a$, it should be the case that $a \in \mathbf{\Theta}^{\mathcal{M}}_{Sharp}(F^0_{Y|X})$. Next, $\mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X}) \supseteq \mathbf{\Theta}^{\mathcal{M}}_{Sharp}(F^0_{Y|X})$ since

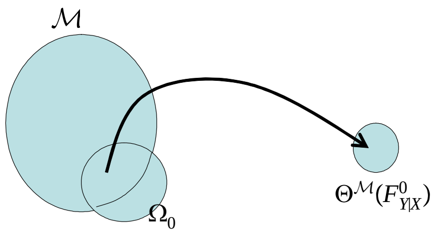


Figure 2.4: Set identification : all the values of the structural feature $(\theta(S))$ generated by any structure that is admitted by the model and observationally equaivalent to the true structure need to lie in the set $\Theta^{\mathcal{M}}(F^0_{Y|X})$. If the structural feature is point identified, that is, $\Theta^{\mathcal{M}}(F^0_{Y|X})$ is singleton, distinct structures that are admitted and observationally equivalent should generate the same value for the structural feature. Note that there can be some parts of the set, $\Theta^{\mathcal{M}}(F^0_{Y|X})$, where $\theta(S)$ never lies. Sharpness of an identified set guarantees that there will be no such parts, in which case, the set can be described as "the smallest set that exhausts all the information from the data and the model" as some authors define sharpness. Two distinct points in the identified may have been generated by two distinct structures, but they should be admitted (consistent with the model) as well as observationally equivalent to each other (consistent with the data) if the identified set is sharp.

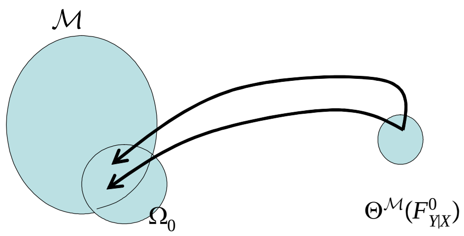


Figure 2.5: Sharpness : showing sharpness involves showing that for each point in the set there exists at least one structure that is admitted (consistent with the model) and observationally equivalent (consistent with the data) to the true structure, S_0. Note that two distinct structures could generate the same value for the structural feature.

$\mathbf{\Theta}^{\mathcal{M}}_{Sharp}(F^0_{Y|X})$ is the smallest identified set without any outer region. Then $\mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X}) = \mathbf{\Theta}^{\mathcal{M}}_{Sharp}(F^0_{Y|X})$ follows. ■

Discussion on Sharpness (examples continued)

1. **Nonparametric Structural Analysis with Discrete Data** : Chesher (2010) and Chapter 5 in this thesis show sharpness of their identified sets. The proofs involve showing that for every point in the identified set, there exists *at least one* admitted and observationally equivalent structure

2. **Treatment Effects** : Heckman, Clement, and Smith (1997), Shaikh and Vytlacil (2005), Fan and Park (2010), and Firpo and Ridder (2009) study distribution of treatment effects defined as difference between the potential outcomes, $Y_1 - Y_0$. Since Hurwicz (1950) mapping, $\mathcal{H}$, can be considered to transform the distribution of $Y_1 - Y_0$ (unobservable) into the distribution of Y (observable), their sharpness proofs involve construction of $F_{Y_1-Y_0|X}$ from the observed distribution, $F_{Y|X}$. For the constructed distribution, $F_{Y_1-Y_0|X}$ to be legitimate, it has to satisfy the properties of distribution functions.

3. **Entry Models** : Ciliberto and Tamer (2007) recognize that the inequality restrictions taken in the entry game do not generate sharp identified sets.

4. **Monotone Binary Choice Models :** Magnac and Maurin (2008)'s identified set is defined as a set of all the points that are observationally equivalent and that satisfy the moment restriction derived in their papers. This is enough since they showed that their moment conditions equivalently express all the restrictions imposed by the model, thus, all the observationally equivalent structures that satisfy the moment conditions should be those that are admitted by the model.

2.1.3 Overidentification, Intersection Bounds and Sharpness

A model defines an identified set and sharp identified sets always exist once a model is given. There can be many such sets (overidentification)[9], for example, where different *values* of IV define different identified sets in Chesher (2005), or where there exist many *IVs* that satisfy the moment conditions in Bontemps, Magnac, and Maurin (2008). Not every identified set is sharp. Also, even though a model may define several identified sets, intersection of them does not guarantee sharpness since every identified set may contain some common *outer regions.*

2.1.4 Overidentification and Specification Tests under Set Identification

The information when there is overidentification, tests regarding the specification of the model can be conducted.

[9] This terminology, "overidentification" in the partial identification context was used in Chesher (2005) and Bontemps, Magnac and Maurin (2007).

Examples

1. **Treatment Effects** : Manski (1990) discusses testability (refutability) of "level-set" assumption - constant treatment effect assumption across different observable characteristics - by taking intersection of each identified intervals and see if the intersection is empty.

2. **Overidentification Tests under Set Identification** : Bontemps, Magnac, and Maurin (2008) develop a Sagan-type specification test of overidentifying restrictions in the case of overidentification when the parameter is partially identified.

2.2 *Refutability* of Structural Features under Set Identification

> "...particularly where the model is to a large degree speculative, empirical confirmation of the validity or usefulness of the model is obtained only to the extent that *observationally restrictive* specifications are upheld by the data....(Koopmans and Reiersol (1950) p180, emphasis added)"

If a model (or the restrictions imposed by a model) can be confirmed by data, this will validate the usefulness and credibility of the model. Some features of an economic model may not be identifiable in which case no amount of empirical information will answer the questions regarding the features of the underlying economic decision processes. To be able to use data as evidence for or against any hypothesis regarding the underlying structure, identifiability of the structural feature which is the object of the hypothesis is essential. The general rule described in the above quote from Koopmans and Reiersol (1950) and Breusch (1986)'s observation on "testability" are adopted.

Let $\mathcal{F}$ be a set of all restrictions and $\mathcal{F}^{YX}$ be a set of restrictions on the distribution of the observables. Then $\mathcal{F}$ would be partitioned into the two sets, $\mathcal{F}^{YX}$ and $\mathcal{F}/\mathcal{F}^{YX}$. Let R denote an element of the set $\mathcal{F}$. R can be a statement regarding either a structure or a distribution of observed variables.

Examples of restrictions on the structure can be monotone treatment response or monotone selection restriction in Manski (1997) and Manski and Pepper (2000). They can be regarding the functional form such as additive separability, linearity, or regarding he distribution of the unobservables such as mean or quantile. In most econometric models restrictions on the structure are not enough for identification.

Often restrictions on the distribution of the observables should be required. Examples of restrictions on the distribution of the observables are various types of rank conditions, or no multicollinearity condition or completeness conditions in Newey and Powell (2003) or Chernozhukov and Hansen (2005). Sometimes existence of a continuous variable plays a key role in identification - identification at infinity, special regressor in Lewbel (2000) (which can be a special case of Manski (1988) and Matzkin (1992)'s conditions for identification). In principle, any such restrictions on observables should be checked whether they are satisfied by the data. that is, $R \in \digamma^{YX}$ is "directly testable", while $R \in \digamma / \digamma^{YX}$ is testable if we can derive an equivalent expression for the restriction in terms of the observable $F^S_{Y|X}$. Any $R \in \digamma^{YX}$ is directly testable, while $R \in \digamma / \digamma^{YX}$ is **testable (confirmable)** if and only if $\exists R' \in \digamma^{YX}$ s.t. $R \Leftrightarrow R'$. Note that some of the restrictions have no testable implications on data.

For the discussion of testability of $R \in \digamma / \digamma^{YX}$ and how to interpret the test results, we adopt Breusch (1986)'s framework. From now on we are concerned with restrictions on structural features which do not have any implications on the distribution of observables.

2.2.1 Breusch (1986)'s Framework of "Testability"

We introduce Breusch (1986)'s framework to determine "testability" of hypotheses on S. Let $\mathcal{H}_0$ be the set of structures that satisfy the null hypothesis. Then $\mathcal{S}$, a set of all structures, is partitioned into two, $\mathcal{H}_0$ and $\mathcal{S}/\mathcal{H}_0$. The testability of the hypothesis is a decision problem regarding whether the true structure S_0 that generates the data is included in $\mathcal{H}_0$ or not using the data.

Determining how to "interpret" the test results structurally requires further clarification of ideas. We adopt the "refutability" and "confirmability" from Breusch (1986) and define them as the following in a general setup. A hypothesis is refutable if, when the true structure, S_0, is not included in $\mathcal{H}_0$, every observationally equivalent structure to S_0 is also not included in $\mathcal{H}_0$.

Definition 2.3 A hypothesis is called refutable if $S_0 \notin \mathcal{H}_0 \Longrightarrow \nexists S \in \mathcal{H}_0 \; s.t. F^0_{Y|X} = F^S_{Y|X}$.

A hypothesis is confirmable if, when the true structure, S_0 is included in $\mathcal{H}_0$, every observationally equivalent structure to S_0 is also included in $\mathcal{H}_0$.

Definition 2.4 A hypothesis is called confirmable if $S_0 \in \mathcal{H}_0 \Longrightarrow \nexists S \notin \mathcal{H}_0 \; s.t. F^0_{Y|X} = F^S_{Y|X}$.

Discussion :

1. A hypothesis is **refutable** if when it is rejected, we can use the data as evidence against the null hypothesis and conclude that the hypothesis is not true. A hypothesis is **confirmable** if when it is accepted, we can conclude that the model (hypothesis) is true.

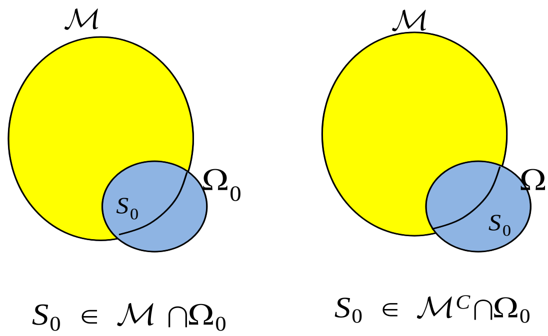


Figure 2.6: **Falsifiability of a model** : this is a problem of deciding whether $S_0 \in \mathcal{M}$, or $S_0 \in \mathcal{M}^C$. This can be restated as $S_0 \in \mathcal{M} \cap \Omega_0$, or, $S_0 \in \mathcal{M} \cap \Omega_0$, since $S_0 \in \Omega_0$ by definition of Ω_0.

2. However, if a hypothesis is **refutable, but not confirmable**, then we cannot conclude that the hypothesis is true even though the hypothesis is not rejected.

In Chapter 3 a refutable implication of endogeneity is discussed. Endogeneity is a structural feature that is not directly observable. By imposing some restrictions, a refutable implication can be derived.

2.3 Falsifiability of a Model

Econometric models characterized by restrictions are used to infer certain information on the true data generating structure, $\theta(S_0)$. Identification analysis assumes that S_0 is in the model, i.e. S_0 satisfies all the restrictions imposed by the model. Otherwise, the identified set by the model would not be informative on $\theta(S_0)$. In this section testability of whether the true structure actually lies in the model ($S_0 \in \mathcal{M} \cap \Omega_0$) is considered. This is a problem of deciding whether $S_0 \in \mathcal{M}$, or $S_0 \in \mathcal{M}^C$.This can be restated as $S_0 \in \mathcal{M} \cap \Omega_0$, or, $S_0 \in \mathcal{M} \cap \Omega_0$, since $S_0 \in \Omega_0$ by definition of Ω_0. See <Figure 2.6>. In this section one way of falsifying a model is discussed.

Let $\mathcal{M}^1$ be a set of structures that satisfy the set of restrictions R^1 and $\mathcal{M}$ be a set of structures that satisfy the set of restrictions $R^{\mathcal{M}}$. Then a model $\mathcal{M}'$ imposing restrictions R^1 and R can be written as

$$\mathcal{M}' = \mathcal{M} \cap \mathcal{M}^1 \qquad (**)$$

Let $\Psi^{\mathcal{M}}$ be a set of distribution functions of observables generated by the structures in $\mathcal{M}$ and $\Psi^{\mathcal{M}'}$ be a set of distribution functions of observables generated by the structures

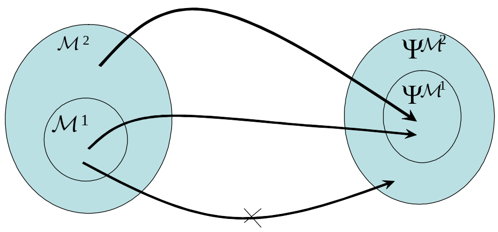


Figure 2.7: **Lemma 3**

in $\mathcal{M}'$.

Koopmans and Reiersol (1950) state that R^1 is "subject to test" if we can test H_0 : $F^0_{Y|X} \in \Psi^{\mathcal{M}'}$. From this we further develop the logic of testability of restrictions and discuss how identification results can be used in "falsifying" a model/restrictions.

Suppose that a model, $\mathcal{M}^1$, identifies a structural feature, $\theta(S)$, by a set $\mathbf{\Theta}^1(F^S_{Y|X})$, and another model, $\mathcal{M}^2$, identifies the same structural feature, $\theta(S)$, by $\mathbf{\Theta}^2(F^S_{Y|X})$. Recall that we define $\Psi^{\mathcal{M}^1} = \{F^S_{Y|X} : S \in \mathcal{M}^1\}$ and $\Psi^{\mathcal{M}^2} = \{F^S_{Y|X} : S \in \mathcal{M}^2\}$. Note that $\Psi^{\mathcal{M}^1 \cap \Omega_0} = \{F^0_{Y|X}\}$.

If $F^0_{Y|X} \notin \Psi^{\mathcal{M}}$, the model $\mathcal{M}$ should be falsified, since the true structure, S_0, that generates $F^0_{Y|X}$ cannot be in $\mathcal{M}$. Falsification of a model is not always possible.

Lemma 2.3 If $\mathcal{M}^1 \subseteq \mathcal{M}^2$, then $\Psi^{\mathcal{M}^1} \subseteq \Psi^{\mathcal{M}^2}$.

Proof. Trivial by definition of $\Psi^{\mathcal{M}}$. ■

Lemma 2.4 If $\Psi^{\mathcal{M}^1} \subseteq \Psi^{\mathcal{M}^2}$, then $\mathbf{\Theta}^1(F^S_{Y|X}) \subseteq \mathbf{\Theta}^2(F^S_{Y|X})$, for $\forall S \in \mathcal{M}^1 \cap \Omega_0$.

Proof. For $S^* \in \mathcal{M}^1 \cap \Omega_0$, $F^{S^*}_{Y|X} = F^0_{Y|X} \in \Psi^{\mathcal{M}^1} \subseteq \Psi^{\mathcal{M}^2}$, and $\theta(S^*) \in \mathbf{\Theta}^1(F^{S^*}_{Y|X})$. Since $\Psi^{\mathcal{M}^1} \subseteq \Psi^{\mathcal{M}^2}$, whenever $F^S_{Y|X} \in \Psi^{\mathcal{M}^1}$ implies that $F^S_{Y|X} \in \Psi^{\mathcal{M}^2}$. Then the definition of $\Psi^{\mathcal{M}^2}$ and identification imply that $S^* \in \mathcal{M}^2 \cap \Omega_0$, thus, $\theta(S^*) \in \mathbf{\Theta}^2(F^0_{Y|X})$ leading to the conclusion that $\mathbf{\Theta}^1(F^0_{Y|X}) \subseteq \mathbf{\Theta}^2(F^0_{Y|X})$. ■

Theorem 2.1 If $\mathcal{M}^1 \subseteq \mathcal{M}^2$, then $\mathbf{\Theta}^1(F^S_{Y|X}) \subseteq \mathbf{\Theta}^2(F^S_{Y|X})$, $\forall S \in \mathcal{M}^1 \cap \Omega_0$.

Proof. The result follows from Lemma 2.3 and Lemma 2.4. ■

Theorem 2.1 is a natural and intuitive result. Consider the following examples.

Example 1 suppose that $\mathcal{M}^1$ imposes linearity with mean independence of the unobserved U, so that the structural relation admitted is of the form, $Y = X\beta + U$, and $\mathcal{M}^2$ admits additively separable structural relation $Y = f(X) + U$,with mean independence of U. Then $\mathcal{M}^1 \subseteq \mathcal{M}^2$. If the true structure lies in $\mathcal{M}^1$, by **Theorem 2.1** for the structural

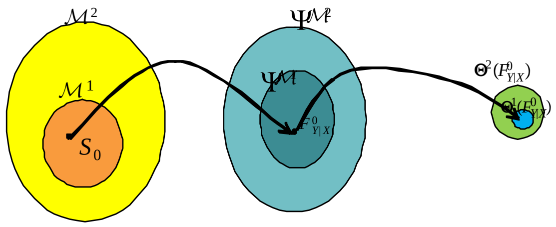


Figure 2.8: **Lemma 4**

feature of partial derivative, $\beta = f'(X)$, implying that $\mathbf{\Theta}^1(F^0_{Y|X}) = \mathbf{\Theta}^2(F^0_{Y|X})$ since both models point identify the partial derivative. If the identified sets did not intersect, the linearity restriction could be refuted.

Example 2 Consider Manski (1990), Manski (1997)'s Monotone Treatment Response (MTR) model, and Manski and Pepper (2000)'s Monotone Treatment Response and Monotone Treatment Selection (MTR-MTS) model. Let $\mathcal{M}^1, \mathcal{M}^2$, and $\mathcal{M}^3$ denote each model. Then $\mathcal{M}^1 \supseteq \mathcal{M}^2 \supseteq \mathcal{M}^3$. **Theorem 2.1** implies that if the true structure satisfies MTR-MTS restrictions, that is, the true structure lies in $\mathcal{M}^3$, we have

$$\mathbf{\Theta}^1(F^0_{Y|X}) \supseteq \mathbf{\Theta}^2(F^0_{Y|X}) \supseteq \mathbf{\Theta}^3(F^0_{Y|X}), \quad \forall S \in \mathcal{M}^3 \cap \Omega_0$$

Note that both MTR and MTS are *not* "directly testable". However, it can be said that if $\mathbf{\Theta}^1(F^0_{Y|X}) \not\supseteq \mathbf{\Theta}^2(F^0_{Y|X})$, then MTR is violated. Likewise, $\mathbf{\Theta}^2(F^0_{Y|X}) \not\supseteq \mathbf{\Theta}^3(F^0_{Y|X})$, then MTS is violated. If $\mathbf{\Theta}^1(F^0_{Y|X}) \not\supseteq \mathbf{\Theta}^3(F^0_{Y|X})$, then either MTR or MTS, or both MTR and MTS are violated.

In **Theorem 2.1** at least one model - either $\mathcal{M}^1$ or $\mathcal{M}^2$ - is overidentifying. However, existence of an overidentifying model is not required to falsify a model. As long as $\Psi^{\mathcal{M}^1} \subseteq \Psi^{\mathcal{M}^2}$ the criterion can be used to falsify $\mathcal{M}^1$ by **Lemma 2.4**. That is, although the two models are just-identifying, if $\Psi^{\mathcal{M}^1} \subseteq \Psi^{\mathcal{M}^2}$, we can falsify $\mathcal{M}^1$.

2.4 Just-identifying Models and Falsifiabilty of Restrictions

We first define just-identification. A just-identifying model loses its identifying power if any of its restrictions is relaxed.

Definition 2.5 A model $\mathcal{M}$ characterized by a set of restrictions R^M **just-identifies** a structural feature $\theta(S)$ if $\nexists \mathcal{M}_1$ characterized by a set of restriction R^1 s.t. (i) $R^M \supset R^1$ and (ii) $\theta(S) \in \mathbf{\Theta}^{\mathcal{M}_1}$, $\forall S \in \mathcal{M}_1 \cap \Omega_0$, where Ω_0 is defined as before.

The set of restrictions, R^M of a model $\mathcal{M}$ which is just-identifying is a minimal set that identifies the structural features.

If a model, $\mathcal{M}$, characterized by a set of restrictions R^M is **not** just-identifying, then there exists a less restrictive model $\mathcal{M}_1$ characterized by a set of restriction R^1 s.t. (i) $R^M \supset R^1$ and (ii) $\theta(S) = \theta(S_0)$ $\forall S \in \mathcal{M}_1 \cap \Omega_0$.

Suppose two models, $\mathcal{M}_1$ and $\mathcal{M}_2$, with $\mathcal{M}_1 \neq \mathcal{M}_2$ are just-identifying the same structural feature. Let the set of restrictions for $\mathcal{M}_1$ be R_1 and that for $\mathcal{M}_2$ be R_2. If $\Psi^{\mathcal{M}^1} \subset \Psi^{\mathcal{M}^2}$, then $\mathcal{M}^1$ can be falsified. In other words, if $\mathcal{M}^1$ is **observationally more restrictive**, and R_1/R_2 is **observationally relevant restrictions**. Then we have the following main result of this chapter.

Definition 2.6 (Koopmans and Reiersol (1950)) $\mathcal{M}^1$ is called **observationally restrictive** if $\Psi^{\mathcal{M}^1} \subset \Psi^{\mathcal{M}^2}$.

Definition 2.7 R_1/R_2 is called a set of **observationally relevant restrictions** if $\Psi^{\mathcal{M}^1} \subset \Psi^{\mathcal{M}^2}$.

Falsifiability of a model can be linked to refutability of restrictions under set identification since a model is characterized by restrictions. If a model is falsified, then some of the restrictions imposed by the model must not be the true description of the true underlying data generating structure. However, which restrictions among all the restrictions imposed by the model are not clear. This can be determined by the following proposition.

Proposition 2.1 If $\Psi^{\mathcal{M}_1} \subset \Psi^{\mathcal{M}_2}$, R_1/R_2, the observationally relevant restrictions, can be refuted.

Proof. The result follows from **Lemma 2.4**. ■

Chapter 3

The Quantile-based Control Function Approach and Testability of Endogeneity

In the first section of this chapter the quantile-based control function approach (QCFA) proposed in Chesher (2003) is introduced, which is a necessary background for later discussion. Section 3.1 is to compare and contrast the results in Chapter 4 and Chapter 5 with Chesher (2003). A simplified derivation of the Chesher (2003) results is introduced in a simple setup as a benchmark. Similar steps will be used in deriving the results in Chapter 4. Section 3.2 has new findings regarding testable implications on endogeneity.

Every variable is assumed to be continuously varying and the structural functions are assumed to be differentiable in this chapter. This assumption is relaxed in Chapter 4 and Chapter 5 where the same QCFA is applied to discrete outcomes and discrete endogeneous variables.

3.1 The Quantile-based Control Function Approach - A Revisit

3.1.1 The Model

The Chesher (2003) setup can be described by the following Restriction A. For simplicity, the case where there is only one endogenous variable is considered. Capital letters indicate random variables and the lower cases indicate their realization. Some variation of this restriction is used to reflect the nature of observational processes of each case in Chapter 4 and Chapter 5.

Restriction A - Triangularity, Continuous Variables, Strict Monotonicity and Differentiability

Scalar random variables W and Y, and a random vector, X of dimension K *are continuously distributed. For any values of X, U, and V, unique values of W and Y are* *determined by the structural equations*

$$W = h(Y, X, U) \quad \text{(S-1)}$$
$$Y = h^Y(X, V) \quad \text{(S-2)}$$

The scalar unobserved indices, U, and V *are jointly continuously distributed and each is normalized uniformly distributed on (0,1). The structural relations* h and h^Y *are strictly monotonic with respect to variation in the unobservable* U and V each. *The structural relations* h and h^Y *are differentiable.*

Chesher (2003) focuses on identification at a point and derives minimum possible restrictions that achieve local identification. Thus, the restrictions suggested are required to hold at the point. Based on Restriction A, further restrictions such as independence or exclusion restrictions will be used to derive identification results in the following discussion.

In the next subsection, some of the implications of this restriction are discussed.

3.1.2 Discussion on Restriction A

Additively Nonseparable Structural Function and *Stochastic* Sensitivities

One of the key implications of the nonseparable functional form is that partial derivatives or partial differences are themselves stochastic objects that have distributions, since $\frac{\partial h(Y,X,U)}{\partial y}$ or $h(y^a, x, u) - h(y^b, x, u)$ contain unobserved heterogeneity. If the structural function is linear, that is, $W = a + bY + cX + U$, then the partial derivative of this linear function with respect to Y is b. Thus, assuming a linear structural relation corresponds to assuming "homogenous" responses. On the other hand, an additively separable structural function, for example, $W = h(Y, X) + U$, allows for heterogeneity in responses, but once conditioning on the observables, there is no difference among the people with different unobserved characteristics as the ceteris paribus effect measured by the partial derivative, $\frac{\partial h(y,x)}{\partial y}$, is determined by observed characteristics only.

Triangularity, Continuous Endogenous Variables, and the Control Function Methods

Triangular[1] simultaneous equations models have been used under the name of "control function approach" when the endogenous variable is continuous[2]. The control function approach is usually used to indicate the way of correcting for endogeneity by adding the residuals from the auxiliary equations for the endogenous variables.

[1] Triangular simultaneous equations systems exclude many interesting economic examples, where outcomes are determined strategically by agents, much studied in the empirical IO literature recently.

[2] See Blundell and Powell (2003, 2004) for the most recent survey of the control function approach - the extensions of the control function approach to nonparametric and semiparametric structural equations for binary/censored outcomes. Note that their treatment is for the continuous endogenous variables.

In the identification analysis of structural functions under endogeneity invertibility of the structural function with respect to the unobservable variable plays a key role for point identification. Invertibility of h^Y with respect to V guarantees the one-to-one mapping between the endogenous variable (Y) and the error (V). When the endogenous variable is continuous, conditioning on the residual obtained from the equation for the endogenous variable is equivalent to conditioning on the specific-quantiles of the endogenous variable due to this *one-to-one mapping.* We call the latter strategy of conditioning on quantiles of the endogenous variables the quantile-based control function approach (QCFA) in contrast with the former strategy of conditioning on the residual.

Scalar Index Unobservables, *Continuous* Unobserved *Types*, and Monotonicity

Only scalar unobserved characteristics are allowed as an argument of the structural relations. The model also admits multiple factors of unobserved heterogeneity as long as they affect the outcome through a scalar index.

Restriction Scalar Index Unobservables (SIU) : *U and V should be scalar in the model. This model admits such cases that* $U = \theta_U(U_1, ..., U_L)$, $V = \theta_V(V_1, ..., V_I)$, *where* $\theta_U : R^L \to (0,1)$, $\theta_V : R^I \to (0,1)$, *for some positive number L and I. Let* $X = [X_1, X_2]'$.

Each unobserved variable is normalized uniform (0,1) and they are assumed to be continuous. This assumption can be natural and general in modeling contract theory in which individuals' unobserved type is assumed to be uniformly distributed.

However, this scalar unobserved index assumption does not admit measurement error models or duration outcomes. For structures with vector unobservables that can**not** be represented by a scalar unobservable, see Chesher (2009), where examples of such case are illustrated. The vector of unobservables is called "excess heterogeneity" in Chesher (2009) - "excess" in the sense that we allow for more unobservable variables than the number of endogenous variables. The distinction of the number of endogenous variables from the number of unobservable variables stems from the analysis of classical simultaneous equations models of the Cowles Commission, and more recent studies on nonparametric identification of simultaneous equations models in Brown (1983), Roehrig (1988), Matzkin (2008), and Benkard and Berry (2006), where the number of unobservables is equal to the number of endogenous variables.

Heterogeneity in sensitivity is recovered by adopting "*quantile*"-based methods, rather than averaging the unobserved characteristics out. Monotonicity of the structural relations in scalar unobserved element is required to use the equivariance property of quantiles. The monotonicity assumption can be justified in many economic examples - see Imbens and Newey (2010) for the examples that justify monotonicity.

3.1.3 Fundamental Identifying Relations - (B), (C), and Derivation of (AB)

Throughout the thesis the focus is on identification of the *sensitivity such as price/income elasticity* of an (endogenous) explanatory variable. The sensitivity or ceteris paribus impacts are measured by partial derivatives/differences of a structural relation, e.g. demand/supply function. To understand and derive the identification of partial derivatives/differences, identification of the structural relations should be understood first.

In this subsection identifying relations are established. The results will be referred to throughout the thesis in the analysis of the identification of the partial derivatives (Chapter 4), partial differences (Chapter 5), and the construction of the distribution of the unobservables in sharpness proofs of Chesher (2005) and Theorem 5.2 (Chapter 5) in Appendix C.

It is impossible to identify the whole structure, $\{h, F_{U|X}\}$ even without endogeneity (Lemma 1 in Matzkin (2003)) due to nonseparability : normalization of $F_{U|X}$ is required for identification of the structural function, h.[3] Then the major focus is on the identification of normalized structural functions, "structural quantile functions", named by Chernozhukov and Hansen (2005). When there is endogeneity problem, Matzkin (2003)'s idea fails to provide identification of independent variations in each argument of the structural function. In such a case Chesher (2003)'s QCFA can be used to identify independent variations in observable arguments in the structural function when the unobservable argument is fixed.

The inverse function of h^Y with respect to v exists by strict monotonicity of h^Y in v. It is denoted by $v = g(y, x)$. Then the following identity can be written. For any x and y on the support of X and Y :

$$y = h^Y(x, g(y, x)). \tag{A}$$

Following Matzkin (2003) under strict monotonicity, the value, $h^Y(x, \tau_V)$ is identified by $Q_{Y|X}(\tau_V|x)$ using the equivariance property of quantiles under the Uniform normalization :

$$\underbrace{Q_{Y|X}(\tau_V|x)}_{\text{"Data"}} = \underbrace{h^Y(x, Q_{V|X}(\tau_V|x))}_{\text{Structural Feature}} \tag{B}$$

Independent variation in each argument of the structural function, $h^Y(\cdot, \cdot)$ can be identified if the two arguments, X and V, are independent.

[3] $h^Y(x, v)$ is identified by $Q_{Y|X}(v|x)$.

Strict monotonicity of h in u also guarantees the following relation

$$\begin{aligned}
\overbrace{Q_{W|YX}(\tau_U|y,x)}^{\text{"Data"}} &= \overbrace{h(y,x,u^*)}^{\text{Structural Feature}} \\
&= h(y,x,Q_{U|VX}(\tau_U|\tau_V,x)) \qquad \text{(C)} \\
&= h(y,x,f(\tau_U,g(y,x),x)) \\
\text{where } f(\tau_U,g(y,x),x) &\equiv Q_{U|VX}(\tau_U|\tau_V,x), \\
y &= Q_{Y|X}(\tau_V|x), \\
u^* &\equiv Q_{U|VX}(\tau_U|\tau_V,x) \\
v &= g(y,x) = Q_{V|X}(\tau_V|x)
\end{aligned}$$

That is, the value of the structural relation evaluated at $Y = y = Q_{Y|X}(\tau_V|x)$, $X = x$, and $U = u^* = Q_{U|YX}(\tau_U|y,x)$, is found by the quantile of the distribution of W given Y and X. Following Lemma 1 in Chesher (2007), we interpret the above relation as "the value of the structural function, $h(y,x,u^*)$, is identified by the functional of the distribution, $Q_{W|YX}(\tau_U|y,x)$". To identify all values of the function, $h(\cdot,\cdot,\cdot)$, it is required to show whether independent variations in each argument of $h(\cdot,\cdot,\cdot)$. The identifying relation, (C) does not show this. How the QCFA achieves independent variations is illustrated in Section 3.1.10 by assuming that there exists an IV.

In **Restriction A** differentiability of h, h^Y, and g are assumed. For identification analysis using (B) and (C), $Q_{Y|X}(\tau_V|x)$ and $Q_{W|YX}(\tau_U|y,x)$ need to be differentiable.

Restriction D (Differentiability) $Q_{W|YX}(\tau_U|y,x)$ is differentiable with respect to y and x, and $Q_{Y|X}(\tau_V|x)$ is differentiable with respect to x.

Remarks

- (B) and (C) are fundamental identifying relations that link "Data" and the structural features. The left hand side of (B) and (C) are called "Data" since $Q_{Y|X}(\tau_V|x)$ and $Q_{W|YX}(\tau_U|y,x)$ are functionals of the distribution of the observables, which can be obtained from data in principle.

- Some of the information regarding endogeneity is contained in $f(\tau_U,x,g(y,x)) \equiv Q_{U|VX}(\tau_U|\tau_V,x)$, where $y = Q_{Y|X}(\tau_V|x)$. When $Q_{U|VX}(\tau_U|\tau_V,x)$ is differentiable, if there is no endogeneity, then

$$\nabla_y Q_{U|VX}(\tau_U|\tau_V,x) = \nabla_g f \cdot \nabla_y g = 0.$$

- The function $f(\cdot,\cdot,\cdot)$ in (C) is introduced to contrast with the result when the outcome is interval censored whose case is considered in **Chapter 4**. When the outcome is interval censored, the value of the structural relation for the interval censored outcome can be found only at a specific quantiles, which depend on (y, x), for example, $\tau(y, x)$. In this case, $f(\tau(y, x), g(y, x)) \equiv Q_{U|YX}(\tau(y, x)|y, x)$. Thus, if $Q_{U|YX}(\tau(y, x)|y, x)$ were invariant with a continuous Y, locally at $Y = y$, $X = x$, we would have

$$\nabla_y Q_{U|YX}(\tau(y, x)|y, x) = \nabla_\tau f \cdot \nabla_y \tau(y, x) + \nabla_g f \cdot \nabla_y g = 0.$$

- Even though the outcome is discrete, the identifying relation (C) holds. However, if the outcome is discrete, the structural function, $h(\cdot, \cdot, \cdot)$ is *not* differentiable, thus, partial derivatives of $h(\cdot, \cdot, \cdot)$ are *not* defined. **Chapter 4** considers interval censored outcome whose latent structural function is differentiable, and the average of a discrete outcome which is assumed to be a differentiable function. Then the QCFA is applied to identify partial derivatives of these differentiable objects with respect to the variables of interest.

From the auxiliary equation (S-2), we have the identity (A), and the identifying relation for $h^Y(\cdot, \cdot)$, (B) :

$$\begin{aligned} y &= h^Y(x, g(y, x)) \quad &\text{(A)} \\ Q_{Y|X}(\tau_V|x) &= h^Y(x, Q_{V|X}(\tau_V|x)) \quad &\text{(B)} \end{aligned}$$

Suppose that the dimension of X, $K = 2$ for simplicity. Differentiating the identity, (A), with respect to y and x_k, $k \in \{1, 2\}$, we get

$$\begin{aligned} 1 &= \nabla_v h^Y \cdot \nabla_y g \\ 0 &= \nabla_{x_1} h^Y + \nabla_v h^Y \cdot \nabla_{x_1} g \\ 0 &= \nabla_{x_2} h^Y + \nabla_v h^Y \cdot \nabla_{x_2} g \end{aligned} \qquad (A')$$

and differentiating (B) with respect to x_k, $k \in \{1, 2\}$ we have

$$\begin{aligned} \overbrace{\nabla_{x_1} Q_{Y|X}(\tau_V|x)}^{\text{"}Data\text{"}} &= \overbrace{\nabla_{x_1} h^Y + \nabla_v h^Y \cdot \nabla_{x_1} Q_{V|X}(\tau_V|x)}^{Structural\ elements} \\ \nabla_{x_2} Q_{Y|X}(\tau_V|x) &= \nabla_{x_2} h^Y + \nabla_v h^Y \cdot \underbrace{\nabla_{x_2} Q_{V|X}(\tau_V|x)}_{(\#)}. \end{aligned} \qquad (B')$$

Suppose we are interested in identification of $\nabla_{x_k} h^Y$, $k \in \{1, 2\}$. Without further restrictions, identification of $\nabla_{x_k} h^Y$ is not feasible. Suppose we assume that $\nabla_{x_2} Q_{V|X}(\tau_V|x) = 0$ (#)[4]. Then $\nabla_{x_k} h^Y$ is identified by $\nabla_{x_k} Q_{Y|X}(\tau_V|x)$, $k \in \{1, 2\}$. Under this assumption

[4] This assumption ($\nabla_{x_k} Q_{Y|X}(\tau_V|x) = 0$, $k \in \{1, 2\}$) would be satisfied if X is independent of V.

Result (AB) is derived from the identity (A) and the identifying relation (B).

Result (AB) : Suppose that the τ_V- quantile of the distribution of V given X is invariant in the small neighborhood of $X = x$, that is, $\nabla_{x_k} Q_{V|X}(\tau_V|x) = 0$, $k \in \{1,2\}$. Then from (B') it can be shown that $\nabla_{x_k} h^Y$ is identified by $\nabla_{x_k} Q_{Y|X}(\tau_V|x)$,

$$\begin{aligned}\underbrace{\nabla_{x_k} h^Y(x, Q_{V|X}(\tau_V|x))}_{\theta(S)} &= \underbrace{\nabla_{x_k} Q_{Y|X}(\tau_V|x),}_{\mathcal{G}(F_{W|YX})} \\ k &\in \{1,2\}.\end{aligned}$$

From this, replacing $\nabla_{x_k} h^Y$ with $\nabla_{x_k} Q_{Y|X}(\tau_V|x)$, $k \in \{1,2\}$ in (A'), we have

$$\begin{aligned}\nabla_y g &= \frac{1}{\nabla_v h^Y} \\ \nabla_{x_1} g &= -\frac{1}{\nabla_v h^Y} \cdot \nabla_{x_1} Q_{Y|X}(\tau_V|x) \\ \nabla_{x_2} g &= -\frac{1}{\nabla_v h^Y} \cdot \nabla_{x_2} Q_{Y|X}(\tau_V|x).\end{aligned} \qquad (AB)$$

The identifying relations (B) and (C), and the relations (AB) will be used in the derivation of the results in this section and **Chpater 4**.

3.1.4 Identification of the *Stochastic* Ceteris Paribus Effects, $\nabla_y h$

Suppose we are interested in the causal effects of a continuous endogenous variable, Y, defined by the partial derivative of h, $\nabla_y h$. Specifying a structural relation reveals different routes of change caused by Y. This can be seen by differentiating (C) :

$$\underbrace{\overbrace{\nabla_y Q_{W|YX}(\tau_U|y,x)}^{\text{"}Data\text{"}}}_{\text{Observed change in } W \text{ due to } Y} = \overbrace{\nabla_y h + \underbrace{\nabla_u h \cdot \nabla_g f \cdot \nabla_y g}_{\text{Indirect effect through } U}.}^{Structural \text{ elements}} \qquad (C'-1)$$

When Y is not independent of the unobserved variable, the observed change in $Q_{W|YX}(\tau_U|y,x)$ due to the change in Y could be caused by two sources - the direct effect of Y on $h(\cdot,\cdot,\cdot)$ and the indirect effect of Y on $h(\cdot,\cdot,\cdot)$ through the effect of U on h. If one could identify the indirect effect, then $\nabla_y h$ can be identified by subtracting the indirect effect from the observed change in W. The following discussion shows how to measure the indirect effect to identify $\nabla_y h$. Note that since the indirect effect would be zero if there is no endogeneity, the indirect effect is called endogeneity bias. The endogeneity bias is discussed in **Section 3.2.2** in more detail.

To discuss ceteris paribus effects on the outcome, we defferentiate (C) with respect $x_k, k \in \{1,2\}$ as well as y:

$$\begin{aligned}
\nabla_y Q_{W|YX}(\tau_U|y,x) &= \nabla_y h \ + \nabla_u h \cdot \nabla_g f \cdot \underbrace{\nabla_y g}_{*}, \\
\nabla_{x_1} Q_{W|YX}(\tau_U|y,x) &= \nabla_{x_1} h + \nabla_u h \cdot (\nabla_g f \cdot \underbrace{\nabla_{x_1} g}_{*} + \nabla_{x_1} f), \qquad (C') \\
\underbrace{\nabla_{x_2} Q_{W|YX}(\tau_U|y,x)}_{\textit{Observed}\text{ Part from Data}} &= \underbrace{\nabla_{x_2} h + \nabla_u h \cdot (\nabla_g f \cdot \underbrace{\nabla_{x_2} g}_{*} + \nabla_{x_2} f).}_{\textit{Unobservable}\text{ Structural Features}}
\end{aligned}$$

Data are informative on the left hand side objects, while the terms on the right hand side are not identifiable without further restrictions. Not all the terms in the right hand side are of interest, thus, some of the terms regarding "structural" elements in the right hand side will be replaced using observed parts, which are embodied in relation (AB) derived in the previous section, or they will be eliminated by imposing homogeneous restrictions (local exogenous restriction), that is, by assuming that the terms are equal to zero.

Next, we replace the terms $(*)$ of the right hand side of (C') using (AB). Replacing $\begin{bmatrix} \nabla_y g \\ \nabla_{x_1} g \\ \nabla_{x_2} g \end{bmatrix}$ with $\frac{1}{\nabla_v h^Y} \begin{bmatrix} 1 \\ -\nabla_{x_1} Q_{Y|X}(\tau_V|x) \\ -\nabla_{x_2} Q_{Y|X}(\tau_V|x) \end{bmatrix}$, it follows that

$$\begin{aligned}
\nabla_y Q_{W|YX}(\tau_U|y,x) &= \nabla_y h \ + \nabla_u h \cdot \frac{\nabla_g f}{\nabla_v h_1^Y}, \qquad (C'') \\
\nabla_{x_1} Q_{W|YX}(\tau_U|y,x) &= \nabla_{x_1} h - \nabla_u h \cdot \frac{\nabla_g f}{\nabla_v h_1^Y} \cdot (\nabla_{x_1} Q_{Y|X}(\tau_V|x) + \nabla_{x_1} f), \\
\nabla_{x_2} Q_{W|YX}(\tau_U|y,x) &= \nabla_{x_2} h - \nabla_u h \cdot \frac{\nabla_g f}{\nabla_v h_1^Y} \cdot (\nabla_{x_2} Q_{Y|X}(\tau_V|x) + \nabla_{x_2} f).
\end{aligned}$$

From here on vector/matrix are introduced. Chesher (2003) discusses a general case with more than one endogeneous variable and the relations in (C') for many endogenous variables case are expressed using a linear equations system using matrices. The identification results in Chesher (2003) are expressed as conditions that the solution to the parameters of interest can be found, similar to the classical simultaneous equations models. By using this simple case I demonstrate using matrices how his results are derived.

Letting

$$\boldsymbol{\nabla}\mathbf{Q}_W \equiv \begin{bmatrix} \nabla_y Q_{W|YX} \\ \nabla_{x_1} Q_{W|YX} \\ \nabla_{x_2} Q_{W|YX} \end{bmatrix}_{3\times 1}, \boldsymbol{\nabla}\mathbf{h} \equiv \begin{bmatrix} \nabla_y h \\ \nabla_{x_1} h \\ \nabla_{x_2} h \end{bmatrix}_{3\times 1}, \nabla\mathbf{g} \equiv \begin{bmatrix} \nabla_y g \\ \nabla_{x_1} g \\ \nabla_{x_2} g \end{bmatrix}_{3\times 1}, \boldsymbol{\nabla}\mathbf{f}_x \equiv \begin{bmatrix} 0 \\ \nabla_{x_1} f \\ \nabla_{x_2} f \end{bmatrix}_{2\times 1}.$$

$\nabla\mathbf{g} \equiv \begin{bmatrix} \nabla_y g \\ \nabla_{x_1} g \\ \nabla_{x_2} g \end{bmatrix}_{3\times 1}$, and $\boldsymbol{\nabla}\mathbf{Q}_Y \equiv \begin{bmatrix} 1 \\ -\nabla_{x_1} Q_{Y|X} \\ -\nabla_{x_2} Q_{Y|X} \end{bmatrix}_{3\times 1}$ and replacing $\nabla\mathbf{g}$ by $\frac{\mathbf{1}}{\nabla_v h^Y}\boldsymbol{\nabla}\mathbf{Q}_Y$, (C'') can be rewritten using matrices as :

$$\nabla \mathbf{Q}_W = \nabla \mathbf{h} + \frac{(\nabla_u h \cdot \nabla_g f)}{\nabla_v h^Y} \cdot \nabla \mathbf{Q}_Y + \nabla \mathbf{f}_x.$$

Bold letters indicate matices. The structural feature to be recovered are $\nabla \mathbf{h}, \boldsymbol{\gamma}$, and $\nabla \mathbf{f}_x$, where $\gamma = \frac{(\nabla_u h \cdot \nabla_g f)}{\nabla_v h^Y}$. Restrictions imposed on $\nabla \mathbf{h}, \boldsymbol{\gamma}$, and $\nabla \mathbf{f}_x$ can be represented as the following

$$\mathbf{A}_h \cdot \nabla \mathbf{h} + \mathbf{A}_{\boldsymbol{\gamma}} \cdot \boldsymbol{\gamma} + \mathbf{A}_f \cdot \nabla \mathbf{f}_x = \mathbf{a}.$$

$\underset{(G\times 3)}{\mathbf{A}_h}$, $\underset{(G\times 1)}{\mathbf{A}_\gamma}$, $\underset{(G\times 2)}{\mathbf{A}_f}$, $\underset{(G\times 1)}{\mathbf{a}}$ are deterministic matrices, which contain the information on the restrictions imposed.

Let

$$\boldsymbol{\Phi} \equiv \begin{bmatrix} \mathbf{1}_3 & \nabla \mathbf{Q}_Y & \mathbf{B}_{3\times 2} \\ \mathbf{A}_h & \mathbf{A}_\gamma & \mathbf{A}_f \end{bmatrix}_{(G+3)\times 6}, \boldsymbol{\Psi} \equiv \begin{bmatrix} \nabla \mathbf{h} \\ \frac{(\nabla_u h \cdot \nabla_g f)}{\nabla_v h^Y} \\ \mathbf{f}_x \end{bmatrix}_{6\times 1},$$

$$\boldsymbol{\phi} \equiv \begin{bmatrix} \nabla \mathbf{Q}_W \\ \mathbf{a} \end{bmatrix}_{(G+3)\times 1}, \text{where } \gamma = \frac{(\nabla_u h \cdot \nabla_g f)}{\nabla_v h^Y} \text{ and } \mathbf{B}_{3\times 2} \equiv \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then we have a system of equations represented by the following

$$\boldsymbol{\Phi}\boldsymbol{\Psi} = \boldsymbol{\phi}.$$

The structural objects of interest are indicated by the vector, $\boldsymbol{\Psi}$. If the rank of $\boldsymbol{\Phi}$ is 6, we can find the solution to $\boldsymbol{\Psi}$. Then identification of $\nabla \mathbf{h}$ can be achieved.

One set of such restrictions is illustrated by the following steps.

1. First, suppose that $\nabla_{x_k} f = 0, k \in \{1, 2\}(**)$ **(Restriction 1 : Local Independence)**. Recall that we define $f(\tau_U, x, g(y, x)) \equiv Q_{U|VX}(\tau_U|\tau_V, x)$, where $y = Q_{Y|X}(\tau_V|x)$ in (C). Thus, $\nabla_{x_1} f = 0, k \in \{1, 2\}$ indicates that $\nabla_{x_k} Q_{U|VX}(\tau_U|\tau_V, x) = 0$. That is, $Q_{U|VX}(\tau_U|\tau_V, x)$ is locally invariant with the values of X. If X is exogenous variable, $\nabla_{x_k} Q_{U|VX}(\tau_U|\tau_V, x) = 0$. Imposing this restriction yields :

$$\begin{aligned} \nabla_y Q_{W|YX}(\tau_U|y, x) &= \nabla_y h + \nabla_u h \cdot \frac{\nabla_g f}{\nabla_v h_1^Y}, \\ \nabla_{x_1} Q_{W|YX}(\tau_U|y, x) &= \nabla_{x_1} h - \nabla_u h \cdot \frac{\nabla_g f}{\nabla_v h_1^Y} \cdot (\nabla_{x_1} Q_{Y|X}(\tau_V|x) + \underbrace{\nabla_{x_1} f}_{**}), \\ \nabla_{x_2} Q_{W|YX}(\tau_U|y, x) &= \underbrace{\nabla_{x_2} h}_{***} - \nabla_u h \cdot \frac{\nabla_g f}{\nabla_v h_1^Y} \cdot (\nabla_{x_2} Q_{Y|X}(\tau_V|x) + \underbrace{\nabla_{x_2} f}_{**}). \end{aligned}$$

2. It is still impossible to express $\nabla_y h$ in terms of observable parts only. Now suppose that $\nabla_{x_2} h = 0$ (($***$) **Restriction 2 : Local Order Condition**). That is, the structural function of the outcome, h, is invariant with the value of X_2 locally at $X_2 = x_2$. This implies that X_2 is locally excluded in h (called *local* order condition to indicate the similarity with the classical linear simultaneous equations analysis). Then we have from (C'')

$$\begin{aligned}
\nabla_y Q_{W|YX}(\tau_U|y,x) &= \nabla_y h + \nabla_u h \cdot \frac{\nabla_g f}{\nabla_v h^Y}, \\
\nabla_{x_1} Q_{W|YX}(\tau_U|y,x) &= \nabla_{x_1} h - \nabla_u h \cdot \frac{\nabla_g f}{\nabla_v h^Y} \cdot \nabla_{x_1} Q_{Y|X}(\tau_V|x), \\
\nabla_{x_2} Q_{W|YX}(\tau_U|y,x) &= -\underbrace{\nabla_u h \cdot \frac{\nabla_g f}{\nabla_v h^Y}}_{****} \cdot \nabla_{x_2} Q_{Y|X}(\tau_V|x).
\end{aligned} \tag{C'''}$$

3. To replace the term, $\nabla_u h \cdot \dfrac{\nabla_g f}{\nabla_v h_1^Y}$($****$), with the observable expression from the third equation of (C''') we have

$$-\underbrace{\nabla_u h \cdot \frac{\nabla_g f}{\nabla_v h^Y}}_{****} = \frac{\nabla_{x_2} Q_{W|YX}(\tau_U|y,x)}{\nabla_{x_2} Q_{Y|X}(\tau_V|x)}, \tag{ Bias}$$

if $\nabla_{x_2} Q_{Y|X}(\tau_V|x) \neq 0$ **(Restriction 3 : Local "Rank" Condition)**

4. Finally, replacing this in the first and second equations of (C'''), the identifying relations for $\nabla_y h$ *and* $\nabla_{x_1} h$ are derived:

$$\begin{aligned}
\nabla_y h &= \nabla_y Q_{W|YX}(\tau_U|y,x) + \frac{\nabla_{x_2} Q_{W|YX}(\tau_U|y,x)}{\nabla_{x_2} Q_{Y|X}(\tau_V|x)}, \\
\nabla_{x_1} h &= \nabla_{x_1} Q_{W|YX}(\tau_U|y,x) - \frac{\nabla_{x_2} Q_{W|YX}(\tau_U|y,x) \cdot \nabla_{x_1} Q_{Y|X}(\tau_V|x)}{\nabla_{x_2} Q_{Y|X}(\tau_V|x)}.
\end{aligned} \tag{TPD}$$

Let $TPD(y,x,\tau_U,\tau_V) \equiv \nabla_y h(y,x,u)$. The structural feature, $\nabla_y h(y,x,u)$, is identified by the functional of data, $\nabla_y Q_{W|YX}(\tau_U|y,x) + \dfrac{\nabla_{x_2} Q_{W|YX}(\tau_U|y,x)}{\nabla_{x_2} Q_{Y|X}(\tau_V|x)}$. This will be referred to as "Three Part Decomposition" indicated by (TPD).

Then the system of equations we need to solve in this illustration can be written as

$$\begin{aligned}
\nabla_y Q_{W|YX}(\tau_U|y,x) &= \nabla_y h \ + \nabla_u h \cdot \frac{\nabla_g f}{\nabla_v h^Y}, \\
\nabla_{x_1} Q_{W|YX}(\tau_U|y,x) &= \nabla_{x_1} h - \nabla_u h \cdot \frac{\nabla_g f}{\nabla_v h^Y} \cdot (\nabla_{x_1} Q_{Y|X}(\tau_V|x) + \nabla_{x_1} f), \\
\nabla_{x_2} Q_{W|YX}(\tau_U|y,x) &= \nabla_{x_2} h - \nabla_u h \cdot \frac{\nabla_g f}{\nabla_v h^Y} \cdot (\nabla_{x_2} Q_{Y|X}(\tau_V|x) + \nabla_{x_2} f). \\
\nabla_{x_2} h &= 0 \\
\nabla_{x_1} f &= 0 \\
\nabla_{x_2} f &= 0 \\
\nabla_{x_2} Q_{Y|X}(\tau_V|x) &\neq 0.
\end{aligned}$$

Note that there are $G = 3$ homogeneous restrictions which will determine the order of $\mathbf{\Phi}$. Setting

$$\mathbf{A}_h = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{A}_\gamma = \mathbf{0}, \mathbf{A}_f = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{a} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

$$\mathbf{\Phi} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -\nabla_{x_1} Q_{Y|X}(\tau_V|x) & 1 & 0 \\ 0 & 0 & 1 & -\nabla_{x_2} Q_{Y|X}(\tau_V|x) & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}_{6\times 6}.$$

Note that even though $\nabla_{x_2} Q_{Y|X}(\tau_V|x) = 0$, since $G = 3$, the order of $\mathbf{\Phi}$ is 6, thus, the necessary order condition is satisfied. However, if $\nabla_{x_2} Q_{Y|X}(\tau_V|x) = 0$, the third row is equal to the sixth row, which will result in $\text{rank}(\mathbf{\Phi}) = 5$. Thus, the *local rank condition*, $\nabla_{x_2} Q_{Y|X}(\tau_V|x) \neq 0$, is required for the system to have solution.

Remarks on the restrictions imposed

The restrictions imposed are local version of order and rank conditions used in the classical simultaneous equations models. Restriction 1 holds if X is independent of U. Restriction 2 is exclusion restrictions imposed locally at a point. Restriction 3 is local rank condition, which implies that X_2 needs to be a determinant of Y. Restriction 1,2 and 3 show that X_2 plays the role of **IV** locally at a point.

Remarks on the rank condition

The rank condition is regarding "identifiability" as well as suggesting constructive identification. In contrast with Matzkin (2008), this rank condition directly suggests constructive identification result as well. Once the specific restrictions, embodied in $\mathbf{A}_h, \mathbf{A}_\gamma, \mathbf{A}_f, \mathbf{a}$, that

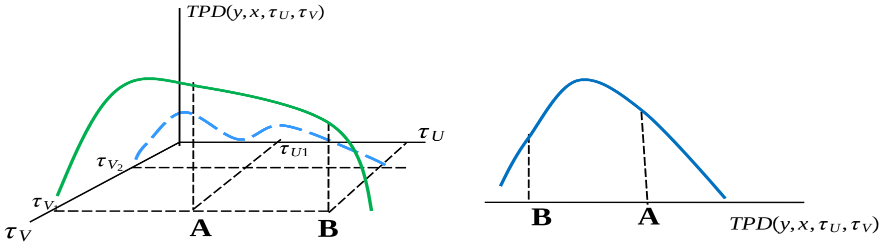


Figure 3.1: The $TPD(y, x, \tau_U, \tau_V)$ is drawn for different values of τ_V by fixing $Y = y, and$ $X = x$, in the left panel. The right panel shows the distribution of $TPD(y, x, \tau_U, \tau_V)$. By using the usual random coefficient models, a similar distribution of the coefficient to the right panel can be drawn. In contrast to them, the QCFA's random ceteris paribus effects are *interpretable* in the sense that *whose* $TPD(\cdot, \cdot, \cdot, \cdot)$ is A or B. An individual who have observed characteristics y and x and who are $\tau_{V1} - ranked$ in V and $\tau_{U1} - ranked$ in U, would have value, indicated by **A**, for example.

satisfy the rank condition is given, the exact form of the functional of the distribution can be found as was described in the illustration by solving the system.

3.1.5 Interpretation of Partial Derivatives - *Stochastic* Ceteris Paribus Effects

Random sensitivity : (TPD) encompasses random coefficient models. The identification of partial derivatives, which are measures of sensitivity, can be used to characterize the distribution of heterogeneous random responses in random coefficient models. Unlike most of the random coefficient models that do not allow *correlated* random coefficients, the random elements in TPD can be correlated with each other as well as other explanatory variables.

The randomness (V, U) is also *interpretable* in the sense that they indicate the *rankings,* (τ_U, τ_V), of the unobservable types which affect the outcome and the endogenous variable. For example, when the outcome is health spending and the endogenous variable is household income, the income elasticity can be recovered for individuals $\tau_V - ranked$ in the income (or unobserved type V) distribution and $\tau_U - ranked$ in health spending (or unobserved type U) distribution. See <Figure 3.1>.

The quantile-based identification strategy can be used to recover heterogeneous causal effects even after conditioning on the observables. This can be informative when the causal effects may be varying with different values of the unobserved characteristic - although the value of the unobserved variable would never be known, there are cases in which "*high*" versus "*low*" types of the unobserved characteristic may have different patterns of sensitivity. In such cases the QCFA could be used to investigate how individuals with different unobserved types show different responses.

3.1.6 Local Identification of Structural Features

If the local restrictions are satisfied at all points in the support, identification of the partial derivatives can be achieved at all points. Two benefits of discussing local identification need to be mentioned. The first benefit is that to identify causal effects of a continuous endogenous variable, a continuous IV is not required to infer certain information from data. Suppose one is interested in how the sensitivity to a continuous endogenous variable varies with different unobserved types, U, for the fixed type V when the observed characteristics are the same. This can be measured by varying τ_U fixing all other things (Y and X). If one wants to recover all the patterns of the sensitivity evaluated at different values of Y and X, then continuous IV would be required.

The second benefit is that local information is enough to "*refute*" (in the Breusch (1986) sense) certain hypothesis. For example, an economic model derives an implication that individuals with high unobserved type (such as effort, degree of risk aversion) would behave or respond differently from individuals with low unobserved type. This can be tested by measuring the sensitivity evaluated at different quantiles of U, other things all fixed. However, this test would not give a *confirmable* conclusion in the Breusch (1986) sense.

3.1.7 Comparison with Roehrig (1988), Benkard and Berry (2005), and Matzkin (2008)

A triangular simultaneous equations model is considered to deal with endogeneity. There have been some attempts to extend the linear classical simultaneous equations analysis into the nonparametric equations model *without triangularity.*

Brown (1983) and Roehrig (1988) assume full independence between exogenous variables and the unobserved variables for identification, while Benkard and Berry (2006) found that the necessary and sufficient condition (called derivative conditions) for full independence is actually not sufficient. Matzkin (2008) proposes different restrictions on the structure which do not require full independence and characterizes observationally equivalence structures and derive rank conditions for "identifiability". She found that for given structural relation satisfying her restrictions there exists a distribution of the unobserved variable that are observationally equivalent to the true structure and that the distribution should be independent of the exogenous variables. Those structures that are admissible and observationally equivalent need to satisfy "independence" condition, however, the "derivative condition" which was shown to be wrong by Benkard and Berry (2006) is not required in her derivation of the results.

A triangular system is a special case of the simultaneous equations systems studied in these papers. Matzkin (2008)'s restrictions are satisfied when Chesher (2003)'s restrictions are satisfied, thus, Matzkin (2008)'s identifiability condition can be applied to the triangular system (as is shown in section 5.2 in Matzkin (2008) where she shows this by imposing "exclusion" restriction which was not imposed in her model). However, the identification results under the non-triangular nonlinear simultaneous equations models in Brown (1983), Roehrig (1988), Benkard and Berry (2006), and Matzkin (2008) all rely on the differentiability of the structural functions, invertibility of the structural functions and continuity of covariates. Thus these results cannot be applied to nonparametric analysis of limited dependent variables.

3.1.8 Imbens and Newey (2009)'s Control Function Approach

When the local restrictions imposed in Chesher (2003) imposed are assumed to hold globally, both Imbens and Newey (2009) and Chesher (2003) use *the same information.* However, they use distinct *identification strategies* - Imbens and Newey (2009) add an extra regressor, $v = F_{Y|Z}(y|z)$, while Chesher (2003) condition on $y = Q_{Y|Z}(v|z)$ to control for endogeneity.

Imbens and Newey (2009) showed that the two control function approaches can produce the equivalent results on partial derivatives when the endogenous variable is continuous and U is a scalar. This is a natural result since $v = F_{Y|Z}(y|z)$ can be understood as the inverse function, $v = g(y, x) = F_{Y|Z}(y|z)$, then inverting it with respect to y, we have $y = g^{-1}(v, z) = Q_{Y|Z}(v|z)$. Thus, the two control function methods utilize exactly the same information. Note that $F_{Y|Z}(y|z)$ guarantees the one to one mapping between Y and V given Z because with continuous Y $F_{Y|Z}(y|z)$ is monotonic in y. The key property to be required for point identification using triangular system is this monotonicity and the existence of IV

Both models can produce the same identification results of partial derivatives. The advantage of using the QCFA can be shown in the next subsection where identification of partial difference is illustrated with an discrete IV. It is not clear how to identify partial difference by Imbens and Newey (2009)'s identification strategy.

Their Theorem 1 still applies to a discrete endogenous variable, as is known with the propensity score for the binary endogenous variable, however, what structural features are identified has not been discussed.

3.1.9 An Illustration of the QCFA with *Discrete* Exogenous Variables

The discussion so far assumes that the structural functions are *differentiable* with respect to every variable. In this subsection, how the identification strategy operates in recovering independent variation in each argument, by using partial differences with respect to a *continuously* varying endogenous variable. This allows for the use of a *discrete* IV. Note that if an endogenous variable is discrete, point identification of the partial difference is *not* achieved. Identification with discrete endogeneous variables will be discussed in Chapter 5.

Chesher (2007) considers identification of partial difference of a structural function evaluated at a point. Chesher (2003) and Chesher (2007) consider the case in which the endogenous variables are continuous. With continuous endogenous variables partial differences are point identified.

Partial differences could be used to measure the causal effects of a *continuously* varying variable when a certain policy change the variable discretely, for example, when household income is observed as a continuously varying, and a certain government subsidy increases income discretely, then one may be interested in the impact on the outcome, say, health spending on children, of the subsidy. One of the benefits focusing on partial difference *at a specific point* is that *discrete* instruments can be used for identification. When the endogenous variable is continuous, at least one continuous IV is required if all the values of partial derivatives are to be recovered.

Recall that X_2 plays the role of IV locally. Denote $Z \equiv X_2$. X_1 is ignored for simplicity. Assume that the exclusion restriction holds globally, so that we exclude Z from the structural relation, h. Also assume that V is independent of Z globally. Note that we assume that U and V are uniformly distributed on $(0,1)$. Suppose the value of the structural function evaluated at (y^a, u^*), where $y^a = Q_{Y|Z}(\tau_V|z^a)$ and $u^* \equiv Q_{U|VZ}(\tau_U|\tau_V, z^a)$, is the parameter of interest.[5] In other words, $h(y^a, u^*)$ is the value of the structural function for an individual with the observed characteristic y^a and the ranking of the unobserved characteristic conditional on $Y = y^a$ and $Z = z^a$ is τ_U. This can be identified by the quantile of the conditional distribution of W given Y and Z (Chesher (2003)).

$$\begin{aligned} h(y^a, u^*) &= Q_{W|YZ}(\tau_U|y^a, z^a), \\ \text{where } y^a &= Q_{Y|Z}(\tau_V|z^a) \\ u^* &\equiv Q_{U|VZ}(\tau_U|\tau_V, z^a) \\ &= Q_{U|YZ}(\tau_U|y^a, z^a), \end{aligned} \tag{3}$$

where $u^* \equiv Q_{U|VZ}(\tau_U|\tau_V, z^a) = Q_{U|YZ}(\tau_U|y^a, z^a)$ due to the one-to-one mapping between the continuous Y and V given the value of Z by the auxiliary equation (S-2).

As the value of Z changes from z^a to z^b, the τ_V-quantile of Y given Z changes from y^a to y^b. Changes in Z cause exogenous variation in Y, because V is fixed at v as Z change *due to independence of* Z *and* V.[6] That is, the change in Y from y^a to y^b caused by change in Z from z^a to z^b is achieved without changing the value of U. See <Figure3.2>

$$\begin{aligned} h(y^a, u^*) - h(y^b, u^*) &= Q_{W|YZ}(\tau_U|y^a, z^a) - Q_{W|YZ}(\tau_U|y^b, z^b), \\ y^a &= Q_{Y|Z}(\tau_V|z^a),\ y^b = Q_{Y|Z}(\tau_V|z^b), \\ &\quad \text{where } z^a \text{ and } z^b \text{ are the values for } Z. \end{aligned}$$

Thus, independent variation of $h(\cdot,\cdot)$ in y by fixing $U = u^*$ is achieved by generating

[5] u^* is not known since U is unobservable, but we assume that u^* is τ_U-quantile of distribution of U given Y and Z.

[6] Note that the value of the structural function $h(y, u^*)$ is found by fixing $u^* = Q_{U|VZ}(\tau_U|\tau_V, z)$ and by changing z. Thus, whether we can recover all the values of the function $h(y, u^*)$ over the whole support will depend on how strongly Y is related with Z as well as whether τ_V- quantile of Y given Z, $Q_{Y|Z}(\tau_V|z)$, would cover the whole points in the support of Y by varying Z.

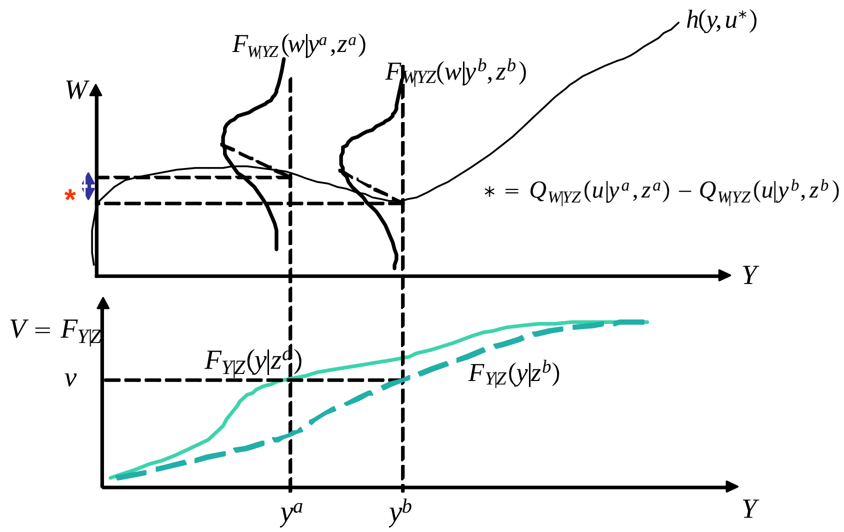


Figure 3.2: The line, $h(y, u^*)$, is drawn by fixing the value of U at u^*. Thus, the causal effect of changing Y from y^a to y^b should be measured by $*$ since on the line, $h(y, u^*)$, u^* is fixed. However, this cannot be identified by Matzkin (2003)'s idea of using quantiles of $F_{W|Y}$ since whenever the values of Y is changed, the change in $F_{W|Y}$ includes the change in W due to the change in U in the presence of endogeneity. Chesher (2003)'s suggestion is to use triangularity to control for the covariation between Y and U. The auxiliary equation (S-2) under the triangularity allows to control the source of endogeneity V when Y is continuous. Continuity of Y and monotonicity of the structural function in the unobservable guarantee that once the values of Y and Z are given the value of V is determined due to the invertibility of the function g. If there exist values z^a and z^b such that $y^a = Q_{Y|Z}(\tau_V|z^a)$ and $y^b = Q_{Y|Z}(\tau_V|z^b)$ then conditional distribution of W given Y and Z, $F_{W|YZ}$, rather than $F_{W|Y}$ will deliver information on exogenous variation in Y. Thus, $*$ is identified using the difference of the quantiles of the two conditional distributions, $F_{W|YZ}(w|y^a, z^a)$ and $F_{W|YZ}(w|y^b, z^b)$. Suppose there is no endogeneity, then Matzkin (2003)'s identification strategy of using quantiles of the conditional distribution of W given Y should be the same as Chesher (2003)'s strategy of using quantiles of the conditional distribution of W given Y and Z. This observation can be used to test exogeneity of an explanatory variable. See Section 3.2.

exogenous variation in y caused by Z.

3.2 Testability of Endogeneity

How to cope with endogeneity, defined as dependence between an explanatory variable and unobserved variables, has been one of the major issues in identification and inference in micro-econometric modelling. In this section I discuss a testable implication of endogeneity of an explanatory variable. The result is motivated by the observation discussed in the last section regarding the endogeneity bias. The endogeneity bias is found when all the variables are continuous and the structural functions are differentiable. The result reported in this section can be applied to a case in which any of relevant variables are discrete.

3.2.1 Endogeneity

Evidence regarding the presence of endogeneity is informative in determining identification and inference methods. So far the literature has more focused on how to identify and make inferences of "ceteris paribus" impacts on the outcome by allowing for endogeneity[7]. However, not only identification and inference procedures under endogeneity involve more steps but also allowing for endogeneity when the variable is actually exogenous may result in efficiency loss[8]. Thus, if one can be sure statistically of exogeneity of an explanatory variable, it could guarantee simpler estimation and more precise inference procedures.

The information regarding endogeneity is contained in the unobservable joint distribution[9] of the unobservable and (possibly) endogenous variables. To deal with this hidden information to judge the exogeneity of an explanatory variable we derive a testable expression of the conditional distribution of the unobserved heterogeneity given the explanatory variables in terms of the observables under some restrictions. This involves identification of the distribution of the unobservable variables. It is shown that if an explanatory variable is exogenous, then the distribution function of the outcome is independent of the IV conditional on the explanatory variable. When the outcome is continuous, the shape of the conditional distribution of the unobservables is "fully" identified (see Appendix C), therefore the test is "confirmable" as well as refutable in Breusch's (1986) sense. However, the test is only "refutable" if the dependent variable is discrete.

A testable implication regarding endogeneity is proposed in this section. A test statistic can be implemented based on this testable implication. See Lee (2010) for one of such

[7] For review, see for example, Hausman (1983) for linear structural equations models, and Blundell and Powell (2003, 2004) for non/semi- parametric discussion under triangularity. There have been studies using single-equation IV models such as many on OLS/2SLS in linear relations, Newey and Powell (2003) for additively separable relations, Chernozhukov and Hansen (2005), and Chesher (2010) for non-additive relations.

[8] This fact is well known in the OLS and 2SLS context. This is also true in the quantile-based control function approach (QCFA) in Chesher (2003) since the casual effects are found by estimating more terms than the case withou endogeneity.

[9] The same information on the joint distribution of the unobservable and explanatory variables is contained in the conditional distribution of the unobservable variable given the explanatory variable when the marginal distribution of the explanatory variable is known. Thus, we focus on the conditional distribution.

attempts.

3.2.2 Identification of the Endogeneity Bias with Continuous Variables

Reproducing $(C'-1)$ in Section 3.1.4

$$\underbrace{\nabla_y Q_{W|YX}(\tau_U|y,x)}_{\text{Observed change in } W \text{ due to } Y} = \nabla_y h + \underbrace{\nabla_u h \cdot \nabla_g f \cdot \nabla_y g}_{\text{Indirect effect through } U}. \qquad (C'-1)$$

The indirect effect is the endogeneity bias : when the value of Y changes, the observed changes in the outcome, W, is not the causal effect of Y on W because the observed change in the outcome contains the indirect effect through U on W. The indirect effect, which is called the endogeneity bias can be identified through the derivation (Bias) in Section 3.1.5 as follows :

$$\begin{aligned} \underbrace{\nabla_u h \cdot \nabla_g f \cdot \nabla_y g}_{\text{Endogeneity Bias}} &= -\frac{\nabla_{x_2} Q_{W|YX}(\tau_U|y,x)}{\nabla_{x_2} Q_{Y|X}(\tau_V|x)}, \\ \text{if} \quad \nabla_{x_2} Q_{Y|X}(\tau_V|x) &\neq 0. \end{aligned} \qquad \text{(Bias)}$$

Note that the indirect effect is composed of the three elements : $\nabla_u h$, the sensitivity of h to $u, \nabla_g f$, the sensitivity of $Q_{U|VX}(\tau_U|\tau_V,x)$ to v, and $\nabla_y g$, the sensitivity of Y to v. If U and Y were independent, $\nabla_u h \cdot \nabla_g f \cdot \nabla_y g$ would be zero, because $\nabla_g f$ would be zero.

$$\underbrace{-\nabla_u h \cdot \nabla_g f \cdot \nabla_y g}_{\text{Indirect effect due to endogeneity}} = \underbrace{\frac{\nabla_{x_2} Q_{W|YX}(\tau_U|y,x)}{\nabla_{x_2} Q_{Y|X}(\tau_V|x)}}_{=0 \text{ if } \boldsymbol{\nabla}_g f=0}$$

The information on $\nabla_g f$, which is not identified directly, is contained in $\dfrac{\nabla_{x_2} Q_{W|YX}(\tau_U|y,x)}{\nabla_{x_2} Q_{Y|X}(\tau_V|x)}$.[10] This is an example of structural features that are not directly testable, but "testable" since there exists an equivalent expression to this that is directly testable as we discussed in Chapter 2.

Use of weak IV : note also that although the degree of endogeneity is not high if IV is weak (small $\nabla_{x_2} Q_{Y|X}(\tau_V|x)$), then the bias measured by $\dfrac{\nabla_{x_2} Q_{W|YX}(\tau_U|y,x)}{\nabla_{x_2} Q_{Y|X}(\tau_V|x)}$ would be large.

By testing whether $\dfrac{\nabla_{x_2} Q_{W|YX}(\tau_U|y,x)}{\nabla_{x_2} Q_{Y|X}(\tau_V|x)} \neq 0$, one can test the existence of endogeneity. Since $\nabla_{x_2} Q_{Y|X}(\tau_V|x) \neq 0$ by local rank conditions, testing the existence of endogeneity would involve testing $\nabla_{x_2} Q_{W|YX}(\tau_U|y,x) = 0$. Note that if this conditional quantile invariance holds at all points and at all quantiles, the conditional distribution of W given Y and X needs to be independent of X_2.

[10] Note that "local" independence ($\nabla_g f = 0$) implies $\nabla_{x_2} Q_{W|YX}(\tau_U|y,x) = 0$ since $\nabla_{x_2} Q_{Y|X}(\tau_V|x) \neq 0$ by local rank condition.

3.2.3 Exogeneity and Conditional Independence

The information contained in the endogeneity bias discussed in the previous section is local and can only be applied to continuous variables with differentiable structural relations. This section discusses a testable implication of endogeneity that can be applied to more general cases without continuity/differentiability.

"Endogeneity" is a structural concept. Without assuming the existence of a "structure" as discussed in Chapter 2, endogeneity is not defined. An endogeneous variable is an explanatory variable which is not independent of the unobserved arguments of the structural relation of concern. A structural relation is assumed to generate the data we observe.

Restriction S (Structural Relation) : Suppose that the outcome of interest W is generated by a structural relation of the following

$$W = h(Y, X, U) \qquad (S-1)$$

The variables W and Y can be discrete, continuous, or mixed discrete continuous random variable. The variable $X = \{X_k\}_{k=1}^{K}$ is a vector of covariates. A vector of latent variates, U is jointly continuously distributed with $F_{U|YX}$.

Definition 3.1 *Exogeneity of Y : Y is called an exogenous variable if $Y \perp U$. Y is endogenous if it is not exogenous.*

Remarks on nonseparability

- A nonseparable nonparametric structural relation is used to deal with discrete or censored outcomes. For the implications or difficulties caused by nonseparability of the structural function see Hahn and Ridder (2009).
- An additively nonseparable structural function requires full independence for identification of the structural function, thus, we define exogeneity using full independence[11].
- The information regarding endogeneity is contained in the joint distribution of U and Y.

Once the identification of the conditional distribution of the unobservables given other covariates is achieved, the test of the hypothesis of exogeneity can be conducted by the

[11] The definition of endogeneity is related with the identification strategy. Whether the structural relation is assumed to be additively separable or not influences what type of restrictions are required to identify the causal effects. For example, with nonparametric structural function with additively separable error, existence of IVs that are mean independent of the regressors will be enough for identification (Newey and Powell (2003), and Newey, Powell, and Vella (1999)), whereas, when we allow for additively nonseparable errors, full independence of IV is required (Matzkin(2003), Chesher (2003), Imbens and Newey (2009), Chernozhukov and Hansen (2005), Chesher (2010) etc)

identification results which link the unobservable structural feature with the observed distribution.

We adopt the definition of conditional independence by Dawid (1979).

Definition 3.2 *Conditional independence (Dawid (1979)) : X and Z are independent conditional on Y if* $F_{X|YZ}(x|y,z) = H(x,y)$, *for all* $x, y, z,$ *for some function H.*

We assume the existence of a "*conditional* instrumental variable" in deriving the testable implication.

Restriction C-IV (Existence of "conditional" IV) : There exists a variable Z such that (i) $U \perp Z \mid Y$ and (ii) $Y = \theta(Z, \Delta)$, where Δ is a vector of determinants of Y, including both observable and unobservable variables.

One of the difficulties in testing endogeneity is the fact that endogeneity is about the dependence between the explanatory variables and the unobserved variables. The information of the dependence is contained in the conditional distribution of the unobservable variables, $F_{UY|X}$, given other variables. Once the identification of the conditional distribution of the unobservables given other covariates is achieved, the test of the hypothesis of exogeneity can be conducted by the identification results which link the unobservable structural feature with the observed distribution.

We first report a "refutable" implication when Y is exogenous. For simplicity we omit X. X can be included as conditioning variables.

Theorem 3.1 *Under Definition 3.1, Restriction S and C-IV, if Y is exogenous, then the distribution of W is independent of Z conditional on Y.*

Proof.

$$\begin{aligned}
F_{W|YZ}(w|y,z) &= \Pr[W \leq w | Y = y, Z = z] \\
&= \Pr[h(Y,U) \leq w | Y = y, Z = z] \\
(*) \quad &= \int_{\{u:h(y,u)\leq w\}} dF_{U|YZ}(u|y,z) \\
&= \int_{\{u:h(y,u)\leq w\}} dF_{U|Y}(u|y) \\
&= \left\{ \begin{array}{ll} \int_{\{u:h(y,u)\leq w\}} dF_U(u) & \text{if} \quad U \perp Y \\ \int_{\{u:h(y,u)\leq w\}} dF_{U|YZ}(u|\theta(z,\Delta),z) & \text{o.w} \end{array} \right\} \\
(**) \quad &= \left\{ \begin{array}{ll} H(w,y) & \text{if} \quad U \perp Y \\ H_\Delta(w,y,z;\Delta) & \text{o.w} \end{array} \right\},
\end{aligned}$$

where the second equality follows from Restriction S and the fourth equality is due to Restriction C-IV. Thus, we conclude that if $U \perp Y$ (Y is exogeneous), then $F_{W|YZ}(w|y,z) = H(w,y)$ (W is conditional independence of Z given Y). ■

Discussion

1. We allow for bi-directional simultaneity in the sense that Δ can include W. Although we specify the structural relations as (4) the test does not involve the estimation of the structural relation.

2. Note that the unobserved variable can be a vector. However, with multi-dimensional unobserved heterogeneity identification of the distribution of the unobserved variables is not achievable[12]. (*) shows that it is impossible to identify $\{h, F_{U|Y}\}$ separately without further restrictions, but the refutable implication can still be derived.

3. This result holds as long as Restriction C-IV holds ; only when an IV, Z satisfies the exclusion as well as relevance conditions. Weak instruments would have an impact on step (**). If the instrument is weak, there would not be much difference in $H(w,y)$ and $H_{\Delta}(w,y,z;\Delta)$ thus the link between the test of conditional independence and the test of exogeneity is weak even in the presence of endogeneity.

3.2.4 Illustration - Endogeneity, Conditional Independence, and Weak IV

I illustrate that the idea can be informally used to test exogeneity by plotting the conditional distribution functions. I also illustrate the possible loss of power due to the use of weak instruments. In each part we generate W, Y, and Z by the following data generating processes :

$$\begin{aligned} Z &\sim Poisson\ (\lambda),\ \lambda = 0.5 \\ Y &= 1(a_0 + a_1 Z + V \geq 0) \\ W &= b_0 + b_1 Y + U \\ \begin{pmatrix} U \\ V \end{pmatrix} | Z &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{UV} \\ \sigma_{UV} & \sigma_U^2 \end{pmatrix} \right) \end{aligned}$$

By varying a_1, we can control the "strength" of IV and by varying σ_{VU}, we control the degree of endogeneity. The distributions of W given Y and Z shown below are drawn using the data generated by the above processes. We draw the cumulative distribution functions, $F_{W|YZ}$, for $Y \in \{0,1\}$, and $Z \in \{0,1\}$ to examine the link between conditional independence and endogeneity, and how the link is affected by the strength of IV.

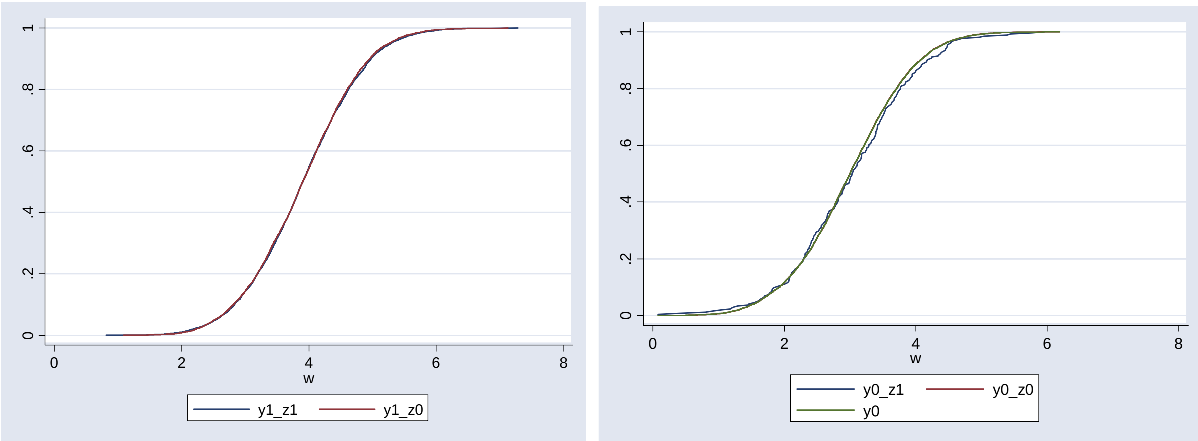
When $\sigma_{UV} = 0$, that is, when there is no endogeneity, the two conditional distributions for different values of Z are the same, while when $\sigma_{UV} \neq 0$, that is, when there is

[12] See Chesher (2009).

endogeneity, the two conditional distributions differ when the instrument is strong, but they do not show much difference when the instrument is weak.

1. Exogenous Y and conditional independence

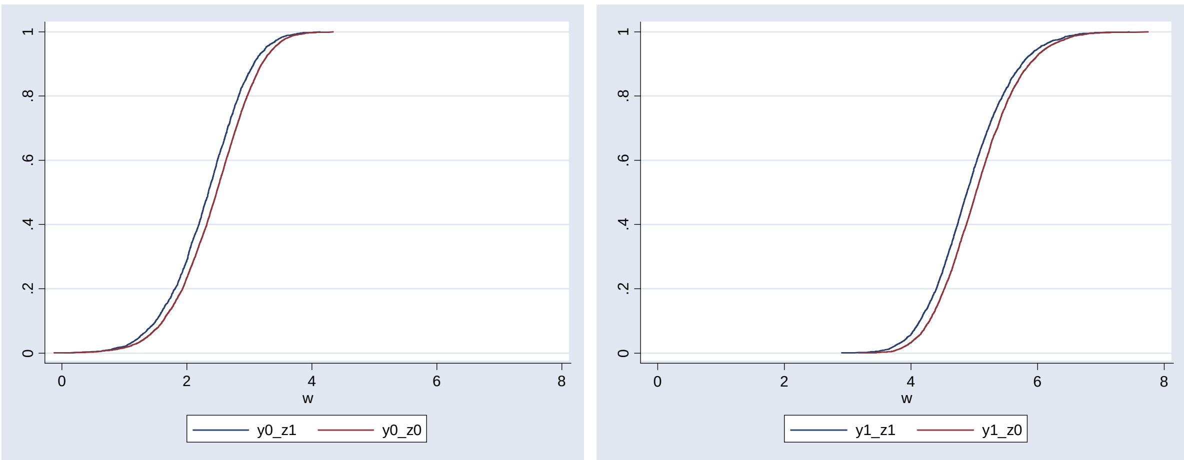
I set $\sigma_{UV} = 0$. The two graphs show the distribution functions of W given Y and Z. The first panel shows whether $F_{W|YZ}$ is independent of Z once we condition on $Y = 1$. It shows that $F_{W|Y=1,Z=1} = F_{W|Y=1,Z=0}$[13]. The second panel is the distribution functions of W given Y and Z for $Y = 0$ for different values of $Z \in \{0, 1\}$.



2. Weak IV, endogenous Y

2.1 $\sigma_{UV} = 0.7, \textit{and } a_1 = 0.3$

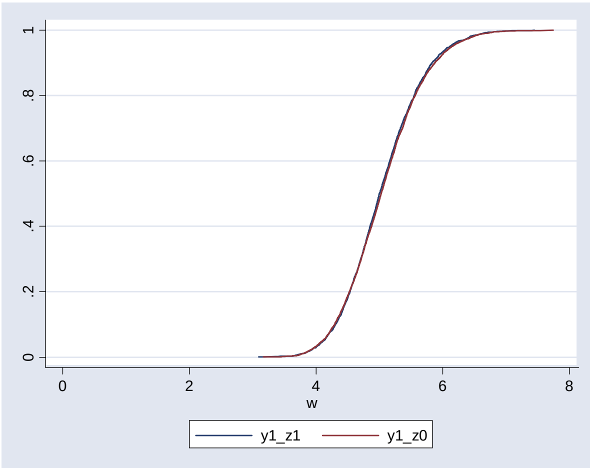
I consider endogenous Y($\sigma_{UV} = 0.7$) and "relatively " weak IV ($a_1 = 0.3$). As long as Z is "relevant" the distribution of outcome seems to be affected by the values of Z.

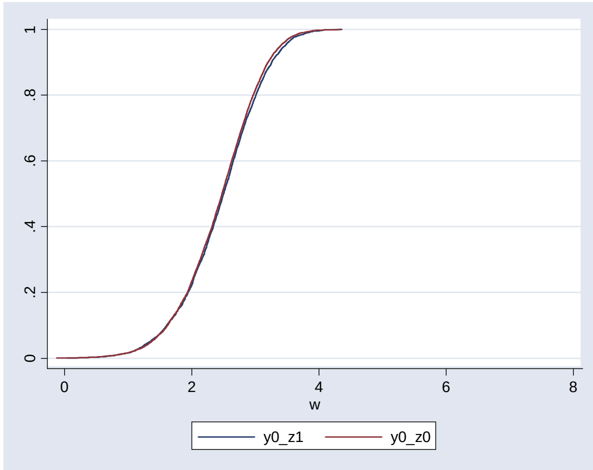


2.2 $\sigma_{UV} = 0.7$ and $a_1 = 0$

[13] Z is distributed by Poisson, but with mean $\lambda = 0.5$, there are a few observations for the values $Z = 2, 3, ...$

When Z is not "relevant", as we expected, the distribution of the outcome is not affected by the irrelevant IV conditioning on Y. Even though Y is endogenous, plotting $F_{W|Y=1Z=1}$ and $F_{W|Y=1Z=0}$ suggests that $F_{W|YZ}$ may be independent of Z. This shows the case in which testing exogeneity via testing conditional independence fails to detect the presence of endogeneity.

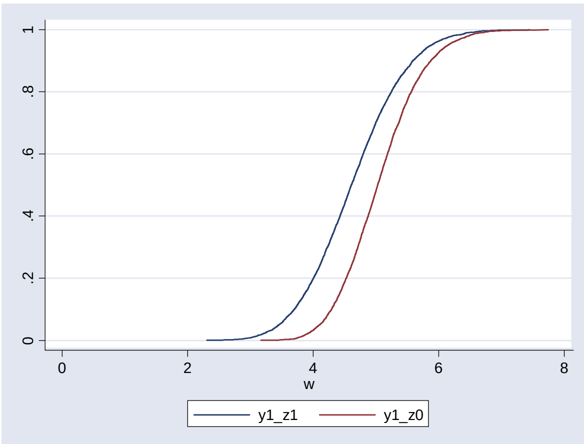


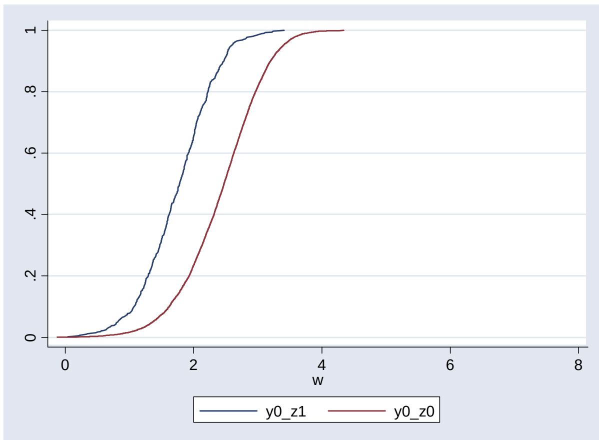


3. Strong IV, endogenous Y

$$\sigma_{UV} = 0.7, and\ a_1 = 1.3$$

Now consider a strong IV and endogenous Y. The conditional distribution is affected by both Y and Z even though Z is excluded from the outcome equation.





Chapter 4

Discrete Outcomes

This chapter is motivated by how to model interval censored outcomes, duration outcomes, and count outcomes allowing for endogeneity without relying on parametric assumptions.

The first section considers a model for interval censored outcomes. When the outcome is interval censored, the observational aspects can be described as discrete outcomes such as count data, ordered discrete data. However, the object of *structural* interest may be different each case in the sense that one may be interested in the uncensored *latent* function. The first section examines what can be recovered regarding the latent structural relation.

Interval censored duration data such as unemployment spells cannot be dealt with by the model in Section 4.1 since duration outcomes require multiple unobserved elements in representing a structural relation[1]. Chesher (2009) defines "excess heterogeneity" as the case where there are more unobserved latent variables than there are observable stochastic outcomes. When Restriction Scalar Index Unobservables (SIU) is relaxed, the QCFA cannot be applied. A random index model with *excess heterogeneity is discussed* in Section 4.2 to deal with such cases.

The last section discusses how to measure sensitivity of the averaged object with the innovation being that the averaged object can be stochastic. When allowing for excess heterogeneity, the objects of interest in many studies are those obtained by integrating out the vector unobserved elements. See Blundell and Powell (2003, 2004), Hoderlein and Mammen (2007), Imbens and Newey (2009), Chernozhukov *et al* (2009), for example. All of these studies do not allow for stochastic elements to be conditioned. The structural object of interest in the last section allows for this.

All the regressors are assumed to be continuous throughout this Chapter. The endogeneity is corrected for by the QCFA by Chesher (2003) discussed in Chapter 3.

When the outcome is discrete, partial derivatives of the structural function are not defined. To measure "ceteris paribus" impacts of a continuous variable, partial derivatives of three different *structural objects* are considered :

[1] When the outcome is duration data, there need to be at least two unobserved variables, which cannot be expressed as a single index, in order to express the duration outcome using a non-additive structural function. See the example in Chesher (2009).

- partial derivatives of *a latent structural function* are considered for the interval censored outcome in section 4.1

- partial derivatives of *random indices* are considered in a model with excess heterogeneity in Section 4.2. When the structural function is differentiable, then ratios of partial derivatives of the structural function can be recovered, once ratios of random indices are recovered.

- partial derivatives of *stochastic average conditional response (SACR) function* are considered for discrete outcomes in Section 4.3. This object is different from the average of the partial derivatives. The object of interest in Section 4.3 is partial derivatives of the mean.

To the best of my knowledge, what is identified and how it is identified with *interval censored outcome* when the latent structural function is additively nonseparable allowing for *endogeneity* has not been discussed. Also, identification of the stochastic index or the stochastic average conditional response function allowing for dependence between the unobserved variables has not been discussed in the literature. The innovation is through the triangularity and strict monotonicity under the existence of an IV, by which certain *stochastic elements* can be recovered from the auxiliary equations for the continuous endogenous variables using the triangularity. This observation is used in Section 4.2 and 4.3. Section 4.2 is a direct extension of Chesher (2009) in which an index restriction is imposed but the index does not have a random element. Section 4.2 shows that allowing for the index to be stochastic is possible when the same QCFA is used as an identification strategy.

In all cases only *ratios* of partial derivatives can be recovered due to lack of variation in the outcome in the three cases. Indeed even under the stronger parametric restrictions when the outcome is discrete, only ratios of coefficients can be identified.

4.1 Interval Censored Outcome

In reality, interval censoring is present everywhere - variables treated as being continuous such as age, expenditure etc, in fact are measured/reported as discrete. The degree of discreteness could depend on the survey design, thus the threshold points ($\{T_m\}_{m=1}^M$) are fixed and known a priori, or the discreteness could depend on the interviewee's memory or intention, hence $\{T_m\}_{m=1}^M$ could also vary with individuals. Examples of the former would be wealth data in the Health and Retirement Survey (HRS) or interval data on food consumption in the BHPS (British Household Panel Survey), and an example of the latter would be unemployment duration data (Han and Hausman (1990), Ridder(1990)), where we would imagine that as the unemployment duration increases the intervals of observed duration will increase. In both cases the econometric studies are conducted by assuming the continuity of the variables - especially the basic building blocks of duration analysis such as the hazard function are defined under differentiability. Thus considering a discrete outcome as the result of interval censoring of a differentiable latent function may be relevant in many applications.

Manski and Tamer (2002) consider identification and inference in cases with both interval censored outcomes and interval censored regressors under both nonparametric and semiparametric setup without consideration of endogeneity. Khan and Tamer (2007) consider censoring with a linear structural function allowing for endogeneity. Bontemps, Magnac, and Maurin (2008) also consider a model that admits interval censoring allowing for endogeneity under the linear structural function. If the latent outcome function is additively separable, the results in threshold crossing models - as in Matzkin (1992) - can be applied to the interval censored outcome model.

Berry and Tamer (2007) use Matzkin (1992)'s identification strategy to identify *ratios of partial derivatives* of additively separable nonparametric latent function under the threshold crossing framework not allowing for endogeneity. However, the identification and inference when the outcome is interval censored whose uncensored process is *nonseparable* has not been studied before. This section proposes an identification result under endogeneity by using a triangular structure.

4.1.1 The Model

To model interval censoring we assume a latent function for the outcome, and assume that there exists a censoring mechanism that transforms the latent function into the observed interval censored outcome.

Restriction IC (Interval censored outcome with scalar unobservable index) *$W^*, Y \equiv \{Y_i\}_{i=1}^N, X \equiv \{X_i\}_{i=1}^K, U$, and $V \equiv \{V_i\}_{i=1}^N$ are random variables, which are continuously distributed. For any values of X, U, and V, unique values of W^* and Y are determined by the structural equations*

$$W^* = h^*(Y, X, U), \quad \text{(S-1*)}$$
$$Y_n = h_n^Y(X, V_n), \quad \text{(S-2)}$$
$$n \in \{1, 2, ..., N\}$$

The unobservable variables, U and $\{V_n\}_{n=1}^N$ are scalar indices and are jointly continuously distributed and each is normalized to be uniform (0, 1). The structural relation h^* is strictly monotonic in each of the unobservable random variables U. Each function $\{h_n^Y\}_{n=1}^N$ is strictly monotonic *with respect to variation in the unobservable $\{V_n\}_{n=1}^N$. This model admits such cases that $U = \theta_U(U_1, ..., U_L)$, $V_n = \theta_{V_n}(V_1, ..., V_{I_n})$, where $\theta_U : R^L \to (0, 1)$, $\theta_{V_n} : R^{I_n} \to (0, 1)$, for some positive number L and I_n, for $n = 1, 2, ..., N$.*

However, the outcome of interest W^* is not observed completely, but it is interval censored by the following censoring mechanism :

$$W = h(Y,X,U) \tag{O}$$
$$= \sum_{m=1}^{M} w_m \cdot \mathbf{1}(T_{m-1}<h^*(Y,X,U) \leq T_m), w_m \in (T_{m-1}, T_m] \ .$$

$\{T_m\}_{m=0}^{M}$ are threshold points by which the observed values of W are determined.

Define $P^m(y,x) \equiv F_{W|YX}(w_m|y,x)$. Then

$$\begin{aligned}
P^m(y,x) &\equiv F_{W|YX}(w_m|y,x) \\
(1) \quad &= F_{W^*|YX}(T_m|y,x) \\
&= \Pr(W^* \leq T_m|Y=y, X=x) \\
(2) \quad &= \Pr(h^*(Y,X,U) \leq T_m|Y=y, X=x) \\
(3) \quad &= \Pr(h^*(Y,X,U) \leq T_m|V=v, X=x) \\
(4) \quad &= \Pr(h^*(Y,X,U) \leq T_m|V=g(y,x), X=x).
\end{aligned}$$

(1) is by the censoring mechanism, and (2) is by the functional relationship specified in Restriction IC. (3) is by strict monotonicity of h_n with respect to V_n, and (4) is by the inverse relationship between V and Y, where $g(y,x) = [g_1(y_1,x),...,g_N(y_N,x)]'$.

The continuous distribution of $F_{W^*|YX}(w^*|y,x)$ is unknown, we only observe the discrete distribution of W given Y and X, $P^m(y,x) \equiv F_{W|YX}(w_m|y,x)$. However, we can obtain partial information on latent $F_{W^*|YX}(w^*|y,x)$ from observable $P^m(y,x)$. Although T_m is unknown, by the interval censoring mechanism defined in Restriction IC, we know

$$P^m(y,x) \equiv F_{W|YX}(w_m|y,x) = F_{W^*|YX}(T_m|y,x).$$

This is the information we can use to identify structural features regarding the latent function h^*. This relation implies that T_m is the $P^m(y,x)-$ quantile of W^* given Y and X. Then by the strict monotonicity of h^* in U, we can connect the latent structural function $h^*(y,x,u)$ with observable distribution $P^m(y,x)$, by the following argument.

Under Restriction IC, we have the following relation :

$$\begin{aligned}
T_m &= h^*(y,x,Q_{U|VX}(P^m(y,x)|\tau_V,x)), \qquad (C-IC) \\
\text{where } y &= Q_{Y|X}(\tau_V|x).
\end{aligned}$$

Remarks

1. Note that when $u \neq Q_{U|VX}(P^m(y,x)|\tau_V,x)$, where $y = Q_{Y|X}(\tau_V|x)$, $m = 1,2,...,M$,

the value of $h^*(y,x,u)$ is not known. With strict monotonicity of h^* in u, we can bound the values of h^* at each point.

2. Eq. $(C-IC)$ is the key relation linking the latent structural function with the observed distribution. $(C-IC)$ is a level set of values of (y,x,u) that produce the same level, T_m. The information regarding the partial derivatives is derived from the level set. Thus, it is no wonder that we can only identify ratios of the partial derivatives.

4.1.2 Identification of Partial Derivatives

The inverse function of each h_n with respect to V_n exists by strict monotonicity. It is denoted by g_n.

Restriction D-IC (Differentiability-IC) *The conditional distribution of W given Y and X, $P^m(y,x)$ is differentiable with respect to y and x, and the conditional distribution function of Y_n given X, and it's quantiles,* $Q_{Y_n|X}(\tau_{Vn}|x)$ are differentiable with respect to x.

Define

$$
\begin{aligned}
f(P^m(y,x), g_1(y_1,x), g_2(y_2,x), ..., g_N(y_N,x), x) &\equiv Q_{U|VX}(P^m(y,x)|g(y,x),x), \\
\text{where } g(y,x) &= [g_1(y_1,x), g_2(y_2,x), ..., g_N(y_N,x)]'.
\end{aligned}
$$

Note that if $Q_{U|VX}$ is independent of x, we have $\nabla_x f = 0$.

Regarding the structural elements the following vectors are defined :

$$
\boldsymbol{\lambda}_{y/u} \equiv \frac{1}{\nabla_u h^*}\begin{bmatrix}\nabla_{y_1} h^* \\ \vdots \\ \nabla_{y_N} h^*\end{bmatrix}_{N\times 1}, \boldsymbol{\lambda}_{x/u} \equiv \frac{1}{\nabla_u h^*}\begin{bmatrix}\nabla_{x_1} h^* \\ \vdots \\ \nabla_{x_K} h^*\end{bmatrix}_{K\times 1},
$$

$$
\mathbf{f}_x \equiv \begin{bmatrix}\nabla_{x_1} f \\ \vdots \\ \nabla_{x_K} f\end{bmatrix}_{K\times 1}, \boldsymbol{\gamma} \equiv \begin{bmatrix}\frac{\nabla_{g_1} f}{\nabla_{v_1} h_1} \\ \vdots \\ \frac{\nabla_{g_N} f}{\nabla_{v_N} h_N}\end{bmatrix}_{N\times 1}.
$$

For the observable elements the following are defined :

$$
\mathbf{F}_y^W \equiv \begin{bmatrix}\nabla_{y_1} P^m \\ \vdots \\ \nabla_{y_m} P^m\end{bmatrix}_{N\times 1}, \mathbf{F}_x^W \equiv \begin{bmatrix}\nabla_{x_1} P^m \\ \vdots \\ \nabla_{x_K} P^m\end{bmatrix}_{K\times 1}, \text{ and } \mathbf{G}_x \equiv \begin{bmatrix}\nabla_{x_1} Q_{Y_1|X} & \cdots & \nabla_{x_1} Q_{Y_n|X} \\ \vdots & \ddots & \vdots \\ \nabla_{x_K} Q_{Y_1|X} & \cdots & \nabla_{x_K} Q_{Y_n|X}\end{bmatrix}_{K\times N}.
$$

Restriction R-IC. *There are G restrictions on $\boldsymbol{\lambda}_{y/u}, \boldsymbol{\lambda}_{x/u}$, $\boldsymbol{\gamma}$, and $\mathbf{f}_x$ as follows,*

$$
\mathbf{A}_y \cdot \boldsymbol{\lambda}_{y/u} + \mathbf{A}_x \cdot \boldsymbol{\lambda}_{x/u} + \mathbf{A}_\gamma \cdot \boldsymbol{\gamma} + \mathbf{A}_f \cdot \mathbf{f}_x = \mathbf{a}.
$$

The arrays $\underset{G\times N}{\mathbf{A}_y}$, $\underset{G\times K}{\mathbf{A}_x}$, $\underset{G\times N}{\mathbf{A}_\gamma}$, $\underset{G\times K}{\mathbf{A}_f}$ *and* $\underset{G\times 1}{\mathbf{a}}$ *are nonstochastic.*

$\mathbf{A}_y, \mathbf{A}_x, \mathbf{A}_\gamma, \mathbf{A}_f$ *and* $\mathbf{a}$ specify specific restrictions imposed in each case. This is local analogy of the classical linear simultaneous equations system. Restriction R-IC can be specified in more detail as will be illustrated later. In the illustration I use the exclusion restriction, the relevance condition (rank condition), and independence restriction as the example of Restriction R. For example, if one is interested in measuring returns to schooling, it is likely that some individuals' unobserved characteristics such as ability, motivation, sociability, etc. determine both the individual's wage and schooling decision. In this case if there exists a variable, such as distance to college, subsidy to schooling, quarter of birth that affects the schooling decision, but does not determine wage, as well as is independent of the unobserved characteristics, then **Restriction R-IC** can be constructed by specifying the matrices $\mathbf{A}_y, \mathbf{A}_x, \mathbf{A}_\gamma, \mathbf{A}_f$ *and* $\mathbf{a}$ accordingly.

Then we define $\mathbf{\Phi}$, $\mathbf{\Psi}$, and $\boldsymbol{\phi}$ to express the system of equations which is to be solved for $\mathbf{\Psi}$, to achieve identification

$$\mathbf{\Phi} \equiv \begin{bmatrix} \mathbf{I}_N & \mathbf{0}_K & \mathbf{I}_N & \mathbf{0}_K \\ \mathbf{0}_N & \mathbf{I}_K & -\mathbf{G}_x & \mathbf{I}_K \\ \mathbf{A}_y & \mathbf{A}_x & \mathbf{A}_\gamma & \mathbf{A}_f \end{bmatrix}_{(G+N+K)\times(2N+2K)}, \mathbf{\Psi} \equiv \begin{bmatrix} \boldsymbol{\lambda}_{y/u} \\ \boldsymbol{\lambda}_{x/u} \\ \gamma \\ \mathbf{f}_x \end{bmatrix}_{(2N+2K)\times 1},$$

$$\boldsymbol{\phi} \equiv \begin{bmatrix} \mathbf{S}_y \\ \mathbf{S}_x \\ \mathbf{a} \end{bmatrix}_{(G+N+K)\times 1} \text{,where } \mathbf{S}_y = \nabla_\tau f \cdot \mathbf{F}_y^W \text{ and } \mathbf{S}_x = \nabla_\tau f \cdot \mathbf{F}_x^W$$

$\mathbf{\Psi}$ contains the structural features of interest. If the solutions to $\mathbf{\Psi}$ can be found then identification is achieved. thus, the rank condition for identification can be stated.

Theorem 4.1 *Under Restriction IC,D-IC, and R-IC,* $\Phi\Psi = \phi$, *and* Ψ *can be found iff* $rank(\Phi) = 2N + 2K$ *for which a necessary condition is* $G \geq N + K$.

Proof. See Appendix A. ■

Note that $\mathbf{S}_y = \nabla_\tau f \cdot \mathbf{F}_y^W$ and $\mathbf{S}_x = \nabla_\tau f \cdot \mathbf{F}_x^W$ contain the unidentified element $\nabla_\tau f$. Therefore, to eliminate this, we take ratios. This is why only ratios of partial derivatives are identified.

Corollary 4.1 *Under Restriction IC,D-IC, and R-IC, ratios of partial derivatives are identified iff* $rank(\Phi) = 2N + 2K$.

Proof. Theorem 4.1 shows identification of $\boldsymbol{\lambda}_{y/u}$ and $\boldsymbol{\lambda}_{x/u}$, but they contain $\nabla_u h^*$. To eliminate $\nabla_u h^*$, we take ratio of the two. Thus, the same rank condition is applied. ■

4.1.3 Illustration - Constructive Identification

Restriction R-IC specifies a general form of restrictions. There can be many options that satisfy Restriction R-IC. In this section we consider the traditional rank and order condition applied to a point - "local" rank and "order" condition.

I consider a case where there is only one endogenous variable, $N = 1$ and $K = 2$. Reproducing $(C - IC)$ we have

$$\begin{aligned} T_m &= h^*(y, x, Q_{U|VX}(P^m(y,x)|\tau_V, x)) \quad (C-IC) \\ &= h^*(y, x, f(P^m(y,x), g(y,x), x)) \\ \text{where } f(P^m(y,x), g(y,x), x) &\equiv Q_{U|VX}(P^m(y,x)|\tau_V, x) \end{aligned}$$

Differentiating $(C - IC)$, with respect to y, and x_k, $k \in \{1, 2\}$, we have

$$\begin{aligned} 0 &= \nabla_y h^* + \nabla_u h^* \cdot (\nabla_\tau f \cdot \underbrace{\nabla_y P^m}_{\text{Observable}} + \nabla_g f \cdot \underbrace{\nabla_y g}_{*}), \qquad (C-IC') \\ 0 &= \nabla_{x_1} h^* + \nabla_u h^* \cdot [(\nabla_\tau f \cdot \underbrace{\nabla_{x_1} P^m}_{\text{Observable}} + \nabla_g f \cdot \underbrace{\nabla_{x_1} g}_{*}) + \nabla_{x_1} f], \\ 0 &= \nabla_{x_2} h^* + \nabla_u h^* \cdot [(\nabla_\tau f \cdot \underbrace{\nabla_{x_2} P^m}_{\text{Observable}} + \nabla_g f \cdot \underbrace{\nabla_{x_2} g}_{*}) + \nabla_{x_2} f]. \end{aligned}$$

The ceteris paribus effect of Y is indicated by $\nabla_y h^*$. Note that in contrast with the continuous outcome case discussed in Chapter 3, the left hand side is all 0, and the information obtainable from data are $\nabla_y P^m$, $\nabla_{x_1} P^m$, and $\nabla_{x_2} P^m$. Without imposing more restrictions, $\nabla_y h^*$ is not identified. Replacing the terms $(*)$ using (AB) from Chapter 3, we have

$$\begin{aligned} 0 &= \nabla_y h^* + \nabla_u h^* \cdot [\nabla_\tau f \cdot \nabla_y P^m + \frac{\nabla_g f}{\nabla_v h^Y}], \qquad (C-IC'') \\ 0 &= \nabla_{x_1} h^* + \nabla_u h^* \cdot [\nabla_\tau f \cdot \nabla_{x_1} P^m - \frac{\nabla_g f}{\nabla_v h^Y} \cdot \nabla_{x_1} Q_{Y|X} + \underbrace{\nabla_{x_1} f}_{**}], \\ 0 &= \underbrace{\nabla_{x_2} h^*}_{***} + \nabla_u h^* \cdot [\nabla_\tau f \cdot \nabla_{x_2} P^m - \frac{\nabla_g f}{\nabla_v h^Y} \cdot \nabla_{x_2} Q_{Y|X} + \underbrace{\nabla_{x_2} f}_{**}]. \end{aligned}$$

The same restrictions as Chapter 3 are adopted to demonstrate the effect of interval censoring on identification. Firstly, we impose $\nabla_{x_k} f = 0, k \in \{1, 2\}$ $(**)$ (local independence restriction), which would hold if X were exogenous. Then local exclusion restriction

($\nabla_{x_2} h^* = 0\ (***)$) is imposed yielding

$$
\begin{aligned}
0 &= \nabla_y h^* + \nabla_u h^* \cdot [\nabla_\tau f \cdot \nabla_y P^m + \frac{\nabla_g f}{\nabla_v h^Y}], \qquad (C - IC'') \\
0 &= \nabla_{x_1} h^* + \nabla_u h^* \cdot [\nabla_\tau f \cdot \nabla_{x_1} P^m - \frac{\nabla_g f}{\nabla_v h^Y} \cdot \nabla_{x_1} Q_{Y|X}], \\
0 &= \nabla_u h^* \cdot [\nabla_\tau f \cdot \nabla_{x_2} P^m - \frac{\nabla_g f}{\nabla_v h^Y} \cdot \nabla_{x_2} Q_{Y|X}].
\end{aligned}
$$

If $\nabla_u h^* \neq 0$, then $\dfrac{\nabla_g f}{\nabla_v h^Y} = \dfrac{\nabla_\tau f \nabla_{x_2} P^m}{\nabla_{x_2} Q_{Y|X}}$, where $\nabla_{x_2} Q_{Y|X} \neq 0$ (local rank condition) from the third equation of $(C - IC'')$. Replacing $\dfrac{\nabla_g f}{\nabla_v h^Y}$ using this in the first and the second equations in $(C - IC'')$, we have

$$
\begin{aligned}
-\frac{\nabla_y h^*}{\nabla_u h^*} &= \nabla_\tau f \cdot (\nabla_y P^m + \frac{\nabla_{x_2} P^m}{\nabla_{x_2} Q_{Y|X}}), \qquad (C - IC''') \\
-\frac{\nabla_{x_1} h^*}{\nabla_u h^*} &= \nabla_\tau f \cdot (\nabla_{x_1} P^m - \nabla_{x_2} P^m \cdot \frac{\nabla_{x_1} Q_{Y|X}}{\nabla_{x_2} Q_{Y|X}}).
\end{aligned}
$$

Note that in contrast with the continuous outcome case considered in Chapter 3, interval censoring causes loss of identifying power when exactly the same restrictions are imposed since $\dfrac{\nabla_y h^*}{\nabla_u h^*}$ is identified, rather than $\nabla_y h^*$. Therefore, the *ratio* of partial derivatives, $\dfrac{\nabla_y h^*}{\nabla_{x_1} h^*}$ is identified by $\dfrac{\nabla_y P^m + \dfrac{\nabla_{x_2} P^m}{\nabla_{x_2} Q_{Y|X}}}{\nabla_{x_1} P^m - \nabla_{x_2} P^m \dfrac{\nabla_{x_1} Q_{Y|X}}{\nabla_{x_2} Q_{Y|X}}}$ as is inferred from $(C - IC''')$.

This can be shown as the following :

Let $\mathbf{F}_y^W \equiv \begin{bmatrix} \nabla_y P^m \end{bmatrix}_{1\times 1}$, $\mathbf{F}_x^W \equiv \begin{bmatrix} \nabla_{x_1} P^m \\ \nabla_{x_2} P^m \end{bmatrix}_{2\times 1}$, $\mathbf{G}_x \equiv \begin{bmatrix} \nabla_{x_1} Q_{Y|X} \\ \nabla_{x_2} Q_{Y|X} \end{bmatrix}_{2\times 1}$,
$\boldsymbol{\lambda}_{y/u} \equiv \dfrac{1}{\nabla_u h^*} \begin{bmatrix} \nabla_y h^* \end{bmatrix}_{1\times 1}$, $\boldsymbol{\lambda}_{x/u} \equiv \dfrac{1}{\nabla_u h^*} \begin{bmatrix} \nabla_{x_1} h^* \\ \nabla_{x_2} h^* \end{bmatrix}_{2\times 1}$,
$\mathbf{f}_x \equiv \begin{bmatrix} \nabla_{x_1} f \\ \nabla_{x_2} f \end{bmatrix}_{2\times 1}$, $\mathbf{f}_g \equiv \begin{bmatrix} \nabla_g f \end{bmatrix}_{1\times 1}$,and $\boldsymbol{\gamma} \equiv \begin{bmatrix} \frac{\nabla_g f}{\nabla_v h^Y} \end{bmatrix}_{1\times 1}$.

Then $(C - IC''')$ can be expressed using these vectors as

$$
\begin{aligned}
\boldsymbol{\lambda}_{y/u} + I_N \cdot \boldsymbol{\gamma} &= -\nabla_\tau f \cdot \mathbf{F}_y^W \\
\boldsymbol{\lambda}_{x/u} - \mathbf{G}_x \cdot \boldsymbol{\gamma} + \mathbf{f}_x &= -\nabla_\tau f \cdot \mathbf{F}_x^W.
\end{aligned}
$$

Then the system of equations that are need to be solved is :

$$
\begin{aligned}
\frac{\nabla_y h^*}{\nabla_u h^*} + \qquad \frac{\nabla_g f}{\nabla_v h^Y} \qquad &= \quad \nabla_\tau f \cdot \nabla_y P^m \\
\frac{\nabla_{x_1} h^*}{\nabla_u h^*} - \nabla_{x_1} Q_{Y|X} \frac{\nabla_g f}{\nabla_v h^Y} + \nabla_{x_1} f &= -\nabla_\tau f \cdot \nabla_{x_1} P^m \\
\frac{\nabla_{x_2} h^*}{\nabla_u h^*} - \nabla_{x_2} Q_{Y|X} \frac{\nabla_g f}{\nabla_v h^Y} + \nabla_{x_2} f &= -\nabla_\tau f \cdot \nabla_{x_2} P^m \\
\frac{\nabla_{x_2} h^*}{\nabla_u h^*} &= 0 \\
\nabla_{x_1} f &= 0 \\
\nabla_{x_2} f &= 0 \\
\nabla_{x_2} Q_{Y|X} &\neq 0.
\end{aligned}
$$

This can be written using the matrices $\mathbf{\Phi}$, $\mathbf{\Psi}$, and $\boldsymbol{\phi}$, as

$$
\text{where } \mathbf{\Phi} \equiv \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & \nabla_{x_1} Q_{Y|X} & -1 & 0 \\ 0 & 0 & 1 & \nabla_{x_2} Q_{Y|X} & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{6\times 6}, \mathbf{\Psi} \equiv \begin{bmatrix} \nabla_{y_1} h^*/\nabla_u h^* \\ \nabla_{x_1} h^*/\nabla_u h^* \\ \nabla_{x_2} h^*/\nabla_u h^* \\ \nabla_{g_1} f/\nabla_{v_1} h^Y \\ \nabla_{x_1} f \\ \nabla_{x_2} f \end{bmatrix}_{6\times 1}, \boldsymbol{\phi} \equiv \begin{bmatrix} \mathbf{S}_y \\ \mathbf{S}_x \\ \mathbf{a} \end{bmatrix}
$$

with the homogeneous restrictions indicated by the following matrices as :

$$
\mathbf{A}_y = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \mathbf{A}_x = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{A}_\gamma = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \mathbf{A}_f = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \text{ and } \mathbf{a} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.
$$

The restrictions used are $\nabla_{x_k} f = 0,\ k \in \{1, 2\}, \nabla_{x_2} h^* = 0$, with $N = 1, K = 2$, and $G = 3$. If $\nabla_{x_2} Q_{Y|X} = 0$, then the third row (r_3) of $\mathbf{\Phi}$ is a linear combination of the fourth (r_4) and the sixth (r_6) rows as $r_3 = r_4 - r_6$. Thus, the local rank condition, $\nabla_{x_2} Q_{Y|X} \neq 0$ is required for $rank(\mathbf{\Phi}) = 6$. The identification condition specified in Corollary 4.1. is satisfied with the restrictions imposed in the illustration.

4.2 Random Index Model with Excess Heterogeneity

The interval censored model in Section 4.1 cannot be applied to interval censored duration data because Restriction SIU does not hold. The model considered here permits a nonseparable and *random* index. As in Chesher (2009) I impose an index structure, but unlike Chesher (2009) I allow for the index to vary with the unobserved factors that affect both the outcome and the endogenous variables. By allowing for the index to be nonseparable and to include the unobserved heterogeneity that affects endogenous variables, we can recover heterogeneous *random* marginal effects - more precisely, ratios of marginal effects. This can relax some of the usual restrictions in duration and count models. The modeling building blocks such as the hazard function or the mean of the count have usually been restricted into specific forms of mixture - e.g. mixed proportional hazard. The random index in this section relaxes the patterns that the unobserved heterogeneity is incorporated into the index, the most general and flexible additively non-separable form.

There have been several studies on index models. The most relevant studies are Han (1987) and Matzkin (1991,1994). Han (1987) considered a general regression model of the form

$$\begin{gathered} Y_i = D(F(X_i\beta, U_i)) \\ \text{where } X_i \perp U_i \\ \text{and } D(\cdot) \text{ is monotonic and} \\ F(\cdot,\cdot) \text{ is strictly monotonic in both arguments.} \end{gathered}$$

This model is nonseparable, so it does allow for heterogeneous marginal effects, but, Han (1987) does not allow for possible endogeneity. The marginal impact of x is $D'F_1(x'\beta, u)\beta$, which varies with U. However, Han (1987) focuses on identification of $\beta., D(\cdot)$ and $F(\cdot,\cdot)$, are not identified separately.

Matzkin (1991, 1994) extended Han (1987)'s general regression model into the nonparametric index of the following form :

$$\begin{gathered} Y_i = D(F(X_i), U_i) \\ \text{where } X_i \perp U_i \\ \text{and } D(\cdot,\cdot) \text{ is monotonic and nonconstant in both arguments} \\ \text{and } F(\cdot) \text{ is strictly monotonic in at least one explanatory variable.} \end{gathered}$$

In both cases the models do not admit endogeneity. Also, the models admit only a single source of stochastic variation. Unlike these studies we incorporate endogeneity into the model by introducing an auxiliary equation for the endogenous variable and by allowing for the possible correlation between the unobservables in both equations. By specifying the data generating process for the endogenous variable using the triangular system we are able to control for endogeneity using the control function method. The endogeneity in this model is accounted for by the QCFA.

We base our model on Chesher (2009), but just incorporate the unobserved heterogeneity (the unobserved type that determines the endogenous variable) into the index function so that we allow for randomness of the partial derivatives of the index.

4.2.1 The Model

Restriction RI - EH (Random Index with Excess Heterogeneity)

W, $Y \equiv \{Y_i\}_{i=1}^N, X \equiv \{X_i\}_{i=1}^K, U$, and $V \equiv \{V_i\}_{i=1}^N$ are random variables, which are continuously distributed. For any values of X, U, and V, unique values of W and Y are determined by the structural equations

$$\begin{aligned} W &= h(\theta(Y, X, V), U_1, ..., U_L) \\ Y_n &= h_n^Y(X, V_n), \\ n &\in \{1, 2, ..., N\} \end{aligned}$$

where θ is a scalar valued function. The unobservable latent variables, $U = [U_1, ..., U_L]$ is a vector and each *V_n is distributed* uniform $(0, 1)$. *h is weakly monotonic in U. $\{h_n^Y\}_{n=1}^N$ is strictly* monotonic *with respect to variation in the unobservable V_n. $\{V_n\}_{n=1}^N$ should be scalar. This model admits such case that $V_n = \theta_{V_n}(V_1, ..., V_{I_n})$, where $\theta_{V_n} : R^{I_n} \to (0, 1)$, for some positive number I_n, for $n = 1, 2, ..., N$.*

The inverse function of each h_n^Y with respect to V_n exists by strict monotonicity. It is denoted by g_n. Let $g = [g_1(Y_1, X), ..., g_N(Y_N, X)]'$. Under Restriction RI - EH the conditional distribution function of W given Y and X is

$$\begin{aligned} F_{W|YX}(w|y, x) &= F_{W|VX}(w|y, v) \\ &= \int_{\{u:h(\theta(y,x,v),U)\leq w\}} dF_{U|VX}(u|g(y, x), x) \qquad \text{(C-RI)} \\ &\equiv s(w, \theta(y, x, g(y, x)), g(y, x), x) \end{aligned}$$

Restriction D - RI (Differentiability) T*he conditional distribution of W given Y and X, $F_{W|YX}(w|y, x)$ is differentiable with respect to y and x, and for $n \in \{1, 2, ..., N\}$ the conditional distribution function of Y_n given X, $F_{Y_n|X}(y_n|x)$ is differentiable with respect to y_n and x.*

Matrices of partial derivatives, all evaluated at *a point* are now defined. For structural elements the followings are defined :

$$\boldsymbol{\lambda}_y \equiv \nabla_\theta s \begin{bmatrix} \nabla_{y_1}\theta \\ \vdots \\ \nabla_{y_N}\theta \end{bmatrix}_{N\times 1}, \boldsymbol{\lambda}_x \equiv \nabla_\theta s \begin{bmatrix} \nabla_{x_1}\theta \\ \vdots \\ \nabla_{x_K}\theta \end{bmatrix}_{K\times 1},$$

$$\mathbf{s}_x \equiv \begin{bmatrix} \nabla_{x_1} s \\ \vdots \\ \nabla_{x_K} s \end{bmatrix}_{K\times 1}, \mathbf{s}_g \equiv \begin{bmatrix} \frac{\nabla_{g_1} s}{\nabla_{v_1} h_1} \\ \vdots \\ \frac{\nabla_{g_N} s}{\nabla_{v_N} h_N} \end{bmatrix}_{N\times 1}, \boldsymbol{\theta}_g = \begin{bmatrix} \frac{\nabla_{g_1} \theta}{\nabla_{v_1} h_1} \\ \vdots \\ \frac{\nabla_{g_N} \theta}{\nabla_{v_N} h_N} \end{bmatrix}_{N\times 1},$$

and $\boldsymbol{\gamma} \equiv \nabla_\theta s \begin{bmatrix} \frac{\nabla_{g_1} \theta + \nabla_{g_1} s}{\nabla_{v_1} h_1} \\ \vdots \\ \frac{\nabla_{g_N} \theta + \nabla_{g_N} s}{\nabla_{v_N} h_N} \end{bmatrix}_{N\times 1} = \nabla_\theta s \cdot (\boldsymbol{\theta}_g + \mathbf{s}_g).$

For the observable elements from the data the followings are defined :

$$\mathbf{F}_y^W \equiv \begin{bmatrix} \nabla_{y_1} F_{W|YX} \\ \vdots \\ \nabla_{y_N} F_{W|YX} \end{bmatrix}_{N\times 1}, \mathbf{F}_x^W \equiv \begin{bmatrix} \nabla_{x_1} F_{W|YX} \\ \vdots \\ \nabla_{x_K} F_{W|YX} \end{bmatrix}_{K\times 1},$$

$$\mathbf{G}_x \equiv \begin{bmatrix} \nabla_{x_1} Q_{Y_1|X} & \cdots & \nabla_{x_1} Q_{Y_n|X} \\ \vdots & \ddots & \vdots \\ \nabla_{x_K} Q_{Y_1|X} & \cdots & \nabla_{x_K} Q_{Y_n|X} \end{bmatrix}_{K\times N}.$$

4.2.2 An Example of Random Index

In modeling count or duration data we need to be careful in how to define "endogeneity". In the regression model the distribution of the dependent variable is determined by the distribution of the unobserved variable - usually a scalar unobserved variable. However, duration data or count data are modeled directly by specifying the hazard function of the duration data, or the mean of the count data, and we do not specify the error term. Therefore, the usual perception of endogeneity, as correlation between an explanatory variable and an error term is not applied.

To incorporate endogeneity we introduce unobserved heterogeneity into the structural function to allow for endogeneity. Then endogeneity could occur if there exists dependence between the unobserved heterogeneity and the explanatory variables.

For example, consider the effects of the wage from the previous job on the unemployment duration. It is possible that some of the unobserved factors such as motivation, ability, personality etc will affect both search efforts and wage. It may be the case that individuals with the high unobserved type that determines the wage behave differently from those with low type. When the unemployment duration is the outcome of concern, possibly interval censored, the previous model in Section 4.1 cannot be applied, because the Single Index Unobservable (SIU) restriction cannot hold with duration outcome. See Chesher (2009) for more discussion.

4.2.3 Identification of Partial Derivatives of the Random Index

Restriction R - RI. *There are G restrictions on λ_y, λ_x, γ, and $\mathbf{s}_x$ as follows,*

$$\mathbf{A}_y \cdot \boldsymbol{\lambda}_y + \mathbf{A}_x \cdot \boldsymbol{\lambda}_x + \mathbf{A}_\gamma \cdot \boldsymbol{\gamma} + \mathbf{A}_s \cdot \mathbf{s}_x = \mathbf{a}.$$

The arrays $\underset{G\times N}{\mathbf{A}_y}$, $\underset{G\times K}{\mathbf{A}_x}$, $\underset{G\times N}{\mathbf{A}_\gamma}$, $\underset{G\times K}{\mathbf{A}_s}$ *and* $\underset{G\times 1}{\mathbf{a}}$ *are nonstochastic.* $\nabla_\theta s$ *is finite and nonzero.*

After the following definitions **Theorem 4.2** can be stated.

$$\Phi \equiv \begin{bmatrix} \mathbf{I}_N & \mathbf{0} & \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_K & -\mathbf{G}_x & \mathbf{I}_K \\ \mathbf{A}_y & \mathbf{A}_x & \mathbf{A}_\gamma & \mathbf{A}_s \end{bmatrix}_{(G+N+K)\times(2N+3K)}, \Psi \equiv \begin{bmatrix} \boldsymbol{\lambda}_y \\ \boldsymbol{\lambda}_x \\ \boldsymbol{\gamma} \\ \mathbf{s}_x \end{bmatrix}_{(2N+3K)\times 1},$$

$$\phi \equiv \begin{bmatrix} \mathbf{F}_y^W \\ \mathbf{F}_x^W \\ \mathbf{a} \end{bmatrix}_{(G+N+K)\times 1}.$$

Theorem 4.2 *Under Restriction RI-EH, D-RI, R-RI* $\boldsymbol{\Phi\Psi} = \boldsymbol{\phi}$, *and* $\boldsymbol{\Psi}$ *is identified iff rank* $(\boldsymbol{\Phi}) = 2N + 3K$ *for which a necessary order condition is* $G \geq N + 2K$.

Proof. See Appendix A. ■

4.2.4 Illustration - Constructive Identification

Suppose for simplicity that $N = 1$ and $K = 2$.
Then we have the following :

$$\begin{aligned} y &= h^Y(x, g(y,x)) && (A) \\ Q_{Y|X}(v|x) &= h^Y(x, Q_V(v|x)) && (B) \\ F_{W|YX}(w|y,x) &= s(w, \theta(y,x,g(y,x)), g(y,x), x) && (C-RI) \end{aligned}$$

From (A) and (B), (AB) can be derived as is shown in Chapter 3 :

$$\begin{aligned} \nabla_y g &= \frac{1}{\nabla_v h^Y} \\ \nabla_{x_1} g &= -\frac{1}{\nabla_v h^Y} \cdot \nabla_{x_1} Q_{Y|X}(\tau_V|x) && (AB) \\ \nabla_{x_2} g &= -\frac{1}{\nabla_v h^Y} \cdot \nabla_{x_2} Q_{Y|X}(\tau_V|x). \end{aligned}$$

Identification of ratios of partial derivatives under exclusion restriction

Now differentiate $(C - RI)$ w.r.t. y and $x_k, k \in \{1, 2\}$, we get

$$\begin{aligned}
\nabla_y F_{W|YX} &= \nabla_\theta s \cdot \nabla_y \theta + \nabla_\theta s \cdot \nabla_v \theta \cdot \nabla_y g + \nabla_g s \cdot \nabla_y g \qquad (C - RI) \\
\nabla_{x_1} F_{W|YX} &= \nabla_\theta s \cdot \nabla_{x_1} \theta + \nabla_\theta s \cdot \nabla_v \theta \cdot \nabla_{x_1} g + \nabla_g s \cdot \nabla_{x_1} g + \nabla_{x_1} s \\
\underbrace{\nabla_{x_2} F_{W|YX}}_{\text{"}Data\text{"}} &= \underbrace{\nabla_\theta s \cdot \nabla_{x_2} \theta + \nabla_\theta s \cdot \nabla_v \theta \cdot \nabla_{x_2} g + \nabla_g s \cdot \nabla_{x_2} g + \nabla_{x_2} s}_{\text{Unobservable Structural elements}} .
\end{aligned}$$

Then replacing $\begin{bmatrix} \nabla_y g \\ \nabla_{x_1} g \\ \nabla_{x_2} g \end{bmatrix}$ with $\dfrac{1}{\nabla_v h^Y} \begin{bmatrix} 1 \\ -\nabla_{x_1} Q_{Y|X}(\tau_V|x) \\ -\nabla_{x_2} Q_{Y|X}(\tau_V|x) \end{bmatrix}$ by using (AB) from Chapter 3, we have

$$\begin{aligned}
\nabla_y F_{W|YX} &= \nabla_\theta s \cdot \nabla_y \theta + (\nabla_\theta s \cdot \nabla_v \theta + \nabla_g s) \cdot \frac{1}{\nabla_v h^Y} \qquad (C - RI'') \\
\nabla_{x_1} F_{W|YX} &= \nabla_\theta s \cdot \nabla_{x_1} \theta \ + (\nabla_\theta s \cdot \nabla_v \theta + \nabla_g s) \cdot \frac{1}{\nabla_v h^Y} \cdot (-\nabla_{x_1} Q_{Y|X}) + \underbrace{\nabla_{x_1} s}_{*} \\
\nabla_{x_2} F_{W|YX} &= \nabla_\theta s \cdot \underbrace{\nabla_{x_2} \theta}_{**} + (\nabla_\theta s \cdot \nabla_v \theta + \nabla_g s) \cdot \frac{1}{\nabla_v h^Y} \cdot (-\nabla_{x_2} Q_{Y|X}) + \underbrace{\nabla_{x_2} s}_{*}.
\end{aligned}$$

Without imposing further restrictions, major structural features of interest, $\nabla_y \theta, \nabla_{x_1} \theta$, and $\nabla_{x_2} \theta$ are not identified. The same restrictions as in Chapter 3 are imposed to demonstrate what effects of the index structure have on identification of causal effects.

Firstly, we assume that X affects the outcome only through the index, θ, that is, $\nabla_{x_1} s = 0$ and $\nabla_{x_2} s = 0$ $(*)$. Imposing also local exclusion restriction, $\nabla_{x_2} \theta = 0 (**)$, we have

$$\begin{aligned}
\nabla_y F_{W|YX} &= \nabla_\theta s \cdot \nabla_y \theta + (\nabla_\theta s \cdot \nabla_v \theta + \nabla_g s) \frac{1}{\nabla_v h^Y} \qquad (C - RI''') \\
\nabla_{x_1} F_{W|YX} &= \nabla_\theta s \cdot \nabla_{x_1} \theta \ + (\nabla_\theta s \cdot \nabla_v \theta + \nabla_g s) \frac{1}{\nabla_v h^Y} \cdot (-\nabla_{x_1} Q_{Y|X}) \\
\nabla_{x_2} F_{W|YX} &= \qquad\qquad + (\nabla_\theta s \cdot \nabla_v \theta + \nabla_g s) \frac{1}{\nabla_v h^Y} \cdot (-\nabla_{x_2} Q_{Y|X}) .
\end{aligned}$$

Since $\dfrac{\nabla_\theta s \nabla_v \theta + \nabla_u s}{\nabla_v h^Y} = -\dfrac{\nabla_{x_2} F_{W|YX}}{\nabla_{x_2} Q_{Y|X}}$ from the third equation in $(C - RI''')$, we now have from the first and second equations of $(C - RI''')$

$$
\begin{aligned}
\nabla_\theta s \nabla_y \theta &= \nabla_y F_{W|YX} + \frac{\nabla_{x_2} F_{W|YX}}{\nabla_{x_2} Q_{Y|X}} \\
\nabla_\theta s \nabla_{x_2} \theta &= \nabla_{x_2} F_{W|YX} - \nabla_{x_2} F_{W|YX} \frac{\nabla_{x_2} Q_{Y|X}}{\nabla_{x_2} Q_{Y|X}} \\
\text{if } \nabla_{x_2} Q_{Y|X} &\neq 0.
\end{aligned}
$$

Note that the structural features, $\nabla_y \theta$, and $\nabla_{x_1} \theta$ are not identified, but ratios of them can be identified. Taking ratio of the above we have

$$
\frac{\nabla_\theta s \nabla_{y_1} \theta}{\nabla_\theta s \nabla_x \theta} = \frac{\nabla_{y_1} F_{W|YX} + \frac{\nabla_{x_2} F_{W|YX}}{\nabla_{x_2} Q_{Y|X}}}{\nabla_{x_1} F_{W|YX} - \nabla_{x_2} F_{W|YX} \frac{\nabla_{x_2} Q_{Y|X}}{\nabla_{x_2} Q_{Y|X}}}.
$$

That is, $\frac{\nabla_{y_1} \theta}{\nabla_x \theta}$ is identified by $\frac{\nabla_{y_1} F_{W|YX} + \frac{\nabla_{x_2} F_{W|YX}}{\nabla_{x_2} Q_{Y|X}}}{\nabla_{x_1} F_{W|YX} - \nabla_{x_2} F_{W|YX} \frac{\nabla_{x_2} Q_{Y|X}}{\nabla_{x_2} Q_{Y|X}}}$. Moreover, if the structural function h is differentiable, then ratios of the structural function can be identified since the ratios of the partial derivatives of the structural function should be equal to the ratios of the random index. Note that from the structural function $W = h(\theta(Y, X, V), U_1, ..., U_L)$, partial derivatives are

$$
\begin{aligned}
\nabla_y h &= \nabla_\theta h \nabla_y \theta \\
\nabla_{x_1} h &= \nabla_\theta h \nabla_{x_1} \theta
\end{aligned}
$$

yielding the following

$$
\frac{\nabla_y h}{\nabla_{x_1} h} = \frac{\nabla_y \theta}{\nabla_{x_1} \theta}.
$$

this can be seen using matrices. $(C - RI''')$ can be expressed using these vectors as

$$
\begin{aligned}
\boldsymbol{\lambda}_y + I_M \cdot \boldsymbol{\gamma} &= \mathbf{F}_y^W \\
\boldsymbol{\lambda}_x - \mathbf{G}_x \cdot \boldsymbol{\gamma} + \mathbf{s}_x &= \mathbf{F}_x^W.
\end{aligned}
$$

Then the system of equations that are need to be solved is :

$$
\begin{aligned}
\nabla_\theta s \cdot \nabla_y \theta + (\nabla_\theta s \cdot \nabla_v \theta + \nabla_g s)\frac{1}{\nabla_v h^Y} &= \nabla_y F_{W|YX} \\
\nabla_\theta s \cdot \nabla_{x_1} \theta + (\nabla_\theta s \cdot \nabla_v \theta + \nabla_g s)\frac{1}{\nabla_v h^Y} \cdot (-\nabla_{x_1} Q_{Y|X}) &= \nabla_{x_1} F_{W|YX} \\
\nabla_\theta s \cdot \nabla_{x_2} \theta + (\nabla_\theta s \cdot \nabla_v \theta + \nabla_g s)\frac{1}{\nabla_v h^Y} \cdot (-\nabla_{x_2} Q_{Y|X}) &= \nabla_{x_2} F_{W|YX} \\
\nabla_{x_2}\theta &= 0 \\
\nabla_{x_1} s &= 0 \\
\nabla_{x_2} s &= 0 \\
\nabla_{x_2} Q_{Y|X} &\neq 0.
\end{aligned}
$$

This can be written as using the matrices $\mathbf{\Phi}$, $\mathbf{\Psi}$, and $\boldsymbol{\phi}$,

$$
\text{where } \mathbf{\Phi} \equiv \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & \nabla_{x_1} Q_{Y|X} & -1 & 0 \\ 0 & 0 & 1 & \nabla_{x_2} Q_{Y|X} & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}_{6\times 6}, \mathbf{\Psi} \equiv \begin{bmatrix} \nabla_\theta s \cdot \nabla_y \theta \\ \nabla_\theta s \cdot \nabla_{x_1} \theta \\ \nabla_\theta s \cdot \nabla_{x_2} \theta \\ \dfrac{\nabla_\theta s \nabla_v \theta + \nabla_u s}{\nabla_v h^Y} \\ \nabla_{x_1} s = 0 \\ \nabla_{x_2} s = 0 \end{bmatrix}_{6\times 1}, \boldsymbol{\phi} \equiv \begin{bmatrix} \mathbf{F}_y^W \\ \mathbf{F}_x^W \\ \mathbf{a} \end{bmatrix}
$$

with the homogeneous restrictions indicated by the following matrices as :

$$
\mathbf{A}_y = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \mathbf{A}_x = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{A}_\gamma = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \mathbf{A}_f = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \text{ and } \mathbf{a} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.
$$

The restrictions used are $\nabla_{x_k} s = 0,\ k \in \{1, 2\}, \nabla_{x_2}\theta = 0$, with $N = 1, K = 2$, and $G = 3$. If $\nabla_{x_2} Q_{Y|X} = 0$, then the third row (r_3) of $\mathbf{\Phi}$ is a linear combination of the fourth (r_4) and the sixth (r_6) rows as $r_3 = r_4 - r_6$. Thus, the local rank condition, $\nabla_{x_2} Q_{Y|X} \neq 0$ is required for $rank(\mathbf{\Phi}) = 6$.

4.3 A Model for Discrete Outcomes with Excess Heterogeneity

When we allow for a vector of unobserved elements in the additively non-separable structural function there have been two approaches suggested. Chesher (2009) imposes index restrictions and the index is included as an argument of a structural function which is additively non-separable with a vector of unobservables. The objects of identification are some features of the index. Imbens and Newey (2009) also allows for a vector of unobservables in the structural function for the outcome, and by using the control function approach they identify the *stochastic "average conditional response* (SACR) function" which is a function of the scalar random elements as well as other observable variables.

Use of an additively nonseparable structural relationship is usually motivated by its flexibility that allows for possibly heterogeneous random ceteris paribus impacts of a "cause" variable on the outcome. However, a more compelling reason for using non-separable form would be found in microeconometric models. Many models used in micro-econometrics are inherently not just nonlinear, but additively nonseparable.

Typically, count outcomes are modeled by adopting a specific parametric distribution such as Poisson or Negative Binomial distributions and using MLE for estimation (see Cameron and Trivedi (1998) for detailed discussion of count data modeling). When exogenous unobserved heterogeneity is modeled into the conditional mean function, quasi-MLE can be used for estimation. However, when the unobserved heterogeneity in the mean function is correlated with the explanatory variables, then quasi-MLE estimators are inconsistent. Therefore, when we suspect the exogeneity of regressors in a count data model, we need to consider a new way of identification and inference.

When the structural relationship is nonseparable and there is an endogeneity problem, then inference based on mean independence causes bias. Consider the following a mean regression model of count outcome with mean independence restriction conditional on IV, Z.

$$\begin{aligned} W &= \lambda(Y) + \varepsilon \qquad (*) \\ E[\varepsilon|Z] &= 0. \end{aligned}$$

If W is in fact generated by a nonseparable structural function, which is implied by the count nature of W as the following,

$$W = h(Y, U),$$

then the model specified in eq.(*) will cause bias since[2]

$$E[h(Y, U) - \lambda(Y)|Z] \neq 0$$

The regression error, defined as the difference between the outcome and the mean regression function, would not contain the information on the uncontrollable unobserved heterogeneity that causes endogeneity when the structural relation is non-additive.

Hahn and Ridder (2009) show that when the structural relation is nonseparable, *conditional moment restrictions* (CMR) do not identify Average Structural Function (ASF) which has been a parameter of interest in many studies (see for example, Blundell and Powell (2003,2004) and Imbens and Newey (2009).

[2] $$\begin{aligned} E[\varepsilon|Z] &= E[W - \lambda(Y)|Z] \\ &= E[h(Y, U) - \lambda(Y)|Z], \text{ when the true model is } W = h(Y, U), \\ &\neq 0. \end{aligned}$$

Note that the equality holds only when Y is not endogenous, or the true model is additively separable.

Two points need to be emphasized in modelling discrete outcomes structurally, in other words to allow for endogeneity. Structural functions need to be additively nonseparable. Nonseparability is required to model unobserved heterogeneity that is the cause of endogeneity. Once a nonseparable structural function is used, conditional mean independence restrictions, that are used in Newey and Powell (2003), cannot be used to identify any *structural* parameters. Alternative options would be to use Chesher (2010)'s single equation IV model or Imbens and Newey (2009). In this note I propose another option to identify any structural parameters when the outcome is discrete using the Quantile-based Control Function Approach (QCFA) by Chesher (2003).

Chesher (2003)'s identification results on partial derivatives are not applicable to discrete outcomes. In this section the *stochastic average conditional response* (SACR) function is defined and identification of SACR and partial derivatives of SACR is demonstrated. The objects of interest need to be distinguished from the averaged object of partial derivatives of structural functions studied for example, in Chernozhukov, Fernandez-val, Hahn, and Newey (2008). It is partial derivatives of the averaged "structural" function. This is a new identification result in the sense that we allow for the correlated unobservable heterogeneity in the nonparametric mixture model. Identifying the conditional mean of the discrete outcome is informative in deriving ceteris paribus impacts, such as price/income elasticities, which is impossible to measure with discrete outcome and continuous regressors due to nondifferentiability[3].

4.3.1 An Example - A Count Outcome Model with Correlated Unobserved Heterogeneity

Identification of income elasticity of demand for health care measured as the number of visits to doctors : in this problem endogeneity is of concern because the wealthy tend to be healthy, and thus without controlling for endogeneity the true causal effect of income would not be measured correctly. Several studies examine the income elasticity of demand for health care under the parametric or semiparametric framework. In their studies how to incorporate unobserved heterogeneity which would be the source of endogeneity is limited, for example, in a multiplicative way into the mean function of count data. The model in this section allows for flexible form of interaction between the unobserved heterogeneity and the other explanatory variables.

4.3.2 The Model

Restriction ODO - EH (Ordered Discrete Outcomes with Excess Heterogeneity)

[3] This should be distinguished from the set-identified results of the marginal effects studied in Chernozhukov, Fernandez-val, Hahn, and Newey (2008).

W is a random variable taking values, $w_1 < w_2 < ... < w_M$, $Y \equiv \{Y_i\}_{i=1}^N, X \equiv \{X_i\}_{i=1}^K, U$, *and* $V \equiv \{V_i\}_{i=1}^N$ *are random variables, which are continuously distributed. For any values of X, U, and V, unique values of W and Y are determined by the structural equations*

$$
\begin{aligned}
W &= h(Y, X, U_1, ..., U_L) && \text{(S-1-EH)} \\
Y_n &= h_n^Y(X, V_n) && \text{(S-2)} \\
n &\in \{1, 2, ..., N\}
\end{aligned}
$$

The unobservable latent variables, $U = [U_1, ..., U_L]$ *is a vector and each component of* $\{V_n\}_{n=1}^N$ *is distributed* uniform $(0, 1)$. h *is weakly monotonic in* U. $\{h_n^Y\}_{n=1}^N$ *are strictly* monotonic *with respect to variation in the unobservable* V_n. $\{V_n\}_{n=1}^N$ *should be scalar. This model admits such a case that* $V_n = \theta_{V_n}(V_1, ..., V_{I_n})$, *where* $\theta_{V_n} : R^{I_n} \to (0, 1)$, *for some positive number* I_n, *for* $n = 1, 2, ..., N$.

The inverse function of each h_n^Y with respect to V_n exists by strict monotonicity. It is denoted by g_n. Let $g = [g_1(Y_1, X), ..., g_N(Y_N, X)]'$.

4.3.3 *Stochastic* Average Conditional Response (SACR) Function for Ordered Discrete Outcomes

We consider the mean of ordered discrete outcomes that are characterized by observed and unobserved factors. Define the Stochastic Average Conditional Response (SACR), $\lambda(y, x, v)$, as the conditional mean of an ordered discrete outcome conditional on all the observable explanatory variables (Y and X) and the vector of unobserved variable, V. Note that this function is obtained by integrating out the excess heterogeneity, a vector U, but it is a *stochastic* object since this is varying with the unobserved variable, V. That can be defined by the following

$$
\begin{aligned}
E_{W|YXV}(W \quad | \quad & Y = y, X = x, V = v) \\
&= \int W dF_{W|YXV}(w|y, x, g(y, x)) \\
&= \int h(Y, X, U_1, ..., U_L) dF_{U|YXV}(u|y, x, g(y, x)) \\
&= \sum_{m=1}^{M} w_m P_m^V(y, x, v) \\
&\equiv \lambda(y, x, g(y, x)) \\
&= \lambda(y, x, v)
\end{aligned}
$$

where $P_m^V(y,x,v) \equiv \Pr(W = w_m|Y = y, X = x, V = v) = \Pr(W = w_m|Y = y, X = x, V = g(y,x))$.

From the definition we have

$$\begin{aligned}
\lambda(y,x,v) &= \sum_{m=0} w_m P_m^V(y,x,v) \\
&= \sum_{m=0} w_m P_m(y,x), \\
\text{where } P_m(y,x) &\equiv \Pr(W = w_m|Y = y, X = x)
\end{aligned}$$

where the second equality is due to the fact that

$$\begin{aligned}
P_m^V(y,x,v) &= \Pr(W = w_m|Y = y, X = x, V = v) \\
&= \Pr(W = w_m|Y = y, X = x, V = g(y,x)) \\
&= \Pr(W = w_m|Y = y, X = x).
\end{aligned}$$

Thus we have the identifying relation[4]

$$\underbrace{P_m(y,x)}_{\text{"}Data\text{"}} = \underbrace{P_m^V(y,x,g(y,x))}_{Structural \text{ element}} \qquad (C-ODO)$$

$P_m^V(y,x,v)$, the *stochastic* conditional probability (SCP), should be distinguished from $P_m(y,x)$. $P_m^V(y,x,v)$ is a structural object, which is unobservable, while $P_m(y,x)$ is observed. Note that $(C-ODO)$ holds always. To achieve independent variations, more restrictions need to be imposed.

Note that $\lambda(y,x,v)$ is not the Average Structural Function (ASF). Hahn and Ridder (2009) show that if the structural function is nonseparable, conditional moment restrictions do not recover the ASF when there is endogeneity. The object of interest in this section should be distinguished from the ASF. Another thing to note is that one can see the implication of endogeneity from $(C-ODO)$. If there is no endogeneity, then $\nabla_v P_m^V = 0$, and there is no indirect effect of Y via V as

$$\nabla_y P_m(y,x) = \nabla_y P_m^V(y,x,v) + \underbrace{\nabla_v P_m^V(y,x,v)\cdot\nabla_y g}_{The\ Indirect \text{ Effect}}.$$

Suppose one is interested in how the average response varies when all the observables are fixed. For example, it may be of interest how the average number of visits to doctors vary with unobserved type (how high types respond differently from low types), when the

[4] Consider $N = 1$ and $K = 2$. Suppose X_2 is the IV for Y. Then by imposing the exclusion restricion explicitly, we have from $(C-ODO)$

$$P_m(y,x_1,x_2) = P_m^V(y,x_1,g(y,x_1,x_2)) \qquad (C-ODO)$$

income level and other characteristics are the same. For this purpose, there should be at least one IV that affects household (or individual) income and that is not a determinant of the number of visits to doctors. For $N = 1$, and $K = 2$, if X_2 is a determinant of Y, but excluded in h, then the independent variation of each coordinate of the SACR can be identified at each point of the support of the conditioning variables, $\lambda(y, x_1, g(y, x_1, x_2))$.

4.3.4 Objects of Interest

The object of interest is sensitivity of the (differentiable) average of the discrete outcome to a continuous endogeneous variable. This should be distinguished from the marginal effects studied in Chernozhukov, Fernandez-Val, Hahn and Newey (2009), which is the average of partial derivatives of a nonseparable structural function. Their object is not defined when the outcome is discrete.

4.3.5 Identification of Partial Derivatives of SACR Function

Using the identification result in Theorem 4.3, we can measure the sensitivity by partial derivatives of $\lambda(y, x, v)$.

Restriction D-ODO (Differentiability-ODO). T*he conditional distribution of W given Y and X, $P_m(y, x)$ is differentiable with respect to y and x.*

Define the following vectors and matrices for the structural features

$$\boldsymbol{\xi}_y \equiv \begin{bmatrix} \nabla_{y_1} P_m^V \\ \vdots \\ \nabla_{y_N} P_m^V \end{bmatrix}_{N\times 1}, \boldsymbol{\xi}_x \equiv \begin{bmatrix} \nabla_{x_1} P_m^V \\ \vdots \\ \nabla_{x_K} P_m^V \end{bmatrix}_{K\times 1}, \boldsymbol{\eta} \equiv \begin{bmatrix} \nabla_{v_1} P_m^V \\ \vdots \\ \nabla_{v_N} P_m^V \end{bmatrix}_{N\times 1},$$

$$\mathbf{h}_V \equiv \begin{bmatrix} \nabla_{v_1} h_1^Y & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \nabla_{v_N} h_N^Y \end{bmatrix}_{N\times N}, \mathbf{h}_x \equiv \begin{bmatrix} \nabla_{x_1} h_1^Y & \cdots & \nabla_{x_1} h_N^Y \\ \vdots & \ddots & \vdots \\ \nabla_{x_K} h_1^Y & \cdots & \nabla_{x_K} h_N^Y \end{bmatrix}_{K\times N}.$$

Define also vectors and matrices for functionals of the distributions of observables as follows :

$$\mathbf{F}_y^W \equiv \begin{bmatrix} \nabla_{y_1} P_m \\ \vdots \\ \nabla_{y_m} P_m \end{bmatrix}_{N\times 1}, \mathbf{F}_x^W \equiv \begin{bmatrix} \nabla_{x_1} P_m \\ \vdots \\ \nabla_{x_K} P_m \end{bmatrix}_{K\times 1},$$

$$\text{and } \mathbf{G}_x \equiv \begin{bmatrix} \nabla_{x_1} Q_{Y_1|X} & \cdots & \nabla_{x_1} Q_{Y_n|X} \\ \vdots & \ddots & \vdots \\ \nabla_{x_K} Q_{Y_1|X} & \cdots & \nabla_{x_K} Q_{Y_n|X} \end{bmatrix}_{K\times N}.$$

Restriction R-ODO. *There are G restrictions on ξ_y, ξ_x,* and *η as follows,*

$$\mathbf{A}_y \cdot \boldsymbol{\xi}_y + \mathbf{A}_x \cdot \boldsymbol{\xi}_x + \mathbf{A}_\eta \cdot \boldsymbol{\eta} = \mathbf{a}$$

The arrays $\underset{G\times N}{\mathbf{A}_y}$, $\underset{G\times K}{\mathbf{A}_x}$, $\underset{G\times N}{\mathbf{A}_\eta}$, *and* $\underset{G\times 1}{\mathbf{a}}$ *are nonstochastic.*

$$\boldsymbol{\Phi} \equiv \begin{bmatrix} \mathbf{I}_N & \mathbf{0}_K & \mathbf{I}_N \mathbf{h}_v^{-1} \\ \mathbf{0}_N & \mathbf{I}_K & -\mathbf{G}_x \mathbf{h}_v^{-1} \\ \mathbf{A}_y & \mathbf{A}_x & \mathbf{A}_\eta \end{bmatrix}_{(G+N+K)\times(2N+K)}, \boldsymbol{\Psi} \equiv \begin{bmatrix} \boldsymbol{\xi}_y \\ \boldsymbol{\xi}_x \\ \boldsymbol{\eta} \end{bmatrix}_{(2N+K)\times 1}, \boldsymbol{\phi} \equiv \begin{bmatrix} \mathbf{F}_y^W \\ \mathbf{F}_x^W \\ \mathbf{a} \end{bmatrix}_{(G+N+K)\times 1}$$

Theorem 4.3 *Under Restriction ODO-EH, D-ODO, and R-ODO,* $\boldsymbol{\Phi\Psi} = \boldsymbol{\phi}$, *and* $\boldsymbol{\Psi}$ *can be found iff rank*$(\boldsymbol{\Phi}) = 2N + K$ *for which a necessary condition is* $G \geq N$.

Proof. See Appendix A. ■

4.3.6 Illustration - Constructive Identification

To illustrate how the identification condition can be used to construct the identified point, suppose $N = 1, K = 2$. Then we have the following results.

Corollary 4.2 *Under Restriction ODO-EH,D-ODO,and R-ODO,* $\nabla_y \lambda(y, x, v)$ *is identified by* $\sum_{m=0} w_m \left\{ \nabla_y P_m(y, x) + \frac{\nabla_{x_2} P_m(y, x)}{\nabla_{x_2} Q_{Y|X}(v|x)} \right\}$.

Proof. From $\lambda(y, x, v) = \sum_{m=0} w_m P_m^V(y, x, v)$, it follows that

$$\nabla_y \lambda(y, x, v) = \sum_{m=0} w_m \nabla_y P_m^V(y, x, v) \qquad (*)$$

Then, to identify $\nabla_y \lambda(y, x, v)$, $\nabla_y P_m^V(y, x, v)$ needs to be identified. Since it is specified how the endogenous variable and the unobservable heterogeneity are related by the triangularity, $y = h^Y(x, v)$ and it is assumed that Y is continuous and h^Y is strictly monotonic in V, we write $v = g(y, x)$, where $g(y, x)$ is the inverse function of $h^Y(x, v)$.

Differentiating $(C - ODO)$ w.r.t. y and $x_k, k \in \{1, 2\}$ yields

$$\begin{aligned} \nabla_y P_m(y, x) &= \nabla_y P_m^V(y, x, v) + \nabla_v P_m^V(y, x, v) \cdot \underbrace{\nabla_y g}_{*} \\ \nabla_{x_1} P_m(y, x) &= \nabla_{x_1} P_m^V(y, x, v) + \nabla_v P_m^V(y, x, v) \cdot \underbrace{\nabla_{x_1} g}_{*} \qquad (C - ODO') \\ \nabla_{x_2} P_m(y, x) &= \nabla_{x_2} P_m^V(y, x, v) + \nabla_v P_m^V(y, x, v) \cdot \underbrace{\nabla_{x_2} g.}_{*} \end{aligned}$$

Using (AB) from Chapter 3, replacing $\begin{bmatrix} \nabla_y g \\ \nabla_{x_1} g \\ \nabla_{x_2} g \end{bmatrix}$ with $\frac{1}{\nabla_v h^Y} \begin{bmatrix} 1 \\ -\nabla_{x_1} Q_{Y|X}(v|x) \\ -\nabla_{x_2} Q_{Y|X}(v|x) \end{bmatrix}$, we have

$$\begin{aligned} \nabla_y P_m(y,x) &= \nabla_y P_m^V(y,x,v) + \nabla_v P_m^V(y,x,v) \cdot \frac{1}{\nabla_v h^Y} \\ \nabla_{x_1} P_m(y,x) &= \nabla_{x_1} P_m^V(y,x,v) + \nabla_v P_m^V(y,x,v) \cdot \frac{-\nabla_{x_1} Q_{Y|X}(v|x)}{\nabla_v h^Y} \\ \nabla_{x_2} P_m(y,x) &= \underbrace{\nabla_{x_2} P_m^V(y,x,v)}_{**} + \nabla_v P_m^V(y,x,v) \cdot \frac{-\nabla_{x_2} Q_{Y|X}(v|x)}{\nabla_v h^Y} \end{aligned}$$

Without imposing restrictions we cannot identify $\nabla_y P_m^V(y,x,v)$. Local exclusion restriction, $\nabla_{x_2} P_m^V(y,x,v) = 0$ $(**)$, is imposed. Then we have

$$\begin{aligned} \nabla_y P_m(y,x) &= \nabla_y P_m^V(y,x,v) + \frac{\nabla_v P_m^V}{\nabla_v h^Y} \\ \nabla_{x_1} P_m(y,x) &= \nabla_{x_1} P_m^V(y,x,v) + \frac{\nabla_v P_m^V}{\nabla_v h^Y} \cdot [-\nabla_{x_1} Q_{Y|X}(v|x)] \quad (C-ODO''') \\ \nabla_{x_2} P_m(y,x) &= \qquad + \frac{\nabla_v P_m^V}{\nabla_v h^Y} \cdot [-\nabla_{x_2} Q_{Y|X}(v|x)]. \end{aligned}$$

Then solving for $\nabla_v P_V^m(y,x,v)$ from the third eq. of $(C-ODO''')$ yields

$$\begin{aligned} \frac{\nabla_v P_m^V}{\nabla_v h^Y} &= -\frac{\nabla_{x_2} P_m(y,x)}{\nabla_{x_2} Q_{Y|X}(v|x)} \qquad \text{(Bias-ODO)} \\ \text{if } \nabla_{x_2} Q_{Y|X}(v|x) &\neq 0 \end{aligned}$$

Using this, Replacing $\frac{\nabla_v P_m^V}{\nabla_v h^Y}$ with $-\frac{\nabla_{x_2} P_m(y,x)}{\nabla_{x_2} Q_{Y|X}(v|x)}$ in the first eq. of $(C-ODO''')$ gives us

$$\nabla_y P_m(y,x) = \nabla_y P_m^V(y,x,v) - \frac{\nabla_{x_2} P_m(y,x)}{\nabla_{x_2} Q_{Y|X}(v|x)}.$$

Thus, we finally have the following the identifying relation :

$$\nabla_y P_V^m(y,x,v) = \nabla_y P_m(y,x) + \frac{\nabla_{x_2} P_m(y,x)}{\nabla_{x_2} Q_{Y|X}(v|x)}.$$

Therefore, the sensitivity of the conditional mean to change in the endogenous variable in eq. (*) is now identified by the following relationship.

$$\begin{aligned} \nabla_y \lambda(y,x,v) &= \sum_{m=0}^{M} w_m \nabla_y P_m^V(y,x,v) \qquad \text{(TPD-ODO)} \\ &= \sum_{m=0}^{M} w_m \left\{ \nabla_y P_m(y,x) + \frac{\nabla_{x_2} P_m(y,x)}{\nabla_{x_2} Q_{Y|X}(v|x)} \right\}. \end{aligned}$$

■

Then the system of equations that need to be solved is :

$$
\begin{aligned}
\nabla_y P_m^V(y,x,v) + \nabla_v P_m^V(y,x,v) \cdot \frac{1}{\nabla_v h^Y} &= \nabla_y P_m(y,x) \\
\nabla_{x_1} P_m^V(y,x,v) + \nabla_v P_m^V(y,x,v) \cdot \frac{-\nabla_{x_1} Q_{Y|X}(v|x)}{\nabla_v h^Y} &= \nabla_{x_1} P_m(y,x) \\
\nabla_{x_2} P_m^V(y,x,v) + \nabla_v P_m^V(y,x,v) \cdot \frac{-\nabla_{x_2} Q_{Y|X}(v|x)}{\nabla_v h^Y} &= \nabla_{x_2} P_m(y,x) \\
\nabla_{x_2} P_m^V(y,x,v) &= 0
\end{aligned}
$$

In other words,

$$
\mathbf{\Phi} \equiv \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & \nabla_{x_1} Q_{Y|X}(v|x) \\ 0 & 0 & 1 & \nabla_{x_2} Q_{Y|X}(v|x) \\ 0 & 0 & 1 & 0 \end{bmatrix}_{4\times 4}, \mathbf{\Psi} \equiv \begin{bmatrix} \nabla_y P_m^V(y,x,v) \\ \nabla_{x_1} P_m^V(y,x,v) \\ \nabla_{x_2} P_m^V(y,x,v) \\ \nabla_v P_m^V(y,x,v) \end{bmatrix}_{4\times 1}, \boldsymbol{\phi} \equiv \begin{bmatrix} \nabla_y P_m(y,x) \\ \nabla_{x_1} P_m(y,x) \\ \nabla_{x_2} P_m(y,x) \\ 0 \end{bmatrix}_{4\times 1},
$$

with $N = 1, K = 2, G = 1$, when the restriction imposed is $\nabla_{x_2} P_m^V(y,x,v) = 0$ (which is indicated in the fourth row of $\mathbf{\Phi}$), and $\nabla_{x_2} Q_{Y|X}(v|x) \neq 0$ (local rank condition) so that $rank(\mathbf{\Phi}) = 4$, satisfying the condition of Theorem 4.4. Once $\nabla_y P_m^V(y,x,v)$ is identified, $\nabla_y \lambda(y,x,v)$ can be identified by (TPT-ODO).

Discussion :

1. A similar three-part decomposition as in Chesher (2003) is obtained.

2. In comparison with the usual control function approach as discussed in Blundell and Powell (2003, 2004), stochastic sensitivity can be identified.

3. Count outcomes are a special case of ordered discrete outcomes with $w_m = m$, for integer $m = 0, 1, 2, \ldots$.

4. Although a binary outcome is not ordered, it can be considered to be a special case with $N = 2$. Thus, for the binary outcome we have

$$
\begin{aligned}
\nabla_y \lambda(y,x,v) &= \nabla_y P(y,x) + \frac{\nabla_{x_2} P(y,x)}{\nabla_{x_2} Q_{Y|X}(v|x)}, \\
\text{where } P(y,x) &= \Pr(W = 1 | Y = y, X = x).
\end{aligned}
$$

5. A testable expression for the exogeneity of Y can be derived even when the outcome is discrete. By testing $H_0 : \nabla_{x_2} P(y,x) = 0$ we could test the exogeneity of Y locally. This can be seen from (Bias-ODO).

$$
\begin{aligned}
\frac{\nabla_v P_m^V}{\nabla_v h^Y} &= -\frac{\nabla_{x_2} P_m(y,x)}{\nabla_{x_2} Q_{Y|X}(v|x)} \qquad \text{(Bias-ODO)} \\
\text{if } \nabla_{x_2} Q_{Y|X}(v|x) &\neq 0
\end{aligned}
$$

If there is no endogeneity, the indirect effect, $\frac{\nabla_v P_m^V}{\nabla_v h^Y} = 0$. In other words, if Y and U are independent, $\nabla_v P_m^V = 0$.

4.3.7 *Stochastic* Conditional Probability(SCP) Function for Categorical Outcomes

When the outcome is not ordered, but categorical, the average does not deliver any meaning. When the outcome is categorical, the *stochastic* conditional probability (SCP) function could be considered to be the structural object of the identification study. Note that the identification results in Theorem 4.3 can be used to identify the SCP.

Chapter 5

Discrete Endogenous Variables

5.1 Introduction

This chapter demonstrates how *additively nonseparable* structural functions are used in measuring heterogeneous causality and provides a model that identifies *individual* treatment effects. This has not been studied using the Hurwicz (1950a) structure. Restrictions are imposed on the *shape* of the Hurwicz (1950a) structure. The novel restriction exploits the fact that the patterns of endogeneity may vary across the level of the unobserved variable. The proposed model does not require differentiability of the structural functions nor continuity of observed variables. The model does not impose weak separability. It can be used to recover some *partial* information on individual-level causal effects of a discrete variable by identifying the partial difference of a nonadditive structural function. In this chapter I assume that every individual is distinguished by their observed characteristics and the rankings in the distribution of their unobserved characteristics, and show that *individual*-specific counterfactual outcomes and causal effects can be partially identified using a control function approach.

5.1.1 Causality, Heterogeneity, and Nonseparable Structural Relations

Suppose we are interested in the impact of a variable (Y) chosen by individuals on their outcome (W) of interest, and suppose the economic decisions on W and Y can be described by the following triangular system

$$\begin{aligned} W &= h(Y, X, U) \\ Y &= h^Y(Z, X, V), \end{aligned} \tag{1}$$

where X is a vector of characteristics that are exogenously given to individuals such as age, gender, and race, Z is an exogenous covariate that is excluded in h, and U and V are normalized scalar indices of unobservable (possibly) multidimensional individual characteristics. Various unobserved factors can affect the outcome and the choice, but they are assumed to do so, only through the scalar indexes taking values between 0 and

1. The structural relations may be derived from some optimization processes such as demand/supply functions. If there is not a well-defined economic theory behind them, then the structural relations can be simply understood as how the outcome and the choice are determined by other relevant (both observable and unobservable) variables. The structural relations deliver the information on "contingent" plans of choice or outcomes when different values of X, Z, U and V are given. Even among the individuals with the same observed characterisics we observe a distribution of the outcome due to the unobserved elements, U and V. The conditional distribution of the outcome, $F_{W|YX}$, is determined by the distribution of the unobserved elements, $F_{U|YX}$ and the structural relation, $h(\cdot, \cdot, \cdot)$.

Causal effects of a variable indicate the effects of the variable only, separated from other possible influences. This counterfactual information is contained in the partial differences of the structural relation. When the outcome is determined by (1), the causal effects of changing the value of Y from y^a to y^b on W, other things being equal (the ceteris paribus effects), would be measured by the partial difference of the structural function, h

$$\Delta(y^a, y^b, x, u) \equiv h(y^a, x, u) - h(y^b, x, u)$$

for some fixed values of $X = x$ and $U = u$. Individuals with different values of X and U may have different values of $\Delta(y^a, y^b, x, u)$, thus, heterogeneity can constitute of both observed and unobserved components.

When Y is binary, the ceteris paribus effect of Y can be expressed by

$$\Delta(1, 0, x, u) = h(1, x, u) - h(0, x, u).$$

Adopting the notation of the potential outcomes framework, let W_{di} denote the hypothetical outcome with $Y = d$ for the individual i whose observed and unobserved characteristics are x and u[1]. Suppose there are binary choices and let $d \in \{0, 1\}$. If we can assume that W_{1i} and W_{0i} are generated by the structural relation then we can write

$$W_{1i} - W_{0i} = h(1, x, u) - h(0, x, u). \tag{2}$$

This way we map the problem in the potential outcomes framework into the structural approach[2]. By this relation the interpretation of $h(1, x, u) - h(0, x, u)$ as the individual-specific treatment response is *justified.*

Identification of causal effects calls for special attention if there is endogeneity or selection problem. Y is called endogenous if U and Y are not independent. The selection problem exists if the distributions of counterfactual outcomes, W_0 and W_1 are different

[1] See Heckman, Florens, Meghir, and Vytlacil (2008) for average effects of continuous treatment, and Angrist and Imbens (1995), and Nekipelov (2009) for average effects of multi-valued discrete treatment.

[2] By the structural approach we mean the sort of analysis in classical simultaneous equations systems model. This should be distinguished from "structural estimation" where the underlying optimization processes such as preferences are fully specified. Rather, the structural approach I am considering simply assumes the existence of decision processes which can be expressed as relationships between variables. Further specification of the decision processes is not required to be specified.

from each other[3]. The identification problem in the potential outcomes approach (identification of the object on the left) is caused by the fact that either W_{1i} or W_{0i} is observed, but not both. Thus, the difference of the two for each individual is never observed and cannot be replaced by observed W_i if there exists the selection problem. Difficulties in identification of the structural function (identification of the object on the right) arise because observed information from the relevant variables does not necessarily guarantee the information on independent variation in each coordinate of the structural relation.

The potential outcomes approach does not utilize the information on the economic processes that generate the potential outcomes. Instead of $W_{1i} - W_{0i}$, this paper focuses on identification of $h(1, x, u) - h(0, x, u)$, by assuming the existence of economic processes and by imposing restrictions on such decision mechanisms. See the recent debate between Deaton (2009) and Imbens (2009)[4]. The proposed model can be used to identify the signs of individual treatment responses. This model would be particularly informative when the signs of individual effects vary across the population, in which case average effects would underestimate the true effects with different signs being cancelled out.

In contrast with the triangular system, switching regression models with a selection equation of the following form have been widely used :

$$\begin{aligned} W_0 &= h_0(0, X, U_0) \\ W_1 &= h_1(1, X, U_1) \\ Y &= g(Z, X, V),\ Y \in \{0, 1\}. \end{aligned} \tag{3}$$

The counterfactual outcomes are determined by distinct functional relations, h_0 and h_1, and the unobserved heterogeneity for the two counterfactual events, U_0 and U_1, are allowed to be different. Individual causal effects would be measured by $h_1(1, X, U_1) - h_0(0, X, U_0)$, not by the partial difference of h_0 nor h_1.

5.1.2 Contributions

This chapter contributes to the nonparametric identification literature by providing new identification results on additively nonseparable structural functions when an endogenous variable is discrete/*binary* by using a control function approach. Non-additive structural functions are used to model heterogeneity. One of the key implications of additively

[3] If the counterfactual distributions are distinct from each other even after controling for observable characteristics, there is selection on unobservables. Selection on unobservables is the case I am considering in this paper.

[4] We advocate the structural approach for two reasons : as Deaton (2009) and Heckman and Urzua (2009) argue econometric models guided by economic models provide clearer interpretation of data analysis. Moreover, assuming the existence of a structure derived from an economic model allows us to use restrictions that may be justified by economic arguments such as monotonicity or concavity of structural relation, which can result in identification of some parameters of interest. In contrast with Imbens (2009)'s arguments, when a specific structural feature is aimed to be recovered (not the whole structure), the structural approach helps, rather than hinders, inference of causal information from data. On the other hand, the applicability may be limited to the extent that the restrictions can be justified since the identifying power comes from such restrictions.

nonseparable functional form is that partial differences are themselves stochastic objects that have distributions. Thus, heterogeneity in individual causal effects can be found by identifying partial differences of a non-additive structural function. However, individual-specific causal effects have not been discussed so far.

On the one hand, in the *structural* approach many studies dealing with endogeneity focus on identification of the structural function itself, rather than its partial differences, however, identification of partial differences is not necessarily guaranteed from the knowledge of identification of structural function when it is non-additive. Existing identification results of a nonadditive structural function are not applicable to identification of partial difference of a nonadditive function with respect to a binary endogeneous variable. Single equation IV models as in Chernozhukov and Hansen (2005) and Chesher (2010) do not guarantee identification of partial differences. Imbens and Newey (2009)'s control function approach is not applicable to discrete endogenous variables. Chesher (2005) report identification resutls of partial difference with respect to an ordered discrete endogenous variable, but it is not applicable to a binary endogenous variable.

On the other hand, individual treatment effect is not recovered from the *potential outcomes* approach since both counterfactual outcomes are never observed. Instead, usually average effects are the focus of interest. Several papers (see Imbens and Rubin (1997), Abadie (2002), and more recently, Chernozhukov, Fernandez-Val, and Melly (2010), Kitagawa (2009), for example) focus on identification of the marginal distribution of the counterfactuals whose information may be useful in recovering QTE, but individual treatment effect cannot be recovered from the information on the marginal distributions of the potential outcomes.

Another distinct feature of the proposed model is that the identifying power does not come from restrictions on data. In this paper nonparametric *shape* restrictions on the structure are imposed, rather than relying on properties of observed variables. Nonparametric identification under endogeneity often relies on the characteristics of IV/exogenous variables - many results exploit continuity, rich support in exogenous variation, large support conditions or certain rank conditions. Such results therefore may have limited applicability since many microeconomic variables are discrete or show limited variation in the support. In contrast with other studies, the new results in this paper can be applied to a discrete, including binary, endogenous variable when the IV is binary or when the IV is weak. The proposed model does not require differentiability of the structural function and thus, can be applied to discrete outcomes. The proposed weak rank condition can be applied to examples such as regression discontinuity designs, a case with a binary endogenous variable or weak IV or a binary IV.

5.1.3 Related Studies

Since Roehrig (1988)'s recognition of the importance of nonparametric identification, there have been many studies that aim to clarify what can be obtained from data without parametric restrictions (see Matzkin (2007) for survey on nonparametric identification and the references therein). When parametric assumptions are avoided, point identification is often not possible[5] with a discrete endogenous variable. In such cases one could aim to define a set in which the parameter of interest can be located. This partial identification idea, which was pioneered by Manski (1990, 1995, 2003), has been actively used in the setup that can be interpreted as a missing data problem - selection or (interval) censoring as examples (Manski (1990), Balke and Pearl (1997), Manski and Pepper(2000), Cross and Manski (2002), Manski and Tamer (2002), Heckman and Vytalcil (1999), Blundell, Gosling, Ichimura, and Meghir (2007), Chernozhukov, Riggobon and Stoker (2009), for example). It has been expanded into other economic models such as consumer demand or labor supply analyses by adopting the restrictions from economic theory recently (Blundell, Browning, and Crawford (2007), Hoderlein and Stoye (2009), and Chetty (2009)). Set identification defined by moment inequalities has been used in entry models (see Berry and Tamer (2007) for the recent survey), panel data models (Honore and Tamer (2006)), discrete outcomes (Chesher (2010)), for example.

Many authors[6] emphasize the existence of heterogeneity in individual responses in practice. The importance of the information regarding individual-specific, possibly heterogeneous causal effects of a binary endogenous variable was recognized earlier. Many interesting parameters are functionals of the distribution of individual treatment effects as Heckman, Smith, and Clements (1997) noted. In contrast with average treatment effects which are found by a linear operator, other functionals such as quantiles require the knowledge of the distribution of the individual treatment effects[7].

One approach to recover individual-specific causal effects has been taken to recover heterogeneity in treatment effects by identifying the distribution of $W_1 - W_0$ directly[8]. Heckman, Smith and Clements (1997) use the Hoeffding-Frechet bounds, and Fan and Park (2009) and Firpo and Ridder (2008) used Makarov bounds to derive information on the distribution of the treatment effects from the "known" marginal distributions of the potential outcomes.

Alternatively, some information regarding heterogeneity can be recovered by using quantiles[9]. One particular object that has been the focus of research is the quantile

[5] Under the "complete" system of equations as Roehrig (1988) and Matzkin (2008), identification analysis relies on differentiability and invertibility of the structural functions. However, differentiability and invertibility fail to hold with discrete endogenous variables. Another well known example is discussed by Heckman (1990) using the selection model - without parametric assumptions point identification is achieved by the identification at infinity argument, which may not hold in practice.

[6] See, for example, Heckman (2000).

[7] When the treatment effects are homogeneous the problem is trivial and the distribution of the treatment effects is degenerate.

[8] The quantiles of treatment effects recovered from the distribution of $W_{1i} - W_{0i}$ are examples of $D\Delta-$treatment effects, while the quantile treatment effects (QTE) are examples of $\Delta D-$treatment effects discussed in Manski (1997). Neither of them is implied by the other, and they deliver different information regarding distributional consequences of any policy. As Firpo and Ridder (2008) nicely discussed, $\Delta D-$treatment effects, such as QTE can deal with the issues such as the impact of a policy on the society (population) in general, while $D\Delta-$treatment effects can be used to address issues such as policy impacts on "individuals".

[9] By estimating quantile treatment effects (QTE) using the Connecticut experimental data Bitler, Gelbach, and Hoynes (2006) found that welfare reforms in the ninties had heterogeneous effects on individuals as predicted by labour supply theory. They conclude that "welfare reform's effects are likely both more varied and more extensive". Average effects may miss much information and can be misleading if the

treatment effect (QTE) defined by Lehman (1974) and Doksum (1974). The QTE can be found from the marginal distributions in principle. Abadie, Angrist, and Imbens (2002) study the QTE under the LATE-type assumptions using a linear quantile regression model, Firpo (2007) under the matching assumption, and Frolich and Melly (2009) under the regression discontinuity design. Chernozhukov and Hansen (2005)'s moment condition based on their IV-QR model provides a way to estimate QTE.

Alternative to these potential outcomes setup, one could use the structural approach. By adopting a triangular structural setup, Chesher (2003,2007) studies identification of $\Delta(y^a, y^b, x, u)$ when Y is continuous, by the quantile-based control function approach (QCFA, hereafter). Chesher (2005) showed how the QCFA proposed by Chesher (2003) can be used to find the intervals that the values of the structural function lie in when the endogenous variable is ordered discrete with more than three points in the support. Jun, Pinkse, and Xu (2010) report *tighter* bounds when a different rank condition from Chesher (2005)'s is used, while other restrictions on the structure in Chesher (2005) are adopted. Jun, Pinkse, and Xu (2010) does not have identifying power for a binary endogenous variable if the IV is binary. Vytalcil and Yildiz (2007) use a triangular system and report a point identification result of average treatment effect of a dummy endogenous variable. They impose weak separability and exclusion restriction. Their result rely on certain characteristics of variation in exogenous variable and excluded variables to achieve point identification. Vytalcil and Yildiz (2007) results does not guarantee identification of partial difference - Jun, Pinkse, and Xu (2009) focus on identification of the structural function, and Vytalcil and Yildiz (2007) focus on identification of average effect, not the structural function. Manski and Pepper (2000) and Bhattacharya, Shaikh, and Vytlacil (2008) have partial identification results on average effects. They exploit different monotonicity restrictions to this paper. More discussions on these studies can be found in Section 5.5.

The remaining part is organized as follows. Section 5.2 introduces the model for "ordered" discrete endogenous variables and contains the main results on identification. Section 5.3 discusses "unordered" binary endogenous variable as a different case of discrete endogenous variable. We also discuss the testability of the restrictions imposed by our model. Section 5.5 discusses several relevant points to the model proposed. I then illustrate the possibly useful information derived from our identification results by examining the effects of the Vietnam-era veteran status on the civilian earnings in section 5.6. Section 5.7 concludes.

---

signs of individual treatment effects are varying across people. However, when experimental data are not available, QTE does not have causal interpretation on *individuals* because individuals' rankings in the two marginal distributions of the potential outcomes may change. Our model could be used to determine who benefits by identifying the signs of treatment effect of individuals with different rankings of the *scalar* unobserved heterogeneity even with observational data.

5.2 Local Dependence and Response Match (LDRM) Model - $\mathcal{M}^{LDRM}$

5.2.1 Restrictions of the Model $\mathcal{M}^{LDRM}$

I introduce a model that interval identifies the value of the structural function evaluated at a certain point in the presence of an endogenous discrete variable by applying the QCFA. The model, $\mathcal{M}^{LDRM}$, is defined as the set of all the structures that satisfy the restrictions[10].

Restriction A-EX : Scalar Unobservables Index (SIU)/Monotonicity/Exclusion[11]

$$\begin{aligned} W &= h(Y, X, U), \\ Y &= h^Y(Z, X, V), \\ \text{with } h^Y(z, x, v) &= y^m, \ P^{m-1}(z, x) < v \leq P^m(z, x), \\ m &\in \{1, 2, ..., M-1\} \end{aligned}$$

The function h is weakly increasing[12] with respect to variation in scalar U. From here on other exogenous variables, X, than Z are ignored. X can be added as a conditioning variables in any steps of discussion without changing the results.

The variable W is a discrete, continuous, or mixed discrete continuous random variable. The conditional distribution of Y given $Z = z$ is discrete with points of support $y^1 < y^2 < ... < y^M$, invariant with respect to z and with positive probability masses $\{p_m(z)\}_{m=1}^M$. Cumulative probabilities $\{P^m(z)\}_{m=1}^M$ are defined as

$$\begin{aligned} P^m(z) &\equiv \sum_{l=0}^{m} p_l(z) = F_{Y|Z}(y^m|z), \quad m \in \{1, 2, ..., M\}, \\ p_0(x) &\equiv 0. \end{aligned}$$

The latent variates are jointly continuously distributed and they are normalized uniformly distributed on $(0, 1)$ independent of Z. The value $y^m, m \in \{2, ..., M-1\}$, is an interior point of support of the distribution of Y.

The function g evaluated at $Z = z$, $g(z, \tau_V)$ is identified by $Q_{Y|XZ}(\tau_V|z)$. The monotonicity restriction on Y is reflected in the threshold crossing structure.

Restriction RC (Rank Condition)[13] There exist instrumental values of Z, $\{z'_m, z''_m\}$,

[10] I adopt this definition of a model as a set of structures satisfying the restrictions imposed, following Koopmans and Reiersol (1950).

[11] Triangularity assumption enables us to avoid the issue of coherency that may be caused due to discrete endogenous variables when the outcome is discrete.

[12] If h^Y is weakly increasing in v, then if h is weakly increasing in u and if h^Y is weakly decreasing, h should be weakly decreasing as well. This monotonicity restriction is one of the two key restrictions in QCFA identification strategy. This enables us to use the equivariance property of quantiles. In many applications this can be justified - under certain regularity conditions many optimization frameworks predict that the equilibrium relations are monotonic in certain variables - law of demand as a typical example.

[13] Restriction RC is related to the "relevance" condition for IV. If Z is a strong IV, Restriction RC is

such that

$$P^m(z'_m) \leq \tau_V \leq P^m(z''_m)$$

for $m \in \{0, 1, 2, ..., M-1\}$.

Restriction C-QI (Conditional Quantile Invariance) : The value of U, $u^* \equiv Q_{U|VZ}(\tau_U|\tau_V, z)$ is invariant with $z \in \overline{z}_m \equiv \{z'_m, z''_m\}$ for $P^m(z'_m) \leq \tau_V \leq P^{m-1}(z''_m)$.

Define $\boldsymbol{V} \equiv (V_L, V_U]$, where $V_L = \max_{z \in \overline{z}_m} P^{m-1}(z)$, and $V_U = \max_{z \in \overline{z}_m} P^{m+1}(z)$.[14] Define also $\boldsymbol{U} \equiv (U_L(z), U_U(z)]$, where $U_L(z) = \min_{\tau_V \in \boldsymbol{V}} Q_{U|VZ}(\tau_U|\tau_V, z)$, and $U_U(z) = \max_{\tau_V \in \boldsymbol{V}} Q_{U|VZ}(\tau_U|\tau_V, z)$. The value, u^*, is not known, but it indicates τ_U- ranked individual's value of U in the conditional distribution of U given V and Z. The case in which $F_{U|VZ}(u^*|v, z)$ is nonincreasing in v, for $u^* \in U$ is called PD (Positive Dependence) and the other case, ND (Negative Dependence). The case in which $h(y^{m+1}, u^*) \geq h(y^m, u^*)$ is called PR (Positive Response) and the other case, NR (Negative Response).

Restriction LDRM (Local (Quantile) Dependence Response Match) : $F_{U|VZ}(u|v, z)$ is weakly monotonic in $v \in \boldsymbol{V}$ for $u \in \boldsymbol{U}$. If $F_{U|VZ}(u|v, z)$ is weakly decreasing in $v \in \boldsymbol{V}$ for $u \in \boldsymbol{U}$, then $h(y^{m+1}, u^*) \geq h(y^m, u^*)$, **(PDPR)** and if $F_{U|VZ}(u|v, z)$ is weakly increasing in $v \in \boldsymbol{V}$ for $u \in \boldsymbol{U}$, then $h(y^{m+1}, u^*) \leq h(y^m, u^*)$, **(NDNR)** for any $u^* \in \boldsymbol{U}$ for $m \in \{0, 1, 2, ..., M-1\}$. See <Figure 5.1>.

5.2.2 Discussion

Restriction A-EX

This is fundamental restrictions imposed in the quantile-based control function method in Chesher (2003). Monotonicity of the structural functions in the scalar indices of unobserved factors and the existence of Z that is excluded from the outcome equation are key features together with independence between U and Z.

There is a tradeoff between using a vector and a scalar unobserved heterogeneity - allowing for a vector unobserved heterogeneity in the structural relation would be more realistic. Several studies report identification results without monotonicity restriction (See Altonji and Matzkin (2005), Hoderlein and Mammen (2007), Imbens and Newey (2009), and Chalak, Schennach, and White (2008), and Chernozhukov, Fernandez-Val,

satisfied. Chesher (2005)'s rank condition is that there exist values of Z, z'_m, and z''_m such that

$$P^m(z'_m) \leq \tau_V \leq P^{m-1}(z''_m)$$

thus, if Chesher (2005)'s rank condition holds, our rank condition also holds since $P^{m-1}(z'') \leq P^m(z'')$. In this sense, Chesher (2005)'s rank condition is stronger than our rank condition. Note also that Chesher (2005)'s strong rank condition is not satisfied when the instrument is weak or when a binary endogenous variable is present.

[14] For a binary endogenous variable $V \equiv [0, 1]$.

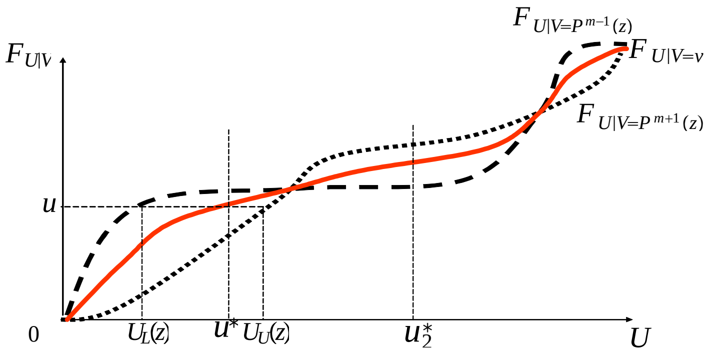


Figure 5.1: Distributions of U given V are drawn for different values of V by assuming monotonity in V. The thick line is the distribution of U given $V = v$. A point in the support of U, u^* can be written as τ_U-quantile of U given $V = v$. "Local" nature of Restriction LDRM : the information on endogeneity is contained in $F_{U|V}$ - if Y is exogenous, then $F_{U|V}$ is invariant with values of V. Monotonicity of $F_{U|V}(u^*|v)$ does not have to be global in U, all that is required is monotonicity in some region U of u. In this graph, for $v' \leq v'' \leq v'''$, $F_{U|V}(u^*|v)$ is decreasing in v, while $F_{U|V}(u_2^*|v)$ is increasing in $v \in V$.

and Newey (2009) for identification analysis without monotonicity). However, what can be identified without monotonicity is objects with the heterogeneity in responses averaged out. On the other hand, the quantile approach under monotonicity can be adopted to recover heterogeneous treatment responses if a scalar (index) unobserved heterogeneity is assumed, however, this may be considered to be restrictive since some of the examples such as models with measurement error cannot be dealt with. See Chesher (2009) for examples where the unobserved elements cannot be collapsed into a scalar index.

Local Dependence and Response Match (LDRM)

Endogeneity is roughly defined as the dependence between an explanatory variable and the unobserved elements in the structural relationship. They can be positively dependent or negatively dependent. "Dependence" is used instead of "correlation" to clarify the *local* information contained in Restriction LDRM. Under the triangularity in the setup of this chapter the source of endogeneity is caused by the dependence between U and V. This information is contained in conditional distribution of $F_{U|V}$.

Restriction LDRM assumes first that $F_{U|V}(u|v)$ is monotonic in v in certain ranges of U and V. Then it restricts the direction of the dependence in that range and the direction of the response - whether the response is positive or negative or zero. For example, college graduates may be different from high school graduates in terms of ability (U) when other observed characteristics are identical. Restriction LDRM is concerned with how the patterns of dependence varies with the level of the unobserved characteristic. It may be the case that individuals with very low ability are not allowed to get into college due to low test scores, on the other hand, individuals with extremely high ability may not choose to go to college if they have better options that will lead to higher income. This example shows the possibility that there is positive dependence with the low level of ability, and negative dependence with the high level.

Discrete Data

The restrictions imposed do not require continuity/differentiability of structural relations, nor rely on continuity of covariates/large support condition. This makes the proposed model more useful since many variables in microeconometrics are in the form of discrete or censored.

5.3 Main Results

5.3.1 Bound on the Value of the Structural Relation

We have the following interval identification for $h(y^m, u^*)$ for $m \in \{1, 2, ..., M-1\}$, where $u^* = Q_{U|VZ}(\tau_U|\tau_V, z)$. For $m = M$, the bound in **Theorem 5.1** is not applied[15].

Theorem 5.1 Under Restriction A-EX,C-QI,RC,and LDRM, there are the inequalities for $m \in \{0, 1, 2, ..., M-1\}$ and $\tau \equiv \{\tau_U, \tau_V\}$

$$\begin{aligned}
q_m^L(\tau, y^m, \overline{z}_m) &\leq h(y^m, u^*) \leq q_m^U(\tau, y^m, \overline{z}_m) \\
\text{where } u^* &= Q_{U|VZ}(\tau_U|\tau_V, z), \\
\text{for some } \tau_U &\in (0,1) \text{ and } \tau_V \in [P^m(z'_m), P^m(z''_m)], \\
z &\in \overline{z}_m = \{z'_m, z''_m\}, \\
q_m^L(\tau, y^m, \overline{z}_m) &= \min\{Q_{W|YZ}(\tau_U|y^m, z'_m), Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)\}, \\
q_m^U(\tau, y^m, \overline{z}_m) &= \max\{Q_{W|YZ}(\tau_U|y^m, z'_m), Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)\}.
\end{aligned}$$

The interval is **not** empty.
Proof. See the Appendix. ■

To identify all the values of the structural function, say, $h(y^1, u^*), h(y^2, u^*), ..., h(y^{M-1}, u^*)$, for fixed u^*, we need to guarantee the rank condition holds for all $m \in \{1, 2, ..., M-1\}$. There should exist two values of Z, $\{z'_m, z''_m\}$ for each m, such that $P^m(z'_m) \leq \tau_V \leq P^m(z''_m)$. Therefore, how closely y and z are related and whether we have enough variation in Z are key to the identification of the whole function.

5.3.2 Sharpness

Suppose a set identifies the value of the structural feature. Then *all* distinct "*admitted*" structures that are "*observationally equivalent*" to the true structure should produce values of the structural feature that are contained in the identified set. All such structures that generate a point in the set, are indistinguishable by data. If the identified set is *not* sharp,

[15] The bounds cannot be applied to $m = M$. This restricts the identification results when $M = 2$, as we will see in the next section.

some of the points in the set are not possible candidates for the value of the structural feature, which would make the identified set less informative. A sharp identified set contains *all and only* such values that are generated by admitted and observationally equivalent structures.

Different points in a sharp identified set may have been generated by different structures, but the distinct structures (i) should all satisfy the restrictions of the model (consistent with the model), (ii) should be observationally equivalent (consistent with the data), and (iii) any point in the interval should be considered to be the possible value of the structural feature. (Lemma 2.1 in Chapter 2).

Common support restriction is imposed for sharpness.

Restriction CSupp (Common Support) The support of the conditional distribution of W given Y and Z has support that is invariant across the values of Y and Z.

Theorem 5.2 Under Restrictions CSupp, A-EX,C-QI,RC,and LDRM, the bound $I(\tau, y^m, \overline{z}) \equiv [q_m^L(\tau, y^m, \overline{z}_m), q_m^U(\tau, y^m, \overline{z}_m)]$, specified in **Theorem 1** for each $m = 0, 1, 2, .., M-1$ and for some $\tau \equiv \{\tau_U, \tau_V\}$, is sharp.

Proof. Use Lemma 2.1 in Chapter 2. See the Appendix. ■

5.3.3 Many Instrumental Values, Overidentification, and Refutability

If there are many pairs of values of Z that satisfy Restriction RC (overidentification), then each pair defines the causal effect for a *different* subpopulation defined by each pair and each identified set is sharp. However, taking intersection of each identified set is *not* a sharp identified set by $\mathcal{M}^{LDRM}$ as is discussed **in Chapter 2**. To use all the information available from data and to justify taking intersection of each set defined by distinct pairs of values of Z in producing *a sharp identified set* in this case, a different restriction is imposed[16].

Let $SUPP(Z)$ be the support of Z. Define $\mathbf{V}_m \equiv [P^m(z_m'), P^m(z_m'')]$ for the pair, $\{z_m', z_m''\}$ that satisfies Restriction RC. Each pair defines different subpopulation over which a causal interpretation is given. Define $\mathcal{Z}_m$ as the set of pairs of $\{z_m', z_m''\}$ that satisfies Restriction RC, $\mathcal{Z}_m \equiv \{\overline{z}_m : P^m(z_m') \leq \tau_V \leq P^m(z_m''), \text{ with } \overline{z}_m = \{z_m', z_m''\}\}$. Let $\mathbf{V}_m(\mathcal{Z}_m) \equiv \{\mathbf{V}_m(\overline{z}_m) : \overline{z}_m \in \mathcal{Z}_m\}$ be a class of the set defined by $\mathcal{Z}_m$. Denote $\mathbb{V} \equiv \cap_{\mathcal{Z}_m} V_m(\overline{z}_m)$.

Restriction C-QI^M (Conditional Quantile Invariance with Many Instrumental Values) : The value of U, $u^* \equiv Q_{U|VZ}(\tau_U|\tau_V, z)$ is invariant with *all* $z \in \overline{z}_m (\in \mathcal{Z}_m)$.

[16] See Manski (1997)'s description of sharp sets as "they exhaust the information from the data".

If the conditional distribution, $F_{U|VZ}(u|v,z)$ is independent of Z, then Restriction **Restriction C-QI^M** always holds.

Corollary 5.1 Under Restriction QCFA,C-QI^M, RC,and LDRM, there are the inequalities for $m \in \{0,1,2,...,M-1\}$, $\tau \equiv \{\tau_U, \tau_V\}$,

$$\begin{aligned}
Q^L_m(\tau, y^m, \mathcal{Z}_m) &\leq h(y^m, u^*) \leq Q^U_m(\tau, y^m, \mathcal{Z}_m) \\
\text{where } u^* &= Q_{U|VZ}(\tau_U|\tau_V, z), \\
\text{for some } \tau_U &\in (0,1) \text{ and } \tau_V \in \mathbb{V} \equiv \cap_m V_m(\overline{z}_m) \\
Q^L_m(\tau, y^m, \mathcal{Z}_m) &= \max_{\overline{z}_m} \ q^L_m(\tau, y^m, \overline{z}_m), \ \overline{z}_m \in \mathcal{Z}_m \\
Q^U_m(\tau, y^m, \mathcal{Z}_m) &= \min_{\overline{z}_m} \ q^U_m(\tau, y^m, \overline{z}_m), \ \overline{z}_m \in \mathcal{Z}_m \\
q^L_m(\tau, y^m, \overline{z}_m) &= \min\{Q_{W|YZ}(\tau_U|y^m, z'_m), Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)\} \\
q^U_m(\tau, y^m, \overline{z}_m) &= \max\{Q_{W|YZ}(\tau_U|y^m, z'_m), Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)\}
\end{aligned}$$

This intersection interval is sharp and **can be empty**.

Proof. Identified intervals for each pair $\overline{z}_m \in \mathcal{Z}_m$, are shown in **Theorem 5.1**. The bound in this corollary is found by taking intersection of all such identified intervals. This intersection bound is sharp. The same sharpness proof of **Theorem 5.2** applies with some modification in (S2) constructed in the proof in Appendix. When there exist many instrumental values that satisfy the rank condition, Restriction RC, the partition, $\{P^l\}_{l=1}^M$ defined in the proof of **Theorem 5.2** can be re-defined as the following :

$$\begin{aligned}
P^l &= \begin{Bmatrix} \min_{z \in SUPP(Z)}\{P^l(z)\}, & if & l < m-1 \\ \max_{z \in SUPP(Z)}\{P^l(z)\}, & if & l > m \end{Bmatrix} \\
P^{m-1} &= \min_{z \in \overline{z}_L}\{P^m(z)\} \\
P^m &= \max_{z \in \overline{z}_U}\{P^m(z)\}, \\
\text{where } \overline{z}_L &\equiv \{z_L : z_L = \min \overline{z}_m, \overline{z}_m \in \mathcal{Z}_m\} \\
\overline{z}_U &\equiv \{z_U : z_U = \max \overline{z}_m, \overline{z}_m \in \mathcal{Z}_m\} \\
\mathcal{Z}_m &\equiv \{\overline{z}_m : P^m(z'_m) \leq \tau_V \leq P^m(z''_m), \text{with } \overline{z}_m = \{z'_m, z''_m\}\}.
\end{aligned}$$

$\overline{z}_L(\overline{z}_U)$ is the set of smaller (larger) values of $\overline{z}_m = \{z'_m, z''_m\} \in \mathcal{Z}_m$. The partition of the support of V, $(0,1)$, is constructed such that $P^1 < P^2 < ... < P^M$. ■

Intersection of identified sets may be empty, and even if it is not empty, the causal interpretation of the intersection bound needs to be given to a different subpopulation.

Suppose that the intersection, $\mathbb{V} \neq \varnothing$. Then the bound defined by **Corollary 5.1** should be interpreted as causal effects for the subpopulation defined by $\mathbb{V}$. If $\mathbb{V} = \varnothing$, no causal interpretation would be possible, even though the intersection bound is not empty since the subpopulation that is affected by the change in the values of Z does not exist. If $\mathbb{V} \neq \varnothing$, but the intersection bound is empty, then this means that some of the restrictions

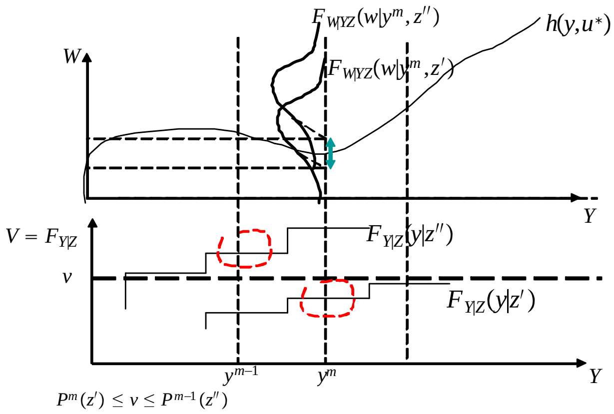


Figure 5.2: Chesher (2005) strong rank condition is that there exist values of Z, z'_m and z''_m such that $P^m(z'_m) \leq v \leq P^{m-1}(z''_m)$: the arrow in the upper panel indicates the Chesher bound. Note that if Chesher (2005)'s strong rank condition holds our rank condition always holds since $P^m(z') \leq v \leq P^{m-1}(z'') \leq P^m(z'')$. Note also that for this rank condition to hold IV should be very strong - Chesher (2005) demonstrate that Angrist and Krueger (1999) quarter of birth IV does not satisfy his rank condition.

in the model are not satisfied[17]. However, which restrictions are misspecified is not known by the fact that the identified set is empty. This way one can falsify the econometric model, rather than a specific restriction. This is one of the examples of the discussion of Chapter 2 can be applied. Another example is discussed in the next section.

5.3.4 Testability of Restriction LDRM

The identifying power of a model comes from the restrictions imposed by the model and the applicability of identification results depends on the credibility of the restrictions imposed. If we could test the restrictions using data, credibility of restrictions can be confirmed. As Koopmans and Reiersol (1950) noted the general rule of testability is that if there exists an observationally more restrictive model than the other such that both models identify the same structural feature, then the restrictions imposed by the observationally more restrictive model can be tested.

Consider Manski (1990), Manski (1997)'s Monotone Treatment Response (MTR) model, and Manski and Pepper (2000)'s Monotone Treatment Response and Monotone Treatment Selection (MTR-MTS) model. Since the models are nested, if the true data generating structure satisfies MTR and MTS, then the identified set by MTR-MTS should be included by the identified set by MTR. Another example is the case with Chesher (2005) model and $\mathcal{M}^{LDRM}$. If the strong rank condition is satisfied, $\mathcal{M}^{LDRM}$ is contained by Chesher (2005) model, thus, $\mathcal{M}^{LDRM}$ is observationally more restrictive. **Theorem 2.1 in Chapter 2** implies that LDRM bound should be equal to or smaller than Chesher (2005) bound.

LDRM restriction is "not directly testable[18]", in other words, LDRM restriction does

[17] I am grateful to Pierre Debois, and Brendon McConell for this point.

[18] Note that LDRM is a restriction imposed on the structural relation and the distribution of the unob-

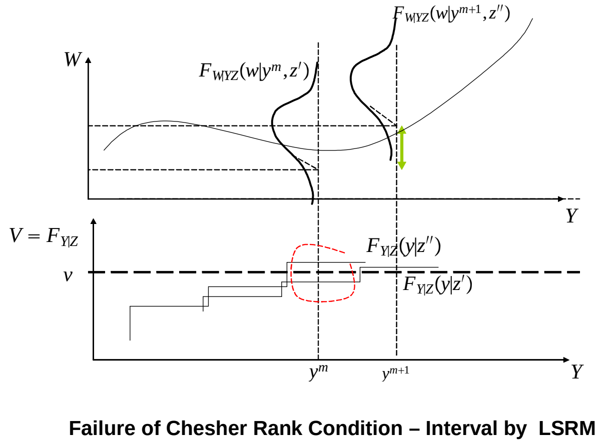


Figure 5.3: Failure of Chesher (2005) strong rank condition : when our rank condition holds we can define the sharp interval by the quantiles of the two distributions $F_{W|YX}(w|y^m, z')$ and $F_{W|YX}(w|y^{m+1}, z'')$ (not $F_{W|YX}(w|y^m, z'')$ as in Chesher (2005)). The arrow indicates the LDRM bound. The graph is drawn for the case with the nonnegative response case. Note that unless Chesher (2005) rank condition holds we are not sure whether u - quantile of $F_{W|YX}(w|y^m, z'')$ is below or above $h(y^m, u^*)$. This is why we cannot define the identified interval by the quantiles of $F_{W|YX}(w|y^m, z'')$ if Chesher (2005)'s rank condition is not satisfied. If Chesher (2005)'s rank condition holds then Chesher (2005) bounds should be equal to or larger than the LDRM bounds. See <Figure 5.4>.

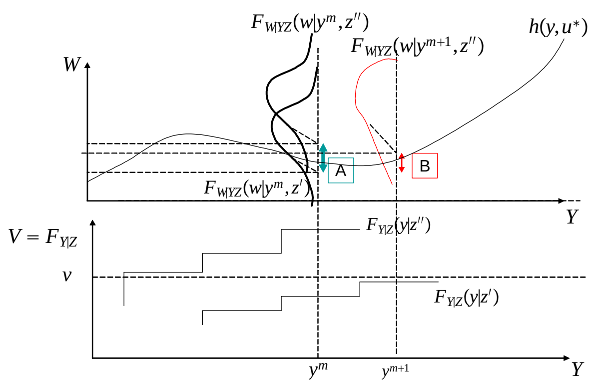


Figure 5.4: Testability of LDRM : when Chesher (2005) rank condition is satisfied Chesher bound(A) should be larger than or equal to LDRM bound(B) - if not, Restriction LDRM is not satisfied by the true structure.

not have any implication on the distribution of the observables, but it can be falsified when the strong rank condition in Chesher (2005) is satisfied. The strong rank condition is "directly testable[19]", thus, once the strong rank condition is satisfied we can say that the model, $\mathcal{M}^{LDRM}$ is observationally more restrictive than the model in Chesher (2005). In this case, the identified interval by $\mathcal{M}^{LDRM}$ should be included by the identified interval by Chesher (2005) if restriction LDRM is satisfied. Therefore, if the bounds constructed by Chesher (2005) are smaller than the bounds formed by the LDRM model, then this implies that restriction LDRM is not the right description of the true underlying structure that generated the data.

We cannot "confirm" restriction LDRM, but we can "refute" the restriction by comparing $Q_{W|YZ}(\tau_U|y^m, z'')$ with $Q_{W|YZ}(\tau_U|y^{m+1}, z'')$.

5.4 Binary Endogenous Variable

Although in many empirical studies, the distribution of the treatment effects can deliver valuable information for any policy design, quantiles of the distribution of differences of potential outcomes, $W_1 - W_0$, have been considered to be difficult to point identify without strong assumptions.[20] In this section I apply the LDRM model to a binary endogenous variable and identify the ceteris paribus impact of the binary variable, or treatment effects. As Chesher (2005) noted, models for an ordered discrete endogenous variable can not directly be applied to binary endogenous variables due to the "unordered" nature of binary variables, however, Restriction LDRM imposes a sense of order to a binary endogenous variable, which enables the model to identify the partial differences. The number of points in the support of Y restricts the identification result.

5.4.1 Bound on the Value of the Structural Relation

The model interval identifies $h(1, u^*)$ and $h(0, u^*)$ as the following corollary.

Corollary 5.2 Under Restriction A-EX,C-QI,RC,and LDRM there are the inequalities for $y \in \{0, 1\}$, $z \in \overline{z} = \{z', z''\}$, and $\tau \equiv \{\tau_U, \tau_V\}$,

$$\begin{aligned}
q^L(\tau, y, \overline{z}) &\leq h(y, u^*) \leq q^U(\tau, y, \overline{z}) \\
\text{where } u^* &= Q_{U|VZ}(\tau_U|\tau_V, z), \\
\text{for some } \tau_U &\in (0, 1) \text{ and } \tau_V \in [P(z'), P(z'')], \\
q^L(\tau, y, \overline{z}) &= \min\{Q_{W|YZ}(\tau_U|0, z'), Q_{W|YZ}(\tau_U|1, z'')\} \\
q^U(\tau, y, \overline{z}) &= \max\{Q_{W|YZ}(\tau_U|0, z'), Q_{W|YZ}(\tau_U|1, z'')\}
\end{aligned}$$

servables. The restrictions imposed on the structure are not testable unless they have implications on the distribution of the observables.

[19] Data are informative about whether the rank condition is satisfied since the rank condition is about the conditional distribution of Y given Z.

[20] Note that in general, quantiles of treatment effects, $Q_{W_1 - W_0|X}(\tau|x) \neq Q_{W_1|X}(\tau|x) - Q_{W_0|X}(\tau|x)$, where the right hand side is the QTE.

The bound is sharp.

Proof. See the Appendix ■

The identified intervals for $h(1,u^*)$ and $h(0,u^*)$ are the same. Nevertheless, this is still informative in the sense that the identified interval restricts the possible range that the values $h(1,u^*)$ and $h(0,u^*)$ lie in, and that under Restriction LDRM either the upper bound or the lower bound on $h(1,u^*)- h(0,u^*)$ should be zero.

Lemma 5.3 Under Restriction A-EX,C-QI,RC,and LDRM,

$$\begin{aligned} PDPR \text{ implies } Q_{W|YZ}(\tau_U|y^{m+1},z''_m) &\geq Q_{W|YZ}(\tau_U|y^m,z'_m), \text{ and} \\ NDNR \text{ implies } Q_{W|YZ}(\tau_U|y^{m+1},z''_m) &\leq Q_{W|YZ}(\tau_U|y^m,z'_m). \end{aligned}$$

Proof. See the Appendix. ■

Corollary 5.2 and **Lemma 5.3** are used to recover heterogeneous treatment responses. **Theorem 5.3** states the partial identification result of heterogeneous treatment effects.

5.4.2 Bound on Partial Difference of the Structural Relation

Theorem 5.3 Under Restriction A-EX,C-QI,RC,and LDRM, $h(1,u^*) - h(0,u^*)$ is identified by the following interval:

$$\begin{aligned} B^L &\leq h(1,u^*) - h(0,u^*) \leq B^U \\ B^U &= \max\{0, Q_{10}^{\Delta}(\tau_U)\} \\ B^L &= \min\{0, Q_{10}^{\Delta}(\tau_U)\}, \\ \text{where } Q_{10}^{\Delta}(\tau_U) &= Q_{W|YZ}(\tau_U|1,z'') - Q_{W|YZ}(\tau_U|0,z') \end{aligned}$$

Proof. Suppose $Q_{W|YZ}(\tau_U|1,z'') \geq Q_{W|YZ}(\tau_U|0,z')$. From **Corollary 5.2** we have

$$\begin{aligned} Q_{W|YZ}(\tau_U|0,z') &\leq h(1,u^*) \leq Q_{W|YZ}(\tau_U|1,z'') \\ Q_{W|YZ}(\tau_U|0,z') &\leq h(0,u^*) \leq Q_{W|YZ}(\tau_U|1,z'') \end{aligned}$$

then we have

$$\begin{aligned} -(Q_{W|YZ}(\tau_U|1,z'') - Q_{W|YZ}(\tau_U|0,z')) &\leq h(1,u^*) - h(0,u^*) \\ &\leq Q_{W|YZ}(\tau_U|1,z'') - Q_{W|YZ}(\tau_U|0,z'). \end{aligned} \tag{3}$$

By **Lemma 5.3**, if $Q_{W|YZ}(\tau_U|1,z'') \geq Q_{W|YZ}(\tau_U|0,z')$, we should have

$$h(1,u^*) - h(0,u^*) \geq 0$$

applying this to (3) yields the result. The case when $Q_{W|YZ}(\tau_U|1,z'') \leq Q_{W|YZ}(\tau_U|0,z')$ can be shown similarly. ■

Whether the treatment effect is positive or negative can be determined by data from the sign of $Q_{10}^{\Delta}(\tau_U) \equiv Q_{W|YZ}(\tau_U|1,z'') - Q_{W|YZ}(\tau_U|0,z')$ based on **Theorem 5.3**. If $Q_{10}^{\Delta}(\tau_U) > 0$, then

$$0 \leq h(1,u^*) - h(0,u^*) \leq Q_{10}^{\Delta}(\tau_U),$$

and if $Q_{10}^{\Delta}(\tau_U) < 0$, then

$$Q_{10}^{\Delta}(\tau_U) \leq h(1,u^*) - h(0,u^*) \leq 0.$$

If $Q_{10}^{\Delta}(\tau_U) = 0$, then $h(1,u^*) - h(0,u^*)$ is point identified as zero. Either the upper bound or the lower bound is always zero.

If Restriction LDRM were true about the underlying structure, then from this restriction we could infer whether the dependence between the two unobservables is positive or negative locally in a certain range by **Lemma 5.3**. If economic arguments can justify the nature of the selection pattern found from data, then this model can be credibly applicable.

5.4.3 Discussion

Heterogeneous Causality Measured by Partial Difference

The major object of interest in this paper is the partial difference of the structural quantile function, $h(1,u^*) - h(0,u^*)$. The value u^* is unknown, but is assumed to be $u^* = Q_{U|VZ}(\tau_U|\tau_V,z)$ for some $\tau_U,\tau_V \in (0,1)$. $h(1,u^*) - h(0,u^*)$ is interpreted as a ceteris paribus impact of Y. When the value of Y changes from 1 to 0, the value of U would change as well if there exists endogeneity.

This is in contrast with other identification results in additively nonseparable models. Other studies identify the values of a *nonseparable* structural function, but their results do not guarantee identification of partial differences. For example, Imbens and Newey (2009)'s control function method does not identify partial difference when the endogenous variable is discrete.

Rank Condition and Causal Interpretation

The rank condition restricts the group for whom the identification of causal impacts is justifiable into those who are ranked between $P(z')$ and $P(z'')$, where $P(z) = \Pr(Y = 0|Z = z)$. $h(1,u^*)-h(0,u^*)$ would be understood as the treatment effects of the τ_U-ranked individuals in the subpopulation whose $V-$ ranking is between $P(z')$ and $P(z'')$. When the value of Z changes from z' to z'', their treatment status changes from $y = 1$ to $y = 0$. We call this group "compliers" following the potential outcomes framework.

Applicability to Regression Discontinuity Designs (RDD)

Recently, many studies (see Lee and Lemieux (2009), for a survey) adopted regression discontinuity design (RDD) to measure causal effects. Under this design if the continuity condition at the threshold point of the "forcing variable" holds, the causal effects of individuals with the forcing variable just above and below the threshold point are shown to be identified.

When the RDD is available, our rank condition[21] is guaranteed to hold, thus, as long as Restriction LDRM is applicable in the context of interest, the proposed model can be applicable to an RD design even when all other variables are not continuous in the treatment - determining variable at the threshold.[22]

5.5 Further Comments

5.5.1 Control Function Methods, and Discrete Endogenous Variables in *Non-additive* Structural Relations

Control function approaches are usually understood as a way to correct endogeneity or selection problem by conditioning on the residuals obtained from the reduced form equations for the endogenous variables in a triangular simultaneous equations system. Control function methods (see Blundell and Powell (2003) for survey) are not considered to be applicable when the structural function is *non-additive* and the endogenous variable is *discrete*. If the structural relation is additively separable, the control function method can be applied to a case with a discrete endogenous variable. (See Heckman and Robb (1986)).

Imbens and Newey (2009)'s control function method is conditioning on the conditional distribution of the endogenous variable given other covariates as an extra regressor for the outcome equation. Chesher (2003) used the QCFA. This uses the same information as Imbens and Newey (2009), but, instead of conditioning on the conditional distributions of the endogenous variable given other covariates, the QCFA can be understood as conditioning on a quantile of the conditional distribution. Imbens and Newey (2009) show that the two control function approaches are equivalent when the endogenous variable is continuous.

When the endogenous variable is discrete, Imbens and Newey (2009)'s approach does not have identifying power.[23] Chesher (2003)'s QCFA fails to produce point identification since the one-to-one mapping between the endogenous variable and the unobserved variable that exists with a continuous endogenous variable does not exist any more with discrete endogenous variable. Rather, with a discrete endogenous variable, a specific value of the

[21] Suppose a threshold point t_0 of a variable T is known by a policy design such that the treatment status (Y) is partly determined by this vairiable. Then we can construct a binary variable Z such that $Z = 1(T > t_0)$. In such a case, our rank condition holds.

[22] For example, age or date of birth, which are used for eligibility criteria, are often only available at a monthly, quarterly, or annual frequency level.

[23] Imbens and Newey (2009) defines a bound, but this is for the case in which the common support assumption fails, not for a discrete endogenous variable.

endogenous variable maps into a set of values of the unobservable variable, called a **V-set**, thus, the QCFA with a discrete variable could be roughly described as conditioning on v- quantiles of the conditional distribution of the endogenous variable given covariates for $v \in$**V-set**. The smaller the V-set is, the smaller would the identified set be. Without imposing further restrictions, a sharp bound cannot be defined. Chesher (2005) suggested to impose monotonicity of $F_{U|V}(u|v)$ in v to define a bound on the value of the structural function. This monotonicity restriction is adopted in this chapter and Jun, Pinkse, and Xu (2010).

5.5.2 Nonparametric *Shape* Restrictions

Identifying power of an econometric model comes from restrictions imposed by the model. The restrictions can be categorized into two : those imposed on the structure, and those on data. One could impose restrictions on data - existence of a variable exhibiting certain characteristics such as large support condition, rank conditions, or completeness conditions.

Alternatively, one could adopt restrictions on the structure. Apart from Chesher (2005) and Jun, Pinkse,a dn Xu (2010), Manski and Pepper (2000)[24] and Bhattacharya, Shaikh and Vytlacil (2008) adopt monotonicity restrictions in the structural relations. Under the MTS (Monotone Treatment Selection) - MTR (Monotone Treatment Response) restriction Manski and Pepper (2000) estimated the upper bounds on the returns to schooling. With monotonicity in response, the lower bound is always zero.

Manski and Pepper (2000) develop their arguments by assuming that both selection and response are increasing, but assuming that both are decreasing also leads to identification of average effects. However, with the LDRM restriction, weakly increasing response should be matched with weakly increasing selection and vice versa. MTR is equivalent to monotone response assumption in our model, and MTS holds if $F_{U|V}(u|v)$ is weakly decreasing in v over the whole support of U. Since LDRM allows the direction (either PDPR or NDNR) of the match to vary over the support of U, while MTR-MTS should be matched - either positive response with positive selection or negative response with negative selection - for the mean. Roughly speaking, the LDRM restriction can be described as a *local* version of MTR-MTS. Manski and Pepper (2000) identifies average treatment effects, thus the heterogeneity in treatment effects can be found for the subpopulation defined by the *observed* characteristics, while LDRM model can recover heterogeneity in treatment effects *even* among observationally identical individuals.

Bhattacharya, Shaikh and Vytlacil (2008) compare Shaikh and Vytlacil (2005) bounds with Manski and Pepper (2000)[25] by applying them to a binary outcome - binary endogenous variable case. Bhattacharya, Shaikh and Vytlacil (2008)'s bounds are found under

[24] Okumura and Usui (2009) impose concavity to Manski and Pepper (2000) framework and show that the identified interval can be shortened. However, when the endogenous variable is binary Okumura and Usui (2009) bounds would be the same as those of Manski and Pepper (2000).

[25] In fact, what they consider is MTR-MIV in Manski and Pepper (2000) with the upper bound of the outcome 1 and the lower bound 0 when the outcome is binary.

the restriction that the binary endogenous variable is determined by an IV monotonically. When IV, Z, and Y are binary, their monotonicity is equivalent to the monotonicity here. Note also that when Y is binary, we can always reorder 0 and 1 due to the "unordered nature" of a binary variable. In contrast with their claim, when Manski and Pepper (2000) is applied to a binary case, the direction of the monotonicity of response and selection does not have to be determined a priori[26]. Data will inform about the direction of the monotonicity, however, the direction of MTR and MTS should be matched in a certain way[27].

The advantage of the LDRM assumption is that it allows the match to vary across the level of the unobserved characteristic unlike MTS-MTR in Manski and Pepper (2000) or Bhattacharya, Shaikh and Vytlacil (2008). The LDRM model would be useful when the direction of the dependence is likely to be different depending on the level of the unobserved characteristic. However, LDRM may not be very informative when the outcome is binary in practice, since the values that the partial difference can take are -1,0, and 1, although it is still legitimate to apply the model to binary outcomes in principle.

5.5.3 Different Approaches to Heterogeneous Treatment Responses

I discuss three different approaches to heterogeneous treatment effects. The information delivered by partial difference need to be distinguished from that by QTE or quantiles of treatment effect. The three approaches answer different policy questions.

Quantile Treatment Effect (QTE) QTE is defined as the horizontal difference between the marginal distributions of the potential outcomes. QTE can be used to investigate the impacts of any policy on, for example, median individuals in the distributions with and without a policy, which can be informative, for example, in the study of changes in inequality. However, QTE should not be interpreted as individual level causal effects because the ranks of individuals may vary across the treatment status. That is, the median ranked individuals in each potential outcome distribution may not be the same individuals. Moreover, even if the rank is preserved across the treatment status, the size of QTE would not necessarily be the same as the quantiles of the treatment effects since

[26] When the endogenous variable is oredered discrete with more than two points in the support, the direction should be assumed a priori to find the bounds.

[27] Following the notation of Manski and Pepper (2000) if data show that $E(y|z=0) \leq E(y|z=1)$, then this is the case where non-decreasing MTR and non-decreasing MTS are matched because

$$\begin{aligned} E(y|z &= 0) = E(y(0)|z=0) \overset{MTR}{\leq} E(y(1)|z=0) \\ \overset{MTS}{\leq} E(y(1)|z &= 1) = E(y|z=1). \end{aligned}$$

Whereas if the data show that $E(y|z=0) \leq E(y|z=1)$, then this is the case where non-increasing MTR matched with non-increasing MTS as follows :

$$\begin{aligned} E(y|z &= 0) = E(y(0)|z=0) \overset{MTR}{\geq} E(y(1)|z=0) \\ \overset{MTS}{\geq} E(y(1)|z &= 1) = E(y|z=1). \end{aligned}$$

The counterfactural $E(y(1)|z=0)$ can be bounded by $E(y|z=0)$ and $E(y|z=1)$, and the data will inform us of which is the upper/lower bound - the direction of the match will be determined by data.

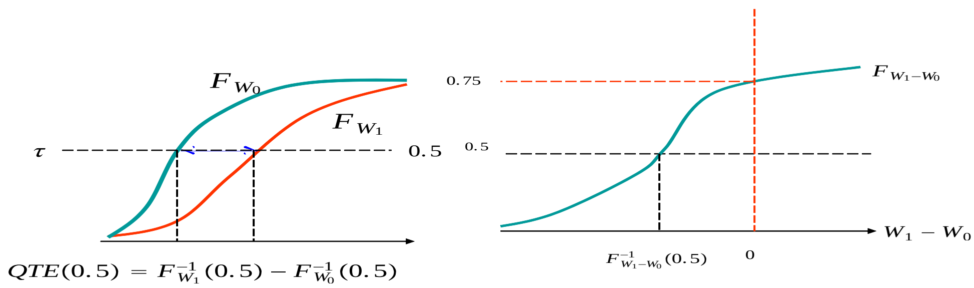


Figure 5.5: The QTE (in the left panel) measures the effects of a treatment on a particular point of distributions of marginal distributions of counterfactuals. It does not necessarily convey the information regarding the causal effects on the individuals unless the rank preservation condition holds. The information on $F_{W_1-W_0}$ (in the right panel) can be useful in finding out the proportion of the population that benefit from the treatment. For example, if the 0.75 quantile of $F_{W_1-W_0}$ is zero, then this means that 25% of the population benefit from the treatment.

$Q_{W_1-W_0} \neq Q_{W_1} - Q_{W_0}$.

Quantiles of treatment effects recovered from the distribution of the treatment effect, $F_{W_1-W_0}$ Another line of studies focuses on the distribution of the treatment effects, $F_{W_1-W_0}$. Their object of identification is $F_{W_1-W_0}$, and the identification results are found by the mathematical bounds such as Hoeffding bounds, or Makarov bounds. These bounds can be found from the marginal distributions of the potential outcomes. Identifying the marginal distributions of the potential outcomes is not simple - Heckman, Smith, and Clements (1997) assumed that the potential outcomes are normally distributed, and Fan and Park (2009) assume that experimental data are available. The studies mentioned above report partial identification of the distribution of the treatment effects. Once $F_{W_1-W_0}$ is found, then functionals of $F_{W_1-W_0}$, such as the quantiles of the treatment effects can be found following the definition of the quantiles.

Heterogeneous treatment responses recovered from partial differences Individual-specific heterogeneous treatment effect, $W_{1i} - W_{0i}$, defined by the potential outcomes framework, can be measured by partial difference under the structural framework, as $W_{1i} - W_{0i} = h(1, x, u) - h(0, x, u)$ for the individual i whose the observed and unobserved characteristics are $X_i = x$ and $U_i = u$.

Comparison of the three In general, QTE and partial differences should be different even with the rank preservation assumption. $h(1, x, u) - h(0, x, u)$ are not the same as the quantiles of $F_{W_1-W_0}$. This is because the quantile parameter(τ) used in our structural framework indicates the ranking of the outcome, W, (which is the same as that of the unobserved heterogeneity (U) under the monotonicity in scalar unobservable variable), while the quantile parameter for $F_{W_1-W_0}$ is the ranking of the treatment effects, W_1-W_0. They may be different.

The knowledge of $F_{W_1-W_0}$, and thus the knowledge of $Q_{W_1-W_0}$ can answer the questions of proportion of the population that benefit from the treatment. Our identification results can answer the question of "who benefits" from the treatment by identifying "who" using the observed characteristics and the ranking of the unobserved heterogeneity. Our results can then recover the proportion of population whose treatment effects are positive.

5.5.4 Inference

The inference results under set identification can be categorized into two[28] : the one is by Horowitz and Manski (2000) or Imbens and Manski (2004), Stoye (2008) and the other is by Chernozhukov, Hong and Tamer (2007), and many others recently. The first line of study estimates the bounds which are explicitly defined by the identification results and deal with the construction of the confidence intervals of the bounds. In the second line of study the identified set is not necessarily defined explicitly, rather they are defined by the (conditional) moment inequality conditions implicitly, and the inference methods are based on the moment inequality conditions. Our identification results do not provide any moment conditions to be adopted, thus, the first line of studies is more relevant to the paper.

The confidence intervals of the bounds with ordered discrete endogenous variables can be found by Imbens and Manski (2004) if there is only one pair of instrumental values. When there are more than two instrumental values, the bounds are found by intersecting the intervals found by each pair. In this case the bounds and the confidence intervals can be found by using Chernozhukov, Lee, and Rosen (2009). Either parametric (see Koenker and Basset (1978) for example) or nonparametric (see Chaudhuri (1991) or Chaudhuri, Doksum, Samarov (1997)) estimation of the quantiles can be applied for the construction of the confidence intervals.

When the endogenous variable is binary, the inference problem on partial difference is somewhat different. The inference problem from the identification results would be (i) estimating the upper bounds or lower bounds as the difference the two quantile functions, $Q_{W|YZ}(\tau_U|1,z'')-Q_{W|YZ}(\tau_U|0,z')$, (ii) testing whether the confidence interval of either the upper bound or the lower bound contains zero, and (iii) constructing the confidence intervals for the identified interval. The major inference issue would be testing whether zero is included in the confidence set of the upper/lower bounds as the model identifies the sign of the treatment effect.

[28] I mention this categorization as it is more relevant to the inference problem in this paper. However, this is not the only possible categorazation ; one can categoraze the inference approaches by whether the confidence set covers a specific point of the parameter of interest, or the identified set itself.

5.6 Empirical Illustration - Heterogeneous Individual Treatment Responses

By heterogeneous treatment responses I mean idiosyncratic treatment effects even after accounting for observed characteristics[29]. Several studies[30] allowed for individual heterogeneity in response. However, identification is achieved by integrating out the heterogeneity[31] in these studies. By identifying average responses, much information regarding the distributional consequences of a policy - heterogeneity in response - would be lost. In this paper individual heterogeneity in response is allowed by use of a non-additive structural relation and the proposed model identifies heterogeneity by identifying partial difference of the structural relation. I demonstrate how "partial" information (the signs and the bounds of treatment effects, not the exact size of them) regarding who benefits (individual heterogeneous response) can be recovered from data by using quantiles rather than averages when "who" is indicated by individual observed characteristics and the ranking in the distribution of the unobserved characteristic[32]. This is illustrated by examining the effects of the Vietnam-era veteran status on the civilian earnings using the data used in Abadie (2002)[33] - a sample of 11,637 white men, born in 1950-1953, from the March Current Population Surveys of 1979 and 1981-1985. Annual labour earnings are used as an outcome, and the veteran status is the binary endogenous variable of concern.

Veterans have been provided with various forms of benefits in terms of insurance, education, etc. How serious the impact of military service on veterans' labour market outcomes, or whether they are compensated for their service enough has been an important political issue and there has not been any consensus on this matter. Angrist (1990) reports negative average impact of veteran status on earnings later in life, which shows that on average military service had a negative impact on earnings possibly due to the loss of labour market experience.

5.6.1 Bounds on Individual-specific Causal Effects of Vietnam-era Veteran Status on Earnings

By applying his identification results of the marginal distribution of the potential outcomes for compliers, Abadie (2002) reports that military service during the Vietnam era reduces

[29] This is called "essential heterogeneity" by Heckman, Urzua, and Vytlacil (2006).

[30] The standard linear IV model cannot identify heterogeneus treatment effects. See Heckman and Navarro (2004) and Heckman and Urzua (2009).

For identification under heterogeneous responses see Heckman, Urzua, and Vytlacil (2006) for binary endogenous variable, and Florens, Heckman, Meghir, and Vytlacil (2008), Athey and Imbens (2006), Imbens and Newey (2009),Chernozhukov, Fernandez-Val, and Newey (2009), Hoderlein and White (2009), among others. There is another line of research using random coefficient models to recover the distribution of the response, see Card (2001) and Heckman and Vytlacil (1999) for example.

[31] The averaged objects however can exhibit a certain degree of heterogeneity by allowing for treatment heterogeneity.

[32] Most welfare programs are designed to support certain groups of people. If "who benefits" from such programs could be recovered from data, this would be informative in judging whether the groups targeted by the policy actually benefit from it.

[33] The data are obtainable in Angrist Data Archive :
http://econ-www.mit.edu/faculty/angrist/data1/data

lower quantiles of the earnings distribution, leaving higher quantiles unaffected. The information from the marginal distribution of the potential outcomes (for compliers) may be used to recover QTE, however, it does not reveal any information on individual-specific impact on earnings of Vietnam-era veteran experience.

Let W be annual labour earnings, Y be the veteran status, and Z be the binary variable determined by the draft lottery. Age, race, and gender are controlled so that the subgroup considered is observationally homogenous. The unobserved variables U and V indicate scalar indexes for "earnings potential" and "participation preference" or "aptitude for the army" each. Note that there can be many factors that determine these indexes, but we assume that these multi-dimensional elements can be collapsed into a "scalar" index.

Selection on Unobservables

Enrollment for military service during the Vietnam era may have been determined by the factors which may have been associated with the unobserved earnings potential. This concern about selection on unobservables is caused by several aspects of decision processes both of the military and of those cohorts to be drafted. On the one hand, the military enlistment process selects soldiers on the basis of factors related to earnings potential. For example, the military prefers high school graduates and screens out those with low test scores, or poor health. As a consequence, men with very low earnings potential are unlikely to end up in the army. On the other hand, for some volunteers military service could be a better option because they expected that their careers in the civilian labour market would not be successful, while others with higher earnings potential probably found it worthwhile to escaping the draft. This shows that the direction of selection could vary with where each individual is located in the distribution of the earnings potential.

Draft Lottery as an Instrument - Exclusion, Rank Condition, and Independence

As in Angrist (1990) the Vietnam era draft lottery is used as an instrument to identify the effects of veteran status on earnings. The lottery was conducted every year between 1970 and 1974. The lottery assigned numbers from 1 to 365 to dates of birth in the cohorts being drafted. Men with the lowest numbers were called to serve up to a ceiling[34]. The ceiling was unknown in advance. I construct a binary IV based on the lottery number the threshold point being chosen as 100 following Abadie(2002).

It would be natural to believe that this IV is not a determinant of earnings, and the unobserved scalar indexes are independent of draft eligibility[35].

[34] The draft eligibility ceilings were 195 for men born in 1950, 125 for men born in 1951, and 95 for men born in 1952. The eligibility ceiling is determined by the Department of Defense depending on the needs in the year.

[35] There has been some discussion on individuals' draft lottery number caused behavior : some men therefore volunteered in the hope of serving under better terms and gaining some control over the timing of their service. If those who change their behavior according to their draft lottery number show certain patterns in their unobserved factors, then the quantile invariance restriction may be violated.

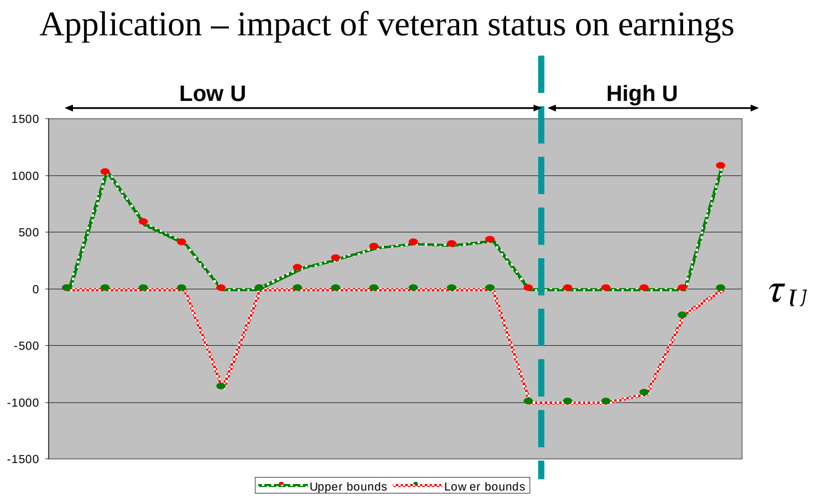


Figure 5.6: LDRM bounds on heterogeneous treatment effects of Vietnam era veteran status among the observationally similar individuals

To apply the identification results in **Theorem 5.3,** I investigate first whether the data satisfy Restriction RC in the model. Consider $X = age, gender, race$. The participation rate[36] among the draft-non-eligible ($Z = 0$) is about 0.14 and the participation rate among the eligible is 0.22.

$$P(Y = 0|Z = 1, X = x) = 0.78 < P(Y = 0|Z = 0, X = x) = 0.86 \qquad \text{(RC)}$$

Thus, $z' = 1$ and $z'' = 0$ in this example. The compliers (or draftees) are defined as those whose V-ranking is between 78% and 86%. Note that the V- ranking is never observed, so we cannot tell whether an individual is a complier or not.

The Result and Causal Interpretation

The bounds on the partial differences, $Q_{W|YZ}(\tau_U|1, z'') - Q_{W|YZ}(\tau_U|0, z')$, are found by the differences in the quantiles of earnings for the veterans who were not eligible and those of non-veterans who were draft-eligible.

LATE can be found by the model in Imbens and Angrist (1994). LATE is found for compliers by integrating out the heterogeneity, therefore, hiding possibly useful information regarding heterogeneity. While Angrist (1990) report negative impact on earnings on average, our quantile based analysis reveals that when age, gender, and race are controlled the veteran status had positive causal impacts for individuals with low earnings potential, but negative causal impacts for individuals with high earnings potential(see <Figure 5.6>).

The costs of military service may be larger than the benefits provided by the government for those with high earnings potential, while the benefits provided may be sufficient for those with low earnings potential. Considering the fact that benefits in the form of

[36] Note that $P(z)$ is not the usual propensity score, and $1 - P(z)$ is the propensity score.

insurance, pension, or education opportunities should be targeted at people with less potentials, the findings indicate that the compensation was enough for this group. However, the Vietnam-era military service may have higher opportunity costs for individuals with high earnings potential. This may be used against conscription.

The results in <Figure 5.6> are interpreted as the causal effects for those who change their participation decision as the value of Z changes. To the extent that we believe the implication from Restriction LDRM on the distribution of the unobservable the bounds would be considered to be informative regarding the population.

5.7 Conclusion

The presence of endogeneity and discreteness of the endogenous variable causes the loss of the identifying power of the quantile-based control function approach (QCFA) in the sense that the model based on the QCFA does not produce point identification. I propose a model that set identifies the structural features when one of the regressors is ordered discrete. I then apply the model to a binary endogenous variable, this structural approach turns out to be useful in defining the bounds on the heterogeneous individual treatment effects, which have not been studied so far under the structural framework without distributional assumptions.

The set identification result of this paper is applied to recover heterogeneous impacts of the Vietnam-era military service on earnings later in life. As we can see in this example, average effects may miss much information in some cases. Even though the proposed model can give only partial information on the individual causal effect, this may be useful in some economic contexts, especially when the sign of the effects may be varying across individuals with different characteristics. The causal interpretation is justified on the group of compliers defined by the pair of instrumental values that satisfy the rank condition. Different pair defines different "compliers". Heterogeneity in responses is recovered for different earnings potentials. If there exist heterogeneity in responses between draftees and volunteers, then our findings cannot be extrapolated into volunteers.

In conclusion, by using nonparametric shape restrictions that can be argued in each economic context, the proposed model provides partial information regarding individual causal effects. This information can be more credible than parametric restrictions to the extent they are justifiable by economic logic. The information on the signs of individual treatment effects is crucial if they vary across the population, since in such a case the average effect would be smaller with different effects with different signs canceled out. This would lead to a misleading conclusion. The model can also be used for robustness checks in data analysis for whether there exists any heterogeneity in causal responses.

Appendices

Appendix A

Appendix of Chapter 4

A.1 Theorem 4.1

Proof. Recall that we have

$$\begin{aligned} y_n &= h_n^Y(x, g_n(y_n, x)), && \text{(A)} \\ G_n(v_n, x) &\equiv Q_{Y_n|X}(v_n|x) = h_n^Y(x, Q_{V_n|X}(v_n|x)) && \text{(B)} \\ T_m &= h^*(y, x, q^m), && \text{(C-IC)} \\ n &\in \{1, 2, ..., N\}, \quad m \in \{1, 2, ..., M\} \end{aligned}$$

Differentiating (A) w.r.t. y_n and x_k, we have

$$\begin{aligned} 1 &= \nabla_v h_n^Y \cdot \nabla_{y_n} g_n, && \text{(A - 1)} \\ 0 &= \nabla_{x_k} h_n^Y + \nabla_v h_n^Y \cdot \nabla_{x_k} g_n, && \text{(A - 2)} \\ n &\in \{1, 2, ..., N\}, \ k \in \{1, 2, ..., K\}. \end{aligned}$$

Note also that h_n^Y is identified by quantiles of Y_n given X, (B). Then we have the following relations by differentiating (B) w.r.t. x_k :

$$\nabla_{x_k} G_n = \nabla_v h_n^Y \cdot \nabla_{x_k} g_n, \ k \in \{1, 2, ..., K\}. \qquad (*)$$

Recall that

$$f(P^m(y, x), g_1(y_1, x), g_2(y_2, x), ..., g_N(y_N, x), x) \equiv Q_{U|VX}(P^m(y, x)|g(y, x), x).$$

Then from eq. (C-IC) we have

$$\begin{aligned} T_m &= h^*(Y, X, q^m) && (**) \\ &= h^*(Y, X, f(P^m(y, x), g_1(y_1, x), g_2(y_2, x), ..., g_N(y_N, x), x)) \end{aligned}$$

First, differentiate (**) with respect to y_n, and x_k, we have[1]

$$
\begin{aligned}
0 &= \nabla_{y_n} h^* + \nabla_u h^* \cdot (\nabla_\tau f \cdot \nabla_{y_n} P^m + \nabla_{g_n} f \cdot \nabla_{y_n} g_n), && \text{(B - 1)}\\
0 &= \nabla_{x_k} h^* + \nabla_u h^* [\sum_{n=1}^{N} (\nabla_\tau f \cdot \nabla_{x_k} P^m + \nabla_{g_n} f \cdot \nabla_{x_k} g_n) + \nabla_{x_k} f], && \text{(B - 2)}\\
n &\in \{1, 2, ..., N\},\ k \in \{1, 2, ..., K\}.
\end{aligned}
$$

When we use the control function approach as in Chesher (2003), how U and $\{V_n\}_{n=1}^N$ move together is indicated by how the quantiles of U given V and X are affected by $V = g(Y, X)$. $\{\nabla_{g_n} f\}_{n=1}^N$ will indicate this information - if U and $\{V_n\}_{n=1}^N$ are independent, then $\nabla_{g_n} f = 0$. It turns out we can identify $\dfrac{\nabla_{g_n} f}{\nabla_{v_n} h_1}$. Now define matrices to state the results

Reproducing the definitions in Section 4.1 we have

$$
\boldsymbol{\lambda}_{y/u} \equiv \frac{1}{\nabla_u h^*} \begin{bmatrix} \nabla_{y_1} h^* \\ \vdots \\ \nabla_{y_N} h^* \end{bmatrix}_{N\times 1}, \boldsymbol{\lambda}_{x/u} \equiv \frac{1}{\nabla_u h^*} \begin{bmatrix} \nabla_{x_1} h^* \\ \vdots \\ \nabla_{x_K} h^* \end{bmatrix}_{K\times 1},
$$

$$
\mathbf{f}_x \equiv \begin{bmatrix} \nabla_{x_1} f \\ \vdots \\ \nabla_{x_K} f \end{bmatrix}_{K\times 1}, \mathbf{f}_g \equiv \begin{bmatrix} \nabla_{g_1} f \\ \vdots \\ \nabla_{g_N} f \end{bmatrix}_{N\times 1}, \boldsymbol{\gamma} \equiv \begin{bmatrix} \frac{\nabla_{g_1} f}{\nabla_{v_1} h_1} \\ \vdots \\ \frac{\nabla_{g_N} f}{\nabla_{v_N} h_N} \end{bmatrix}_{N\times 1}
$$

$$
\mathbf{g}_y \equiv \begin{bmatrix} \nabla_{y_1} g_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \nabla_{y_N} g_N \end{bmatrix}_{N\times N}, \mathbf{g}_x \equiv \begin{bmatrix} \nabla_{x_1} g_1 & \cdots & \nabla_{x_1} g_N \\ \vdots & \ddots & \vdots \\ \nabla_{x_K} g_1 & \cdots & \nabla_{x_K} g_N \end{bmatrix}_{K\times N},
$$

$$
\mathbf{h}_V \equiv \begin{bmatrix} \nabla_{v_1} h_1^Y & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \nabla_{v_N} h_N^Y \end{bmatrix}_{N\times N}, \mathbf{h}_x \equiv \begin{bmatrix} \nabla_{x_1} h_1^Y & \cdots & \nabla_{x_1} h_N^Y \\ \vdots & \ddots & \vdots \\ \nabla_{x_K} h_1^Y & \cdots & \nabla_{x_K} h_N^Y \end{bmatrix}_{K\times N},
$$

[1] With U continuously distributed, we have

$$\tau = F_{U|VX}(Q_{U|VX}(\tau|g(y,x)|g(y,x)).$$

Differentiating it w.r.t. τ, we have

$$\nabla_\tau Q_{U|VX}(\tau|g(y,x),x) = \frac{1}{f_{pdf}(Q_{U|VX}(\tau|g(y,x),x))},$$

where $f_{pdf}(u|v,x) = \nabla_u F_{U|VX}(u|v,x)$.

From $q^m_{U|VX} \equiv Q_{U|VX}(P^m(y,x)|g(y,x),x) \equiv f(P^m(y,x), g_1(y_1,x), g_2(y_2,x), ..., g_N(y_N,x), x)$, $\nabla_\tau f = \nabla_\tau Q_{U|VX}(\tau|g(y,x),x)$. Then we know from the property of quantiles described above,

$$\nabla_\tau Q_{U|VX}(\tau|g(y,x),x) = \frac{1}{f_{pdf}(Q_{U|VX}(\tau|g(y,x),x))},$$

where $f_{pdf}(u|v,x) = \nabla_u F_{U|VX}(u|v,x)$.

Thus, we can see that $\nabla_\tau f$ is unknown. However, with continuously distributed U, we know that $\nabla_\tau f \neq 0$.

$$\mathbf{F}_y^W \equiv \begin{bmatrix} \nabla_{y_1} P^m \\ \vdots \\ \nabla_{y_m} P^m \end{bmatrix}_{N\times 1}, \mathbf{F}_x^W \equiv \begin{bmatrix} \nabla_{x_1} P^m \\ \vdots \\ \nabla_{x_K} P^m \end{bmatrix}_{K\times 1}, \text{ and } \mathbf{G}_x \equiv \begin{bmatrix} \nabla_{x_1} G_1 & \cdots & \nabla_{x_1} G_N \\ \vdots & \ddots & \vdots \\ \nabla_{x_K} G_1 & \cdots & \nabla_{x_K} G_N \end{bmatrix}_{K\times N},$$

$$\mathbf{\Phi} \equiv \begin{bmatrix} \mathbf{I}_N & \mathbf{0}_K & \mathbf{I}_N & \mathbf{0}_K \\ \mathbf{0}_N & \mathbf{I}_K & -\mathbf{G}_x & \mathbf{I}_K \\ \mathbf{A}_y & \mathbf{A}_x & \mathbf{A}_\gamma & \mathbf{A}_f \end{bmatrix}_{(G+N+K)\times(2N+2K)}, \mathbf{\Psi} \equiv \begin{bmatrix} \boldsymbol{\lambda}_{y/u} \\ \boldsymbol{\lambda}_{x/u} \\ \boldsymbol{\gamma} \\ \mathbf{f}_x \end{bmatrix}_{(2N+2K)\times 1},$$

$$\boldsymbol{\phi} \equiv \begin{bmatrix} \mathbf{S}_y \\ \mathbf{S}_x \\ \mathbf{a} \end{bmatrix}_{(G+N+K)\times 1} \text{,where } S_y = \nabla_\tau f \cdot \mathbf{F}_y^W \text{ and } S_x = \nabla_\tau f \cdot \mathbf{F}_x^W.$$

Then (A - 1) and (A - 2) will be written as the following since $G_x = h_x$ by (*) :

$$\begin{aligned} \mathbf{I}_N &= \mathbf{g}_y \cdot \mathbf{h}_v \\ -\mathbf{G}_x &= \mathbf{g}_x \cdot \mathbf{h}_v. \end{aligned}$$

Thus, Replacing g_y and g_x by the following in (B - 1) and (B - 2)

$$\begin{aligned} \mathbf{g}_y &= \mathbf{I}_N \cdot \mathbf{h}_v^{-1} \\ \mathbf{g}_x &= -\mathbf{G}_x \cdot \mathbf{h}_v^{-1} \end{aligned}$$

we obtain :

$$\begin{aligned} \boldsymbol{\lambda}_{y/u} &= -\nabla_\tau f \cdot \mathbf{F}_y^W - \mathbf{I}_N \cdot \boldsymbol{\gamma} \\ \boldsymbol{\lambda}_{x/u} &= -\nabla_\tau f \cdot \mathbf{F}_x^W + \mathbf{G}_x \cdot \boldsymbol{\gamma} - \mathbf{f}_x \end{aligned}$$

Then by stacking these equations using the restrictions of Restriction R-IC, the rank condition follows from the fact that Ψ has 2N+2K elements. ■

A.2 Theorem 4.2

Proof. Recall that we have from Section 4.2

$$\begin{aligned} y_n &= h_n^Y(x, g_n(y_n, x)), && \text{(A)} \\ G_n(v_n, x) &\equiv Q_{Y_n|X}(v_n|x) = h_n^Y(x, Q_{V_n|X}(v_n|x)) && \text{(B)} \\ T_m &= h^*(y, x, q^m), && \text{(C-IC)} \\ n &\in \{1, 2, ..., N\}, \quad m \in \{1, 2, ..., M\} \end{aligned}$$

Now differentiate $F_{W|YX}$ and G_m w.r.t. $y_m, m = 1, 2, ..., M$ and $x_k, k = 1, 2, ..., K$ to get

$$\begin{aligned}\nabla_{y_m} F_{W|YX} &= \nabla_\theta s \cdot \nabla_{y_m}\theta + \nabla_\theta s \cdot \nabla_{g_m}\theta \cdot \nabla_{y_m} g_m + \nabla_{g_m} s \cdot \nabla_{y_m} g_m \ , \qquad \text{(A)}\\ \nabla_{x_k} F_{W|YX} &= \nabla_\theta s \cdot \nabla_{x_k}\theta \ + \nabla_\theta s \cdot \nabla_{v_m}\theta \cdot \nabla_{x_k} g_m \ \ + \nabla_u s \cdot \nabla_{x_k} g_1 \ + \nabla_{x_k} s \ .\end{aligned}$$

$$\nabla_{x_k} G_m = \nabla_{x_k} h_m.$$

Differentiating (A) w.r.t. y_m and x_k, will yield

$$\begin{aligned}1 &= \nabla_{v_m} h_m \cdot \nabla_{y_m} g_m \qquad \text{(B)}\\ 0 &= \nabla_{x_k} h_m + \nabla_{v_m} h_m \cdot \nabla_{x_k} g_m\end{aligned}$$

$m = 1, 2, ..., M$ and $k = 1, 2, ..., K$.Then from (A) and using the fact that h_m is identified the quantiles of the conditional distribution of Y_m given X, that is, $\nabla_{x_k} G_m = \nabla_{x_k} h_m$ for all $m = 1, 2, ..., M$ and $k = 1, 2, ..., K$.

$$\begin{aligned}F_y^W &= \lambda_y + g_y \cdot (\nabla_\theta s \cdot \theta_g + s_g)\\ F_x^W &= \lambda_x \ + g_y \cdot (\nabla_\theta s \cdot \theta_g + s_g) \ \ + s_x\end{aligned}$$

$$\mathbf{G}_x = \mathbf{h}_x$$

Reproducing the definitions in Section 4.2 we have

$$\boldsymbol{\lambda}_y \equiv \nabla_\theta s \begin{bmatrix} \nabla_{y_1}\theta \\ \vdots \\ \nabla_{y_M}\theta \end{bmatrix}_{N\times 1}, \boldsymbol{\lambda}_x \equiv \nabla_\theta s \begin{bmatrix} \nabla_{x_1}\theta \\ \vdots \\ \nabla_{x_K}\theta \end{bmatrix}_{K\times 1},$$

$$\mathbf{s}_x \equiv \begin{bmatrix} \nabla_{x_1} s \\ \vdots \\ \nabla_{x_K} s \end{bmatrix}_{K\times 1}, \mathbf{s}_g \equiv \begin{bmatrix} \nabla_{g_1} s \\ \vdots \\ \nabla_{g_N} s \end{bmatrix}_{N\times 1}, \boldsymbol{\theta}_g = \begin{bmatrix} \nabla_{g_1}\theta \\ \vdots \\ \nabla_{g_N}\theta \end{bmatrix}_{N\times 1},$$

and $\boldsymbol{\gamma} \equiv \nabla_\theta s \begin{bmatrix} \frac{\nabla_{g_1}\theta + \nabla_{g_1} s}{\nabla_{v_1} h_1} \\ \vdots \\ \frac{\nabla_{g_M}\theta + \nabla_{g_M} s}{\nabla_{v_M} h_M} \end{bmatrix}_{N\times 1} = \nabla_\theta s \theta_g + s_g,$

$$\mathbf{g}_y \equiv \begin{bmatrix} \nabla_{y_1} g_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \nabla_{y_N} g_N \end{bmatrix}_{N\times N}, \mathbf{g}_x \equiv \begin{bmatrix} \nabla_{x_1} g_1 & \cdots & \nabla_{x_1} g_N \\ \vdots & \ddots & \vdots \\ \nabla_{x_K} g_1 & \cdots & \nabla_{x_K} g_N \end{bmatrix}_{K\times N},$$

$$\mathbf{h}_V \equiv \begin{bmatrix} \nabla_{v_1} h_1^Y & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \nabla_{v_M} h_N^Y \end{bmatrix}_{N\times N}, \mathbf{h}_x \equiv \begin{bmatrix} \nabla_{x_1} h_1^Y & \cdots & \nabla_{x_1} h_N^Y \\ \vdots & \ddots & \vdots \\ \nabla_{x_K} h_1^Y & \cdots & \nabla_{x_K} h_N^Y \end{bmatrix}_{K\times N},$$

$$\mathbf{F}_y^W \equiv \begin{bmatrix} \nabla_{y_1} F_{W|YX} \\ \vdots \\ \nabla_{y_N} F_{W|YX} \end{bmatrix}_{N\times 1}, \mathbf{F}_x^W \equiv \begin{bmatrix} \nabla_{x_1} F_{W|YX} \\ \vdots \\ \nabla_{x_K} F_{W|YX} \end{bmatrix}_{K\times 1}, \mathbf{G}_x \equiv \begin{bmatrix} \nabla_{x_1} Q_{Y_1|X} & \cdots & \nabla_{x_1} Q_{Y_n|X} \\ \vdots & \ddots & \vdots \\ \nabla_{x_K} Q_{Y_1|X} & \cdots & \nabla_{x_K} Q_{Y_n|X} \end{bmatrix}_{K\times N}.$$

Now, from (B), we have

$$\begin{aligned} \mathbf{I}_N &= \mathbf{g}_y \cdot \mathbf{h}_V \\ -\mathbf{G}_x &= \mathbf{g}_x \cdot \mathbf{h}_V. \end{aligned}$$

Solving for g_y and g_x, then replacing them in (A)

$$\begin{aligned} \mathbf{g}_y &= \mathbf{I}_N \cdot \mathbf{h}_V^{-1} \\ \mathbf{g}_z &= -\mathbf{G}_x \cdot \mathbf{h}_V^{-1}. \end{aligned}$$

$$\begin{aligned} \mathbf{F}_y^W &= \boldsymbol{\lambda}_y + \mathbf{I}_M \cdot \mathbf{h}_V^{-1}(\nabla_\theta s \cdot \boldsymbol{\theta}_g + \mathbf{s}_g) \\ \mathbf{F}_x^W &= \boldsymbol{\lambda}_x \ - \mathbf{G}_x \cdot \mathbf{h}_V^{-1}(\nabla_\theta s \cdot \boldsymbol{\theta}_g + \mathbf{s}_g) \quad + \mathbf{s}_x \end{aligned}$$

we have

$$\begin{aligned} \mathbf{F}_y^W &= \boldsymbol{\lambda}_y + \mathbf{I}_N \cdot \boldsymbol{\gamma} \\ \mathbf{F}_x^W &= \boldsymbol{\lambda}_x \ - \mathbf{G}_x \cdot \boldsymbol{\gamma} \ + \mathbf{s}_x. \end{aligned}$$

Thus, the result follows. ■

A.3 Theorem 4.3

Proof. From $\lambda(y,x,v) = \sum_{m=0} m P_m^V(y,x,v)$, we have

$$\begin{aligned} \nabla_{y_n}\lambda(y,x,v) &= \sum_{m=1}^{M} m \cdot \nabla_{y_n} P_m^V(y,x,v) \\ n &= 1,2,...,N \end{aligned}$$

To identify $\nabla_{y_n}\lambda(y,x,v)$, $\nabla_{y_n}P_m^V(y,v)$ needs to be measured. Since how the endogenous variable and the unobservable heterogeneity are related is specified by auxiliary equations, $Y_n = h_n(X, V_n)$, $n = 1,2,...,N$, and by strict monotonicity of h_n in V_n, the inverse function exists, $V_n = h_n^{-1}(Y_n, X) \equiv g_n(Y_n, X)$. Thus, we have

$$P_m(y,x) = P_m^V(y,x,g(y,x)) \qquad (C-ODO)$$

Also recall from section 3.1 we have

$$\begin{aligned} y_n &= h_n(x, g_n(y,x)) && \text{(A)} \\ G_n(v_n, x) &\equiv Q_{Y|X}(v_n|x) = h_n^Y(x, Q_{V_n|X}(v_n|x)) && \text{(B)} \\ n &\in \{1,2,...,N\}. \end{aligned}$$

Differentiate (A) w.r.t. y_n and x_k respectively yields

$$\begin{aligned} 1 &= \quad + \nabla_v h_n^Y \cdot \nabla_{y_n} g_n \\ 0 &= \nabla_{x_k} h_n + \nabla_v h_n^Y \cdot \nabla_{x_k} g_n \\ n &\in \{1, 2, ..., N\},\ k \in \{1, 2, ..., K\}. \end{aligned}$$

Differentiating (B) w.r.t. x_k yields

$$\begin{aligned} \nabla_{x_k} G_n(v_n, x) &= \nabla_{x_k} h_n^Y + \nabla_v h_n^Y \cdot \nabla_{x_k} Q_{V_n|X}(v_n|x) \\ k &\in 1, 2, ..., K \end{aligned}$$

Differentiating $(C - ODO)$ w.r.t. y_n and x_k respectively yields

$$\begin{aligned} \nabla_{y_n} P_m(y, x) &= \nabla_{y_n} P_m^V(y, x, v) + \nabla_{v_n} P_m^V(y, x, v) \cdot \nabla_{y_n} g_n(y_n, x) \qquad (*) \\ \nabla_{x_k} P_m(y, x) &= \nabla_{x_k} P_m^V(y, x, v) + \nabla_{v_n} P_m^V(y, x, v) \cdot \nabla_{x_k} g_n(y_n, x) \\ n &\in \{1, 2, ..., N\},\ k \in \{1, 2, ..., K\} \end{aligned}$$

We need to find out conditions that $\nabla_{y_n} P_m^V(y, x, v), \nabla_{x_k} P_m^V(y, x, v), n \in \{1, 2, ..., N\},\ k \in \{1, 2, ..., K\}$ can be solved for. Using matrices $(*)$ is rewritten so that the rank condition similar to the classical simultaneous equations system.

Reproducing the definitions of vectors and arrays in Section 4.3 as follows

$$\boldsymbol{\xi}_y \equiv \begin{bmatrix} \nabla_{y_1} P_m^V \\ \vdots \\ \nabla_{y_N} P_m^V \end{bmatrix}_{N\times 1}, \boldsymbol{\xi}_x \equiv \begin{bmatrix} \nabla_{x_1} P_m^V \\ \vdots \\ \nabla_{x_K} P_m^V \end{bmatrix}_{K\times 1}, \boldsymbol{\eta} \equiv \begin{bmatrix} \nabla_{v_1} P_m^V \\ \vdots \\ \nabla_{v_N} P_m^V \end{bmatrix}_{N\times 1},$$

$$\mathbf{g}_y \equiv \begin{bmatrix} \nabla_{y_1} g_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \nabla_{y_N} g_N \end{bmatrix}_{N\times N}, \mathbf{g}_x \equiv \begin{bmatrix} \nabla_{x_1} g_1 & \cdots & \nabla_{x_1} g_N \\ \vdots & \ddots & \vdots \\ \nabla_{x_K} g_1 & \cdots & \nabla_{x_K} g_N \end{bmatrix}_{K\times N},$$

$$\mathbf{h}_V \equiv \begin{bmatrix} \nabla_{v_1} h_1^Y & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \nabla_{v_N} h_N^Y \end{bmatrix}_{N\times N}, \mathbf{h}_x \equiv \begin{bmatrix} \nabla_{x_1} h_1^Y & \cdots & \nabla_{x_1} h_N^Y \\ \vdots & \ddots & \vdots \\ \nabla_{x_K} h_1^Y & \cdots & \nabla_{x_K} h_N^Y \end{bmatrix}_{K\times N},$$

$$\mathbf{F}_y^W \equiv \begin{bmatrix} \nabla_{y_1} P_m \\ \vdots \\ \nabla_{y_m} P_m \end{bmatrix}_{N\times 1}, \mathbf{F}_x^W \equiv \begin{bmatrix} \nabla_{x_1} P_m \\ \vdots \\ \nabla_{x_K} P_m \end{bmatrix}_{K\times 1},$$

$$\text{and } \mathbf{G}_x \equiv \begin{bmatrix} \nabla_{x_1} G_1 & \cdots & \nabla_{x_1} G_N \\ \vdots & \ddots & \vdots \\ \nabla_{x_K} G_1 & \cdots & \nabla_{x_K} G_N \end{bmatrix}_{K\times N}.$$

Then we have from the identify

$$
\begin{aligned}
\mathbf{I}_N &= \mathbf{g}_y \cdot \mathbf{h}_v \\
-\mathbf{G}_x &= \mathbf{g}_x \cdot \mathbf{h}_v
\end{aligned}
$$

thus, yielding

$$
\begin{aligned}
\mathbf{g}_y &= \mathbf{I}_N \cdot \mathbf{h}_v^{-1} \\
\mathbf{g}_x &= -\mathbf{G}_x \cdot \mathbf{h}_v^{-1}
\end{aligned}
$$

By replacing $\mathbf{g}_y$ and $\mathbf{g}_x$ in (*) we have

$$
\begin{aligned}
\mathbf{F}_y^W &= \boldsymbol{\xi}_y + \mathbf{I}_N \cdot \mathbf{h}_v^{-1}\boldsymbol{\eta} \\
\mathbf{F}_x^W &= \boldsymbol{\xi}_x - \mathbf{G}_x \cdot \mathbf{h}_v^{-1}\boldsymbol{\eta}
\end{aligned}
$$

Therefore, for the arrays defined as the following

$$
\boldsymbol{\Phi} \equiv \begin{bmatrix} \mathbf{I}_N & \mathbf{0}_K & \mathbf{I}_N\mathbf{h}_v^{-1} \\ \mathbf{0}_N & \mathbf{I}_K & -\mathbf{G}_x\mathbf{h}_v^{-1} \\ \mathbf{A}_y & \mathbf{A}_x & \mathbf{A}_\eta \end{bmatrix}_{(G+N+K)\times(2N+K)}, \boldsymbol{\Psi} \equiv \begin{bmatrix} \boldsymbol{\xi}_y \\ \boldsymbol{\xi}_x \\ \boldsymbol{\eta} \end{bmatrix}_{(2N+K)\times 1}, \boldsymbol{\phi} \equiv \begin{bmatrix} \mathbf{F}_y^W \\ \mathbf{F}_x^W \\ \mathbf{a} \end{bmatrix}_{(G+N+K)\times 1},
$$

we have the identification condition for the linear equations to have solutions to $\boldsymbol{\Psi}$ since $\boldsymbol{\Psi}$ has $(2N+K)$ elements. ■

Appendix B

Appendix of Chapter 5

B.1 Theorem 5.1

Proof. Recall that $\mathbf{V} \equiv (V_L, V_U]$, where $V_L = \max_{z \in \overline{z}_m} P^{m-1}(z)$, and $V_U = \max_{z \in \overline{z}_m} P^{m+1}(z)$, and where $\overline{z}_m$ is the set of values of the values of Z that satisfy the rank condition.

Suppose that $Q_{U|VZ}(\tau_U|\tau_V, z)$ is weakly increasing in $\tau_V \in \mathbf{V}$. Then by Lemma 2 in Chesher (2005) we have for $Y = y^m$,

$$\begin{aligned} h(y^m, Q_{U|VZ}(\tau_U|V_L, z'_m)) &\leq Q_{W|YZ}(\tau_U|y^m, z'_m) & \text{(A-1)} \\ &\leq h(y^m, Q_{U|VZ}(\tau_U|P^m(z'_m), z'_m)) \\ h(y^m, Q_{U|VZ}(\tau_U|V_L, z''_m)) &\leq Q_{W|YZ}(\tau_U|y^m, z''_m) & \text{(A-2)} \\ &\leq h(y^m, Q_{U|VZ}(\tau_U|P^m(z''_m), z''_m)) \end{aligned}$$

and for $Y = y^{m+1}$

$$\begin{aligned} h(y^{m+1}, Q_{U|VZ}(\tau_U|P^m(z'_m), z'_m)) &\leq Q_{W|YZ}(\tau_U|y^{m+1}, z'_m) & \text{(A-3)} \\ &\leq h(y^{m+1}, Q_{U|VZ}(\tau_U|V_U, z'_m)) \\ h(y^{m+1}, Q_{U|VZ}(\tau_U|P^m(z''_m), z''_m)) &\leq Q_{W|YZ}(\tau_U|y^{m+1}, z''_m) & \text{(A-4)} \\ &\leq h(y^{m+1}, Q_{U|VZ}(\tau_U|V_U, z''_m)) \end{aligned}$$

Under Restriction RC, $P^m(z'_m) \leq \tau_V \leq P^m(z''_m)$, when $Q_{U|VZ}(\tau_U|\tau_V, z)$ is weakly increasing in v, then :

$$\begin{aligned} Q_{U|VZ}(\tau_U|\tau_V, z''_m) &\leq Q_{U|VZ}(\tau_U|P^m(z''_m), z''_m) \\ Q_{U|VZ}(\tau_U|P^m(z'_m), z'_m) &\leq Q_{U|VZ}(\tau_U|\tau_V, z'_m) \end{aligned}$$

and because h is weakly increasing in U,

$$\begin{aligned} h(y^m, Q_{U|VZ}(\tau_U|\tau_V, z''_m)) &\leq h(y^m, Q_{U|VZ}(\tau_U|P^m(z''_m), z''_m)) & \text{(B-1)} \\ h(y^m, Q_{U|VZ}(\tau_U|P^m(z'_m), z'_m)) &\leq h(y^m, Q_{U|VZ}(\tau_U|\tau_V, z'_m)). & \text{(B-2)} \end{aligned}$$

Combining (A-4) and (B-1) we can find the upper bound on $h(y^m, Q_{U|VZ}(\tau_U|\tau_V, z''_m))$

$$\begin{aligned} h(y^m, Q_{U|VZ}(\tau_U|\tau_V, z''_m)) &\leq h(y^m, Q_{U|VZ}(\tau_U|P^m(z''_m), z''_m)) \\ &\leq h(y^{m+1}, Q_{U|VZ}(\tau_U|P^m(z''_m), z''_m)) \\ &\leq Q_{W|YZ}(\tau_U|y^{m+1}, z''_m) \end{aligned}$$

The first inequality is due to (B-1) and the second inequality is due to Restriction LDRM, and the third inequality is due to (A-4).

The lower bound on $h(y^m, Q_{U|VZ}(\tau_U|\tau_V, z'_m))$ can be found by (A-3) and (B-2) :

$$Q_{W|YZ}(\tau_U|y^m, z'_m) \leq h(y^m, Q_{U|VZ}(\tau_U|P^m(z'_m), z'_m)) \leq h(y^m, Q_{U|VZ}(\tau_U|\tau_V, z'_m)).$$

The first inequality is due to (A-3), the second is due to (B-2).

Finally, under the conditional quantile invariance C-QI and exclusion Restrictions A-EX, there is for $z \in \{z'_m, z''_m\}$ for $u^* = Q_{U|VZ}(\tau_U|\tau_V, z)$,

$$Q_{W|YZ}(\tau_U|y^m, z'_m) \leq h(y^m, u^*) \leq Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)$$

Similarly, when $Q_{U|VZ}(\tau_U|\tau_V, z)$ is weakly decreasing in $\tau_V \in \mathbf{V}$, we have

$$Q_{W|YZ}(\tau_U|y^{m+1}, z''_m) \leq h(y^m, u^*) \leq Q_{W|YZ}(\tau_U|y^m, z'_m)$$

■

B.2 Corollary 5.2

Proof. We adopt Lemma 2 in Chesher (2005) when $m = 1$ with $P^0(z) = 0$ and $P^1(z) = P(z)$, where $P(z) = \Pr(Y = 1|Z = z)$ and when $m = 2$ with $P^2(z) = 1$ and $P^1(z) = P(z)$.

Suppose that $Q_{U|VZ}(\tau_U|v, z)$ is weakly increasing in v. Then we have

$$\begin{aligned} h(0, Q_{U|VZ}(\tau_U|0, z')) &\leq Q_{W|YZ}(\tau_U|0, z') && \text{(A-1)} \\ &\leq h(0, Q_{U|VZ}(\tau_U|P(z'), z')) \\ h(0, Q_{U|VZ}(\tau_U|0, z'')) &\leq Q_{W|YZ}(\tau_U|0, z'') && \text{(A-2)} \\ &\leq h(0, Q_{U|VZ}(u|P(z''), z'')) \\ h(1, Q_{U|VZ}(\tau_U|P(z'), z')) &\leq Q_{W|YZ}(\tau_U|1, z') && \text{(A-3)} \\ &\leq h(1, Q_{U|VZ}(\tau_U|1, z')) \\ h(1, Q_{U|VZ}(\tau_U|P(z''), z'')) &\leq Q_{W|YZ}(\tau_U|1, z'') && \text{(A-4)} \\ &\leq h(1, Q_{U|VZ}(\tau_U|1, z'')) \end{aligned}$$

We use (A-1) and (A-4).

$$\begin{aligned} Q_{W|YZ}(\tau_U|0,z') &\leq h(0,Q_{U|VZ}(\tau_U|P(z'),z')) && \text{(A-1)} \\ h(1,Q_{U|VZ}(\tau_U|P(z''),z'')) &\leq Q_{W|YZ}(\tau_U|1,z'') && \text{(A-4)} \end{aligned}$$

Under Restriction RC, $P(z') \leq \tau_V \leq P(z'')$, when $Q_{U|VZ}(\tau_U|v,z)$ is weakly increasing in v, then :

$$\begin{aligned} Q_{U|VZ}(\tau_U|\tau_V,z'') &\leq Q_{U|VZ}(\tau_U|P(z''),z'') \\ Q_{U|VZ}(\tau_U|P(z'),z') &\leq Q_{U|VZ}(\tau_U|\tau_V,z') \end{aligned}$$

and because h is monotonic in u and weakly increasing,

$$\begin{aligned} h(1,Q_{U|VZ}(\tau_U|\tau_V,z'')) &\leq h(1,Q_{U|VZ}(\tau_U|P(z''),z'')) && \text{(B-1)} \\ h(1,Q_{U|VZ}(\tau_U|P(z'),z')) &\leq h(1,Q_{U|VZ}(\tau_U|\tau_V,z')). && \text{(B-2)} \end{aligned}$$

Combining (A-4) and (B-1) we can find the upper bound for $h(1,Q_{U|VZ}(\tau_U|\tau_V,z''))$

$$h(1,Q_{U|VZ}(\tau_U|\tau_V,z'')) \leq h(1,Q_{U|VZ}(\tau_U|P(z''),z'')) \leq Q_{W|YZ}(\tau_U|1,z'')$$

Use the Restriction LDRM : $h(1,u) \geq h(0,u)$, for all values of z and u in the support of Z and $u \in \mathbf{U}$. Applying Restriction LDRM to (B-2)

$$h(0,Q_{U|VZ}(\tau_U|P(z'),z')) \leq h(1,Q_{U|VZ}(\tau_U|P(z'),z')) \leq h(1,Q_{U|VZ}(\tau_U|\tau_V,z')). \quad \text{(C)}$$

Applying (A-1) to (C), we have the lower bound for $h(1,Q_{U|VZ}(\tau_U|\tau_V,z'))$

$$Q_{W|YZ}(\tau_U|0,z') \leq h(1,Q_{U|VZ}(\tau_U|\tau_V,z')).$$

Finally, under the conditional independence restriction and exclusion Restriction C-QI and QCFA, there is for $z \in \{z',z''\}$ for $u^* = Q_{U|VZ}(\tau_U|v,z)$

$$Q_{W|YZ}(\tau_U|0,z') \leq h(1,u^*) \leq Q_{W|YZ}(\tau_U|1,z'') \quad \text{(D-1)}$$

Consider next the identification of $h(0,u^*)$.

Under Restriction RC, $P(z') \leq \tau_V \leq P(z'')$, when $Q_{U|VZ}(\tau_U|v,z)$ is weakly increasing in v, then :

$$\begin{aligned} Q_{U|VZ}(\tau_U|\tau_V,z'') &\leq Q_{U|VZ}(\tau_U|P(z''),z'') \\ Q_{U|VZ}(\tau_U|P(z'),z') &\leq Q_{U|VZ}(\tau_U|\tau_V,z') \end{aligned}$$

and because h is monotonic in U and weakly increasing,

$$\begin{aligned} h(0, Q_{U|VZ}(\tau_U|\tau_V, z'')) &\leq h(0, Q_{U|VZ}(\tau_U|P(z''), z'')) \quad \text{(B-3)} \\ h(0, Q_{U|VZ}(\tau_U|P(z'), z')) &\leq h(0, Q_{U|VZ}(\tau_U|\tau_V, z')). \quad \text{(B-4)} \end{aligned}$$

using (A-4) and (B-3), and Restriction LDRM we can find the upper bound for $h(0, Q_{U|VZ}(\tau_U|\tau_V, z''))$

$$\begin{aligned} & h(0, Q_{U|VZ}(\tau_U|\tau_V, z'')) \overset{(a)}{\leq} h(0, Q_{U|VZ}(\tau_U|P(z''), z'')) \\ & \overset{(b)}{\leq} h(1, Q_{U|VZ}(\tau_U|P(z''), z'')) \\ & \overset{(c)}{\leq} Q_{W|YZ}(\tau_U|1, z'') \end{aligned}$$

(a) is due to (B-3), (b) follows from Restriction LDRM, and (c) is from (A-4).

Applying (A-1) to (B-4) we have

$$Q_{W|YZ}(\tau_U|0, z') \overset{(a)}{\leq} h(0, Q_{U|VZ}(\tau_U|P(z'), z')) \overset{(b)}{\leq} h(0, Q_{U|VZ}(\tau_U|\tau_V, z')).$$

(a) follows from (A-4) and (b) is from (B-4). Thus, the lower bound for $h(0, Q_{U|VZ}(\tau_U|\tau_V, z'))$

$$Q_{W|YZ}(\tau_U|0, z') \leq h(0, Q_{U|VZ}(\tau_U|\tau_V, z')).$$

Finally, by Restriction C-QI and A-EX, there is for $z \in \{z', z''\}$

$$Q_{W|YZ}(\tau_U|0, z') \leq h(0, u^*) \leq Q_{W|YZ}(\tau_U|1, z'') \quad \text{(D-2)}$$

Note that the identified intervals for $h(0, u^*)$ and $h(1, u^*)$ are the same as we see in (D-1) and (D-2). ■

B.3 Lemma 5.1

Proof. We show the case in which $Q_{W|YZ}(\tau_U|y^m, z'_m) \leq Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)$. The other case can be shown similarly. We need to show that PDPR implies $Q_{W|YZ}(\tau_U|y^m, z'_m) \leq Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)$. Let Q''_{m+1} and Q'_m indicate the values of τ_U- quantiles, $Q'_m \equiv Q_{W|YZ}(\tau_U|y^m, z'_m)$ and $Q''_{m+1} \equiv Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)$. Then by definition of quantiles we have

$$\begin{aligned} \tau_U &= F_{W|YZ}(Q'_m|Y = y^m, Z = z'_m) \\ &= \Pr(W \leq Q'_m|Y = y^m, Z = z'_m) \quad \text{(A)} \\ &= \Pr(h(y^m, U) \leq Q'_m|Y = y^m, Z = z'_m) \\ &= \Pr(U \leq h^{-1}(y^m, Q'_m)|Y = y^m, Z = z'_m) \end{aligned}$$

similarly for Q''_{m+1}, we have

$$\tau_U = \Pr(U \leq h^{-1}(y^{m+1}, Q''_{m+1})|Y = y^{m+1}, Z = z''_m) \tag{B}$$

where h^{-1} is defined as (C*) in Appendix C. Suppose PDPR. Then we have

$$\begin{aligned} \tau_U &= \Pr(U \leq h^{-1}(y^{m+1}, Q''_{m+1})|Y = y^{m+1}, Z = z''_m) \qquad \text{(C-1)} \\ &= \Pr(U \leq h^{-1}(y^{m+1}, Q''_{m+1})|V \in (P^m(z''_m), P^{m+1}(z''_m)]) \\ &\leq \Pr(U \leq h^{-1}(y^{m+1}, Q''_{m+1})|V \in (P^{m-1}(z''_m), P^m(z''_m)]) \\ &= \Pr(U \leq h^{-1}(y^{m+1}, Q''_{m+1})|Y = y^m, Z = z''_m) \\ &= \Pr(U \leq h^{-1}(y^{m+1}, Q''_{m+1})|Y = y^m, Z = z'_m) \equiv \widetilde{u} \qquad \text{(C-2)} \end{aligned}$$

where the first equality is by (B), the second equality follows from that the event $\{V \in (P^m(z''_m), P^{m+1}(z''_m)]\}$ is equivalent to the event $\{Y = y^{m+1}, Z = z''_m\}$. The first inequality is due to PD ($F_{U|VZ}(u|v, z)$ is non-increasing in $v \in \mathbf{V}$), and the third equality results from the same logic as in the second equality. The last equality is due to Restriction C-QI. Then $\tau_U \leq \widetilde{u}$.

From (A) and (C-2), we have

$$\tau_U \overset{by\ (A)}{=} \Pr(U \leq \underbrace{h^{-1}(y^m, Q'_m)}_{u^*}|Y = y^m, Z = z'_m) \leq \Pr(U \leq \underbrace{h^{-1}(y^{m+1}, Q''_{m+1})}_{u^{**}}|Y = y^m, Z = z'_m) \overset{by\ (C\text{-}2)}{=} \widetilde{u}$$

since $\tau_U \leq \widetilde{u}$, which implies that

$$u^* \equiv h^{-1}(y^m, Q'_m) \leq h^{-1}(y^{m+1}, Q''_{m+1}) \equiv u^{**}$$

by the nondecreasing property of distribution function, i.e., if $a \leq a'$, $F_{A|B}(a|b) \leq F_{A|B}(a'|b)$. Then we have

$$\begin{aligned} Q'_m &= h(y^m, u^*) \\ Q''_{m+1} &= h(y^{m+1}, u^{**}) \end{aligned}$$

By PDPR and monotonicity of h in u, we have by inverting h^{-1}

$$\begin{aligned} Q'_m &= h(y^m, u^*) \leq h(y^{m+1}, u^*) \\ &\leq h(y^{m+1}, u^{**}) = Q''_{m+1} \end{aligned}$$

where the first inequality is due to PDPR and the second inequality is due to monotonicity of h in u. Thus, we have shown that $Q'_m \equiv Q_{W|YZ}(\tau_U|y^m, z'_m) \leq Q''_{m+1} \equiv Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)$. The other case can be shown similarly. ■

Appendix C

Proofs of Sharpness

Sharpness in **Theorem 5.2** in Chapter 5 is proved in this appendix following **Lemma 2.1** in Chapter 2. To contrast the role of Restriction LDRM, and to describe the logic behind the construction of the structure, sharpness of Chesher (2005) is discussed first briefly. Define

$$h^{-1}(y^m, x, w) \equiv \sup_u \{u : h(y^m, x, u) \leq w\}. \tag{C*}$$

This implies

$$h(y^m, x, h^{-1}(y^m, x, w)) \leq w \tag{C**}$$

with equality holding when $h(y^m, u)$ is strictly increasing in u.

Under the triangular system with the single index unobservables restriction, a variation of **(HR-SIU-C) (Chapter 1)** of the following exists when the endogenous variable is continuous

$$\underbrace{F_{U|VX}(h^{-1}(y, x, w)|g(y, x), x)}_{\textbf{Structure}} \overset{\overset{\textbf{generates}}{\Longrightarrow}}{\underset{\underset{I\text{identification}}{\Longleftarrow}}{=}} \underbrace{F_{W|YX}(w|y, x)}_{\textbf{Data}}.$$

By the interaction between h and $F_{U|VX}$, the distribution of the observables is generated. From this relationship alone h and $F_{U|VX}$ cannot be separately recovered from data. Identification of the structure $S \equiv \{h, F_{U|VX}\}$ is achieved by imposing several restrictions discussed in Chapter 3. For exposition, the identification result **(C) in Chapter 3** are reproduced ignoring exogenous covariates X other than the IV, Z :

$$\begin{aligned}
\underbrace{h(y, u^*)}_{\text{Structural feature, } \theta(S)} &= h(y, Q_{U|VZ}(\tau_U|\tau_V, z)) && \text{(C)}\\
&= \underbrace{Q_{W|YZ}(\tau_U|y, z)}_{\text{Functional of data, } \mathcal{G}(F_{W|YZ})} \\
\text{where } u^* &\equiv Q_{U|VZ}(\tau_U|\tau_V, z), \\
y &= Q_{Y|Z}(\tau_V|z). && \text{(C.1)}
\end{aligned}$$

Chesher (2003) did not consider identification of the distribution of the unobserved variables. However, the QCFA can be used for this purpose. The distribution of the unobservables is identified up to monotone transformation : only the shape is identified by the conditional distribution of the outcome given other covariates. Recall that U and V are each normalized Uniform $(0,1)$. Once the conditional distribution of U given V and X is identified, the joint distribution of U and V given other covariates, X, can be recovered since $F_{UV|X}(u,v|x) = F_{U|VX}(u|v,x)F_{V|X}(v|x)$, where $F_{V|X}(v|x)$ is assume to be uniform (0,1). The dependence between the unobservable variables is indicated by the conditional distribution, $F_{U|VX}$, which would *not* be uniform $(0,1)$ in the presence of endogeneity. Thus we focus on the identification of the conditional distribution, $F_{U|VX}(u|v,x)$.

For notational simplicity, we ignore other covariates, X, than an IV, Z. X can be added as conditioning variables in any steps without changing the results. The following identification of the value of the distribution of the unobservables can be stated as in (D). The value of $F_{U|VZ}(u|v,z)$ is identified by $F_{W|YZ}(w|y,z)$ when $v = g(v,z)$ and $u = h^{-1}(y,w)$. Note that once the values of W, Y, and Z are given, the values of U and V are determined by the structural relations, h^{-1} and g.

$$
\begin{aligned}
\underbrace{F_{U|VZ}(u|v,z)}_{\theta(S)} &= \Pr(U \leq u|v,z) \\
&= \Pr(U \leq u|g(y,z),z) \\
&= \Pr(U \leq h^{-1}(y,w)|g(y,z),z) \qquad \text{(D)} \\
&= \Pr(U \leq h^{-1}(y,w)|Y=y, Z=z) \\
&= \Pr(h(y,U) \leq h(y,h^{-1}(y,w))|Y=y,Z=z) \\
&= \Pr(W \leq w|Y=y,Z=z) \\
&= \underbrace{F_{W|YZ}(w|y,z)}_{\mathcal{G}(F_{W|YZ})} \\
\text{where } v &= g(y,z) \\
u &= h^{-1}(y,w)
\end{aligned}
$$

where the second equality follows from $v = g(y,x)$, the third equality is due to (C^*), the fourth equality follows from the fact that the event $\{V = g(y,x) \cap X = x\}$ is equivalent to the event $\{Y = y \cap X = x\}$ under the triangular structure, the sixth equality follows from $(a-2)$. However, there can be many pairs of $\{h, F_{U|VX}\}$, that generate $F_{W|YX}(w|y,x)$, thus, the distribution of the unobservables is only identified up to a monotone transformation.

All the values of the distribution of the unobservable variables is *completely* identified when W, Y and Z are *continuous.* Recall that the unobserved variables U and V are assumed to be *continuously* varying, and normalized uniform $(0,1)$. Continuity of U and V is the reason why loss of identifying power of the QCFA arises when the endogenous variable is discrete[1].

[1] I am grateful to Hide Ichimura for pointing this out. However, assuming that the unobserved types, U and V, take infinitely many values which are represented by numbers in (0,1) is, I suppose, a natural

When the endogenous variable is discrete, the identification in (C) fails, and the *whole*[2] shape of the conditional distribution of the unobservable variables can not be identified. The information on the distribution of the observables from data does not pin down the values of the structural function nor the shape of the distribution of the unobservables. This is mainly because of the fact that in general

$$\begin{aligned} u^* &\equiv Q_{U|VZ}(\tau_U|\underset{\Uparrow}{\tau_V}, z) \neq Q_{U|YZ}(\tau_U|\underset{\Uparrow}{y}, z) \\ \text{where } y &= Q_{Y|X}(\tau_V|x), \end{aligned} \tag{1}$$

when Y shows discrete variation, which results in $h(y, u^*) \neq Q_{W|YZ}(\tau_U|y, z)$.

From these identification results, the sharpness proofs start by constructing a structure, $\{h_a, F^a_{U|VZ}\}$, using the distribution of the observables, $F_{W|YZ}$. In the proofs, the construction of the structure is based on the identification results in (C) and (D) making adjustments where necessary reflecting the discreteness of Y and the restrictions imposed by each model.

C.1 Sharpness of Chesher (2005) Bound

The loss of identifying power of the QCFA arises by the fact that knowing the values of Y and Z does not pin down the value of V. Since U and V are dependent, the QCFA cannot control the "endogeneity" problem completely. Chesher (2005)'s idea is that if data satisfy a certain rank condition, the value of the structural function can be bounded by restricting the pattern of dependence between U and V in a certain way. That is, by imposing monotonicity of $F_{U|V}(u|v)$ in v, the possible range of the value of $h(y, u^*)$ is recovered from data.

The monotonicity restriction together with the strong rank condition does not have the identifying power when the endogenous variable is *binary,* because the strong rank condition in Chesher (2005) cannot be satisfied with binary endogenous variable. The model in Chapter 5 imposes more restriction in the sense that the pattern of the dependence between the unobserved types, U and V, should be matched with the patterns of the response in a certain way and the direction of the match can be varying with the level of the unobserved type, U. The benefit of this model is that the strong rank condition in Chesher (2005) can be relaxed thereby making it applicable with a binary endogenous variable, but the additional restriction, LDRM should be justified to be applied, instead.

In this section a candidate structure to show sharpness of Chesher (2005) bounds is discussed to clarify the role of Restriction LDRM in the model in Chapter 5. A modified version of the candidate structure to reflect the additional restriction, Restriction LDRM is used in the next section to show sharpness of LDRM bounds.

presumption. Assuming particular forms of discrete unobserved types as in Heckman and Singer (1984), which is the alternative to the continuous types, seems to be arbitrary.

[2] This is also related to the continuity of U and V. If they are discrete, there may be a way to point identify the structural function.

Recall that the identification problem of concern in Chesher (2005) is to find out the value of the structural function, h, evaluated at $Y = y^m$, $X = x$, and $U = Q_{U|VX}(\tau_U|\tau_V, z)$. The value of U is never known, but it is indicated as $u^* \equiv Q_{U|VX}(\tau_U|\tau_V, z)$. Other covariates X than Z, which is an IV, is ignored without affecting the results. They can be added as additional conditioning variables in any part of the proofs. The following lemma shows what happens when the endogenous variable is discrete and the unobservables are *continuous.* **(HR-SIU-C) (from Chapter 1)** have the following variation with a discrete endogenous variable.

Lemma C.1 ***Observational Equivalence*** *(Lemma 1 in Chesher (2005)) Under Restriction A-EX and C-QI, the conditional distribution of W given* $Y = y^m$ *and* $Z = z \in \overline{z}_m$ *with* $\overline{z}_m$ *is the set of values of Z that satisfy Chesher (2005)'s strong rank condition*

$$\begin{aligned} \overbrace{F_{W|YZ}(w|y^m, z)}^{\textbf{\textit{Data}}} &= \frac{1}{p_m(z)} \overbrace{\int_{P^{m-1}(z)}^{P^m(z)} F_{U|VZ}(h^{-1}(y^m, w)|s, z)ds}^{\textbf{\textit{Structure}}} \\ \textit{where } p_m(z) &= \Pr(Y = y_m|Z = z) \qquad\qquad (\textbf{HR-SIU-D}) \\ \textit{for } z &\in \overline{z}_m \equiv \{z'_m, z''_m\}. \end{aligned}$$

This lemma is the key in the construction of the distribution of the unobservables when Y is discrete.[3] There can be many pairs, $\{h, F_{U|VZ}\}$ that produce the same observed data $F_{W|YZ}$: the shape of the distribution of the unobservables ($F_{U|VZ}$) is *not* determined by the distribution of the observed variables completely, in contrast with when the endogenous variable is continuous. Not all such observationally equivalence structures satisfy all the restrictions imposed by the model.

The candidate structure is constructed based on the identification result (C) and **Lemma C.1**. The candidate structural relation (h_a) is chosen as a certain quantile of the distribution of the observables, $F_{W|YZ}$, since the value of the structural relation is identified by the quantile of the distribution of the observables. The candidate for the distribution of the unobservables ($F^a_{U|VZ}$) is chosen as a step function in v from the above observation. It is assumed that both U and V are continuously varying, but how $F_{U|VZ}(u|v, z)$ is varying for a range of V is hidden from the observed data, the constructed distribution is constant over a certain range of V. Suppose the integrand, $F_{U|VZ}(h^{-1}(y^m, w)|s, z)$, is constant for $s \in (P^{m-1}(z), P^m(z))$. Then the relation **(HR-SIU-D)** can be written, for example, as

$$\begin{aligned} F_{W|YZ}(w|y^m, z) &= F_{U|VZ}(h^{-1}(y^m, w)|v, z), \\ \text{for } v &\in (P^{m-1}(z), P^m(z)). \end{aligned}$$

[3]See Chesher (2010) for the proof of sharpness in the structural approach. Note that in his proof the key relation is

$$F^a_{WY|Z}(w^*, y|z) = F^a_{UY|Z}(h_a^{-1}(y, w^*), y|z)$$

since how Y is determined given Z is not specified as it is under triangularity. The proof in Chesher (2010) is done by constructing the distribution of the unobservables using the observables, and the construction of the structural function is not required since the information on the structural relation is included in the threshold crossing function ($P^m(y)$). The proof is concerned with constructing $F_{U|YZ}$, and using $F_{Y|Z}$ the object of interest $F_{UY|Z}$ can be recovered.

In **Part 1** a structure, a structural relation (h_a) and the distribution of the unobservables ($F^a_{U|VZ}$), is suggested as a candidate structure that generated the data. In **Part 2**, the three conditions in **Lemma 2.1 in Chapter 2** will be discussed.

Part 1 - Construction of an admitted and observationally equivalent(o.e.) structure $S^a \equiv \{h_a, F^a_{U|VZ}\}$, such that $\theta(S^a) = a$

If there were no restrictions imposed on h or $F_{U|VZ}$, then *any form of h and $F_{U|VZ}$ can be used as the candidate structure.* This is why in many studies sharpness proof is omitted. Otherwise, at least one candidate structure needs to be constructed such that the imposed restrictions are satisfied and this admissibility needs to be proven formally.

The candidate structure is constructed such that all the values of h and $F_{U|VZ}$ can be determined. Some of the restrictions such as Restriction LDRM imposed in the model in Chapter 5 are regarding *local* properties of the structure, while some of the restrictions such as monotonicity of h in u or whether the constructed distribution of the unobservables is weakly increasing should be shown for all the points in the support of U. To show such restrictions all the values in the support of the arguments of the structural function and the distribution of the unobservables need to be determined by the construction. Note also that there can be other ways of construction. The distribution of observables, $F_{W|YZ}$, is used in the construction of h_a and $F^a_{U|VZ}(u|v,z)$, such that by the interaction of h_a and $F^a_{U|VZ}(u|v,z)$, $F_{W|YZ}$ can be generated (the Hurwicz (1950a) relation).

1. Construction of h_a

The structural function is constructed as

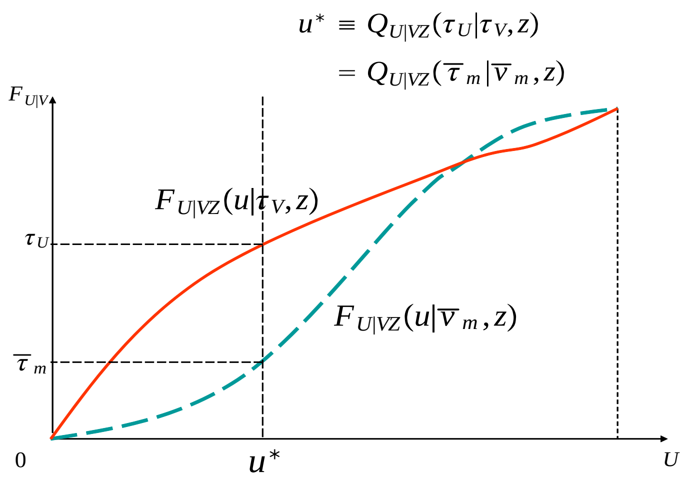
$$\begin{aligned} h_a(y^m, u^*) &\equiv Q^0_{W|YZ}(\overline{\tau}_m|y^m, z) \text{ for some } \overline{\tau}_m \text{ and } \overline{v}_m \qquad \text{(S1)} \\ \text{where } u^* &\equiv Q^a_{U|VZ}(\tau_U|\tau_V, z) \\ &= Q^a_{U|YZ}(\overline{\tau}_m|\overline{v}_m, z) \\ \text{for } m &= 1, 2, ..., M. \end{aligned}$$

For given $V = \tau_V$ and $Y = y^m$, by varying $\tau_U \in (0,1)$, all values of $h_a(y^m, u^*)$ are determined by (S1). All the values of the structural function for given τ_U can be defined as follows :

$$\begin{aligned} h_a(y, u^*) &\equiv \sum_{m=1}^{M} [Q^0_{W|YZ}(\overline{\tau}_m|y^m, z'_m)]1(y = y_m) \\ \text{where } u^* &\equiv Q^a_{U|VZ}(\tau_U|\tau_V, z) \\ &= Q^a_{U|VZ}(\overline{\tau}_m|\overline{v}_m, z) \end{aligned}$$

Remarks :

- When Y is *continuous*, the value of $h_a(y, u^*)$, where $u^* \equiv Q_{U|VZ}(\tau_U|\tau_V, z)$, is identified by $Q^0_{W|YZ}(\tau_U|y, z)$, where $Y = y = Q_{Y|Z}(\tau_V|z)$.



- The equality fails to hold with discrete Y as is discussed in (1) in Appendix C. There remains certain ambiguity regarding which value of V corresponds to $Y = y^m$ given Z.

- It should be noted that even though exclusion of Z in the structural function h is assumed, the value of Z can affect the value of Y via the auxiliary equation for Y in the triangular system.

- All the values of $h_a(y, u)$ in the support of Y and U are defined (globally defined).

  – In (S1), for given value, u^*, the value of the structural function $h_a(y^m, u^*)$ can take M values. The *value*, $Q^0_{W|YZ}(\overline{\tau}_m|y^m, z'_m)$ is assigned as the value of $h_a(y^m, u^*)$, where $u^* \equiv Q_{U|VZ}(\tau_U|\tau_V, z)$, for $m = 1, 2, .., M$. (S1) does not indicate that $h_a(y^m, u^*)$ is a function of Z.

  – In (S1), on the other hand, for given $Y = y^m$, different values of $h_a(y^m, u^*)$ can be assigned by varying the value of τ_U.

2. Construction of $F^a_{U|VZ}$ and proof of proper distribution

For a given structural relation, h_a, and given the values of $Y = y^m$ and an arbitrary value of $U = u \in (0, 1)$ can be written as

$$h_a^{-1}(y^m, w^\#)$$

for some $w^\#$ by (C*). Then we can find $w^1, w^2, ..., w^M$ for the fixed value u such that

$$w^l = h_a(y^l, u), \text{ for } l = 1, 2, ..., M$$

so that

$$h_a^{-1}(y^m, w^\#) = h_a^{-1}(y^1, w^1) = h_a^{-1}(y^2, w^2) = \cdots = h_a^{-1}(y^M, w^M)$$

for continuous W .

Let $SUPP(Z)$ be the support of Z. For an arbitrary value $u \in (0,1)$, u is expressed as $u = h_a^{-1}(y^m, w^{\#})$, for some $w^{\#}$. For a given $z \in SUPP(Z)$, for any $u, v \in (0,1) \times (0,1)$, $F^a_{U|VZ}(u|v,z)$ is constructed as follows :

$$
\begin{aligned}
F^a_{U|VZ}(u|v,z) &= F^a_{U|VZ}(\underbrace{h_a^{-1}(y^m, w^{\#})}_{u}|v,z) \\
&\equiv \left(\begin{array}{lll}
F^0_{W|Y}(w^1|y^1,z), & \text{if} & 0 < v \le P^1 \\
F^0_{W|Y}(w^2|y^2,z), & \text{if} & P^1 < v \le P^2 \\
& \cdots & \\
F^0_{W|Y}(w^{m-1}|y^{m-1},z), & \text{if} & P^{m-2} < v \le P^{m-1} \\
F^0_{W|Y}(w^{\#}|y^m,z), & \text{if} & P^{m-1} < v \le P^m \\
F^0_{W|Y}(w^{m+1}|y^{m+1},z), & \text{if} & P^m < v \le P^{m+1} \\
& \cdots & \\
F^0_{W|Y}(w^M|y^M,z), & \text{if} & P^{M-1} < v \le 1
\end{array} \right) \\
& \text{where } w^1, w^2, ..., w^M \text{ are found such that} \\
w^l &= h_a(y^l, u), \\
P^l &= \max_{z \in SUPP(Z)} \{P^l(z)\}, \ l \ne m-1, m \\
P^{m-1} &= \min_{z \in \bar{z}_m} \{P^m(z)\} \text{ and } P^m = \max_{z \in \bar{z}_m} \{P^m(z)\} \\
l &= 0,1,2,...,M, \text{ with } P^0 = 0, P^M = 1
\end{aligned} \tag{S2}
$$

Remarks

- For a given value v, if $v \in (P^{l-1}(z), P^l(z)]$, assign $Y = y^l$, as the conditioning value, $l = 1,2,...,M$.
- If $u = h_a^{-1}(y^m, w^{\#})$ and $v \in (P^{l-1}(z), P^l(z)]$, where $l \ne m$ then find the value, w^l such that
$$w^l = h_a(y^l, u)$$
and assign $F^a_{U|VZ}(u|v,z)$ the value $F^0_{W|Y}(w^l|y^l,z)$.
- If $u = h_a^{-1}(y^m, w^{\#})$ and $v \in (P^{m-1}(z), P^m(z)]$, then assign the value, $F^0_{W|Y}(w^{\#}|y^m,z)$, to $F^a_{U|VZ}(u|v,z)$.
- $\{P^l\}_{l=1}^M$ is a weakly increasing sequence. The partition of the support of $V, (0,1)$, by $\{P^l\}_{l=1}^M$ is determined once a variable Z is given.Therefore, the partition does not vary with different values of Z, but the assigned value vary with the values of Z.
- $P^{m-1} = \min_{z \in \bar{z}_m} \{P^m(z)\}$ and $P^m = \max_{z \in \bar{z}_m} \{P^m(z)\}$ is chosen to guarantee the conditional quantile invariance restriction, which locally holds for τ_U quantile of U given V and Z, for the range of V specified by the *rank* condition.

- If W is discrete, $F^a_{U|VZ}(u|v,z)$ should be a step function in u as well as in v. For notational simplicity, we assume that W is continuous. Other parts in the proof are not affected when W is discrete, but in each part of the proof extra complication of notation needs to be introduced.

Proof of proper distribution

It is required to check whether the constructed distribution is proper : since each $F^0_{W|YZ}(w|y^l,z)$, for all $l \in \{1,2,...,M\}$ is a proper distribution, $F^0_{W|YZ}(w|y^l,z)$ lies between zero and one, thus, the values of constructed distribution, $F^a_{U|VZ}(u|v,z)$, lie between zero and one, but to guarantee the nondecreasing property of $F^a_{U|VZ}(u|v,z)$ in u for given v and z, we need to show that as u increases, w increases for given v and z. This can be shown by the following Lemma.

Lemma C.2 *For given v and z, $F^a_{U|VZ}(u|v,z)$ weakly increases in u.*

Proof. Consider two distinct values u' and u''. We express u' and u'' using h_a^{-1}, for given $Y = y^m$ as the following

$$\begin{aligned} u' &= h_a^{-1}(y^m, w') \\ u'' &= h_a^{-1}(y^m, w'') \end{aligned}$$

Fix $V = v$ and $Z = z$ and suppose that $V = v$ and $Z = z$ corresponds to $Y = y^l$, $l = 1,2,...,M$. Then by (S2) we have for some τ', τ'', w'_l, and w''_l

$$\begin{aligned} \tau' &= F^a_{U|VZ}(u'|v,z) \\ &\overset{(S2)}{=} \left\{ \begin{matrix} F^0_{W|YZ}(w'_l|y^l,z), & \text{if } l \neq m-1, m \\ F^0_{W|YZ}(w'|y^l,z), & \text{if } l = m-1,\ m \end{matrix} \right\} \\ \text{where } u' &= h_a^{-1}(y^m, w') = h_a^{-1}(y^l, w'_l) \end{aligned} \tag{1-1}$$

and

$$\begin{aligned} \tau'' &= F^a_{U|VZ}(u''|v,z) \\ &\overset{(S2)}{=} \left\{ \begin{matrix} F^0_{W|YZ}(w''_l|y^l,z), & \text{if } l \neq m-1, m \\ F^0_{W|YZ}(w''|y^l,z), & \text{if } l = m-1,\ m \end{matrix} \right\} \\ \text{where } u'' &= h_a^{-1}(y^m, w'') = h_a^{-1}(y^l, w''_l). \end{aligned} \tag{1-2}$$

If we can show that $w''_l \geq w'_l$, when $u'' \geq u'$, then the proof is done because then the assigned value following (S2) for $F^a_{U|VZ}(u''|v,z)$ is larger than $F^a_{U|VZ}(u'|v,z)$. Suppose

$$\begin{aligned} u'' &= Q^a_{U|VZ}(\tau''|v,z) \\ &\geq Q^a_{U|VZ}(\tau'|v,z) = u' \\ \text{for } \tau'' &\geq \tau'. \end{aligned}$$

Then $w_l'' \geq w_l'$, since from (1-1) and (1-2)

$$w_l'' = Q^0_{W|YZ}(\tau''|y^l, z) \geq Q^0_{W|YZ}(\tau'|y^l, z) = w_l'$$

whenever $u'' = Q^a_{U|VZ}(\tau''|v, z) \geq Q^a_{U|VZ}(\tau'|v, z) = u'$, that is, whenever $\tau'' \geq \tau'$. ■

Part 2

Part 2 - A. $h^a(y^m, u^*) = w^*$, $\forall w^* \in I(\tau, y^m, \overline{z}_m)$, where $u^* = Q_{U|VZ}(\tau_U|\tau_V, z)$, and $I(\tau, y^m, \overline{z}_m)$ is the Chesher (2005) bound where $\overline{z}_m$ is the set of values of Z that satisfy Chesher (2005)'s strong rank condition.

Note that under Restriction Common Support, any point in the identified interval, $w^* \in I(\tau, y^m, \overline{z}_m)$ can be written as (see <Figure C.1>)[4]

$$w^* = Q^0_{W|YZ}(\overline{\tau}_m|y^m, z_m') \text{ for some } \overline{\tau}_m \geq \tau_U.$$

That is,

$$\overline{\tau}_m = F^0_{W|YZ}(w^*|y^m, z_m') \text{ for some } \overline{\tau}_m \geq \tau_U$$

Note also that for **any** $v \in (P^{m-1}, P^m]$ by construction from (S2)

$$F^a_{U|VZ}(\underbrace{h_a^{-1}(y^m, w^*)}_{\overline{\tau}_m\text{-quantile of } F^a_{U|VZ}} |v, z_m') \overset{(S2)}{=} F^0_{W|YZ}(w^*|y^m, z_m') = \overline{\tau}_m,$$

thus, by definition of quantiles,

$$h_a^{-1}(y^m, w^*) = Q^a_{U|VZ}(\overline{\tau}_m|v, z_m') \text{ for } \textbf{some } v \in (P^{m-1}, P^m] \qquad (h_a - a)$$

For a given value, w^*, in the identified interval, $\overline{\tau}_m(\geq \tau_U)$ is determined by w^*. Then $(h_a - a)$ holds for a range of values of $v \in (P^{m-1}, P^m]$. Now we choose $\overline{v}_m \in (P^{m-1}, P^m]$ such that

$$\begin{aligned} u^* &\equiv Q^a_{U|VZ}(\tau_U|\tau_V, z_m') \qquad (h_a - b) \\ &= Q^a_{U|VZ}(\overline{\tau}_m|\overline{v}_m, z_m'). \end{aligned}$$

Then by inverting h_a^{-1} in $(h_a - a)$, for given w^* and $\overline{\tau}_m(\geq \tau_U)$, we have

$$\begin{aligned} w^* &= h_a(y^m, Q^a_{U|VZ}(\overline{\tau}_m|\overline{v}_m, z_m')) \\ &= h_a(y^m, u^*). \end{aligned}$$

Thus, this construction guarantees that the constructed structural function crosses

[4] Alternatively, one can find $\overline{\tau}_m$ such that $w^* = Q^0_{W|YZ}(\overline{\tau}_m|y^m, z_m'')$ for some $\overline{\tau}_m \leq \tau_U$

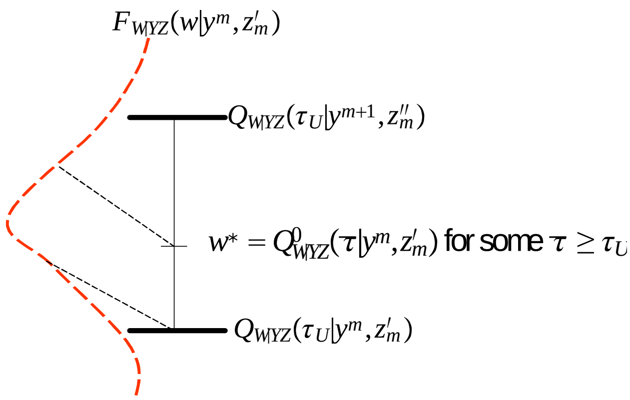


Figure C.1: Any point in the interval, $w^* \in I(\tau, m, \overline{z}_m)$, can be expressed using the quantiles of $F_{W|YZ}(w|y^m, z'_m)$ under the common support restriction. If the common support condition does not hold, then some of the points in the identified set, one cannot express them as a τ_m-quantile of W given Y and Z.

the arbitrary value in the identified interval

$$w^* = h_a(y^m, u^*),$$

that is, there exists a structural relation (that satisfies all the restrictions imposed by the model, which will be shown in the next section) which crosses an arbitrary point, w^*, in the identified interval.

Part 2 - B : Observational equivalence ($F^a_{W|YZ} = F^0_{W|YZ}$) : See Part 2 - B in Section C.2.

Part 2 - C : Admissibility by the Model in Chesher (2005)

0. Rank condition : this can be shown using data. This restriction is assumed to be satisfied.

1. Monotonicity of $h_a(y^m, u)$ **in** u : See Part 2 - C.1 in **Section C.2**.

2. Conditional Quantile Invariance : See Part 2 - C.2 in **Section C.2.**

3. Monotonicity of $F_{U|VZ}$ in v : See Part 2 - C.3 - (1) in **Section C.2**.

In the next section sharpness stated in Theorem 5.2 is shown. Note that LDRM model adopts all the restrictions imposed in Chesher (2005), but relaxes the Chesher (2005)'s strong rank condition and imposes one more restriction Restriction LDRM. The rank condition can be tested if data are available. If Chesher (2005)'s rank condition holds, the weak rank condition in the model LDRM always holds, in this sense, it is as if Chesher (2005)'s rank condition is imposed. Thus, the model in Chapter 5 is nested by Chesher (2005) model if Chesher (2005)'s rank condition holds.

The sharpness proof of the bound in Chapter 5 requires to show Restriction LDRM in addition to the restrictions in Chesher (2005). For that purpose the construction of the distribution of the unobservables should be modified slightly.

C.2 Theorem 5.2

Notation : The case in which $F_{U|VZ}(u^*|v,z)$ is nonincreasing in v, for $u^* \in \mathbf{U}$ is called PD (Positive Dependence) and the other case, ND (Negative Dependence). The case in which $h(y^{m+1},u^*) \geq h(y^m,u^*)$ is called PR (Positive Response) and the other case, NR (Negative Response).

Let $I(\tau, y^m, \overline{z}_m)$ denote the identified interval for PDPR. Sharpness of the other case can be shown similarly.

Following **Lemma 2.1. in Chapter 2**, what is required to show sharpness is to construct a structure (S^a) such that (A) for any value, $w^* \in I(\tau, y^m, \overline{z}_m)$, $w^* = h^a(y^m, u^*)$, and to show that (B) the constructed structure is observationally equivalent to the true structure ($F^a_{W|YZ} = F^0_{W|YZ}$) and (C) is admitted by LDRM model ($S^a \in \mathcal{M}^{LDRM}$). In **Part 1** we construct a structure $S^a \equiv \{h_a, F^a_{U|VZ}(u|v,z)\}$ and in **Part 2** we show (A),(B), and (C) .

Part 1. Construction of a candidate structure :

1-A Construction of a structural function

The same structural relation, (S1), for Chesher (2005) bound is used.

1-B Construction of the distribution of the unobservables.

For a given structural relation, h_a, and given the values of $Y = y^m$ and an arbitrary value of $U = u \in (0,1)$ can be written as

$$h_a^{-1}(y^m, w^{\#})$$

for some $w^{\#}$ by (C*). Then we can find $w^1, w^2, ..., w^M$ for the fixed value u such that

$$w^l = h_a(y^l, u), \text{ for } l = 1, 2, ..., M$$

so that

$$h_a^{-1}(y^m, w^{\#}) = h_a^{-1}(y^1, w^1) = h_a^{-1}(y^2, w^2) = \cdots = h_a^{-1}(y^M, w^M)$$

for continuous W .

Let $SUPP(Z)$ be the support of Z. For an arbitrary value $u \in (0,1)$, u is expressed as $u = h_a^{-1}(y^m, w^{\#})$, for some $w^{\#}$. For a given $z \in SUPP(Z)$, for any $u, v \in (0,1) \times (0,1)$, $F^a_{U|VZ}(u|v,z)$ is constructed as follows :

$$
\begin{aligned}
F^a_{U|VZ}(u|v,z) &= F^a_{U|VZ}(\underbrace{h_a^{-1}(y^m,w^{\#})}_{u}|v,z) \\
&\equiv \begin{pmatrix}
F^0_{W|Y}(w^1|y^1,z), & \text{if} & 0<v\le P^1 \\
F^0_{W|Y}(w^2|y^2,z), & \text{if} & P^1<v\le P^2 \\
 & \cdots & \\
F^0_{W|Y}(\underset{\upuparrows}{w^{\#}}|y^{m-1},z),\ (*) & \text{if} & P^{m-2}<v\le P^{m-1} \\
F^0_{W|Y}(\underset{\upuparrows}{w^{\#}}|y^{m},z),\ (*) & \text{if} & P^{m-1}<v\le P^{m} \\
F^0_{W|Y}(w^{m+1}|y^{m+1},z), & \text{if} & P^m<v\le P^{m+1} \\
 & \cdots & \\
F^0_{W|Y}(w^M|y^M,z), & \text{if} & P^{M-1}<v\le 1
\end{pmatrix} \qquad (\text{S2}') \\
& \text{where } w^1,w^2,...,w^M \text{ are found such that} \\
w^l &= h_a(y^l,u), \\
P^l &= \max_{z\in SUPP(Z)}\{P^l(z)\},\ l\neq m-1,m \\
P^{m-1} &= \min_{z\in\overline{z}_m}\{P^m(z)\} \text{ and } P^m=\max_{z\in\overline{z}_m}\{P^m(z)\} \\
l &= 0,1,2,...,M,\ \text{with } P^0=0, P^M=1
\end{aligned}
$$

Remarks

- For any given value v, if $v\in(P^{l-1},P^l]$, uses$Y=y^l$, as the conditioning value.
- If u is expressed as $h_a^{-1}(y^m,w^{\#})$ for some w, in the identified interval, and $v\in(P^{l-1},P^l]$, where $l\neq m-1$ and m, then find a value, w^l such that

  $$w^l=h_a(y^l,u)$$

  then assign the value $F^a_{U|VZ}(u|v,z)\equiv F^0_{W|YZ}(w^l|y^l,z)$.
- In (*) in (S2′) if $u=h_a^{-1}(y^m,w^{\#})$ and $v\in(P^{m-2},P^{m-1}]$, then assign $F^a_{U|VZ}(u|v,z)\equiv F^0_{W|YZ}(w^{\#}|y^{m-1},z)$. Note the value, $w^{\#}$, (indicated by $\upuparrows$) is assigned in contrast with in the previous section for Chesher (2005) bound.
- In (*) in (S2′) if $u=h_a^{-1}(y^m,w^{\#})$ and $v\in(P^{m-1},P^m]$, then assign $F^a_{U|VZ}(u|v,z)\equiv F^0_{W|YZ}(w^{\#}|y^m,z)$.
- Note also that this is a special case of the Chesher (2005) setup - that is, the restrictions imposed in Chesher (2005) model can be shown to be satisfied, once the restrictions in LDRM model are shown to hold.

Part 2

Part 2 - A : For any value, $w^* \in I(\tau, y^m, \overline{z}_m)$, $w^* = h_a(y^m, u^*)$. This follows from Part 2 - A in Section C.1.

Part 2 - B : Observational equivalence[5] $(F^a_{W|YZ} = F^0_{W|YZ})$

We need to show that $F^a_{W|YZ} = F^0_{W|YZ}$, for $S^a = \{h_a, F^a_{U|VZ}\}$ constructed as in **Part 1** : for $p^a_m = P^m - P^{m-1}$, for all $m \in \{1, 2, ..., M\}$,

$$\begin{aligned} F^a_{W|YZ}(w|y^m, z) &= \frac{1}{p^a_m}\int_{P^{m-1}}^{P^m} F^a_{U|VZ}(h_a^{-1}(y^m, w)|s, z)ds \\ &= \frac{1}{p^a_m}\int_{P^{m-1}}^{P^m} F^0_{W|YZ}(w|y^m, z)ds \\ &= F^0_{W|YZ}(w|y^m, z) \end{aligned}$$

the first equality is due to Lemma 1 in Chesher (2005), the second equality is due to construction in (S2′), that is, $F^a_{U|VZ}(h_a^{-1}(y^m, w)|v, z) = F^0_{W|YZ}(w|y^m, z)$, for $v \in (P^{m-1}, P^m]$ and the last equality is due to integration over the constant and the definition of p^a_m.

Part 2 - C : Admissibility by the model $S^a \in \mathcal{M}^{LDRM}$

0. Rank condition : this can be shown using data. We suppose this restriction is satisfied.

1. Monotonicity of $h_a(y^m, u)$ in u

I consider whether $h_a(y, u)$ is nondecreasing in u. Recall that

$$h_a(y^m, u) = h_a(y^m, Q^a_{U|VZ}(\overline{\tau}_m|\overline{v}_m, z)) \overset{\text{by } (S1)}{\equiv} Q^0_{W|YZ}(\overline{\tau}_m|y^m, z)$$

by choosing $\overline{v}_m$ such that $u = Q^a_{U|VZ}(\tau_U|\tau_V, z) = Q^a_{U|VZ}(\overline{\tau}_m|\overline{v}_m, z) = Q_{U|YZ}(\overline{\tau}_m|y^m, z)$, for $\forall \tau_U, \tau_V, \overline{\tau}_m \in (0, 1)$ and $\overline{v}_m \in (P^{m-1}, P^m]$.

- First, fix $\overline{v}_m$, then $h_a(y^m, u)$ is weakly increasing in u since higher $\overline{\tau}_m$ implies higher $u = Q_{U|VZ}(\overline{\tau}_m|\overline{v}_m, z)$, as well as higher $Q_{U|YZ}(\overline{\tau}_m|y^m, z)$.
- Next fix $\overline{\tau}_m$, if we observe higher u, then it is because of higher $\overline{v}_m$ if $F_{U|V}(u|\overline{v}_m, z)$ is nonincreasing in $\overline{v}_m$ and lower $\overline{v}_m$ if $F_{U|VZ}(u|\overline{v}_m, z)$ is nondecreasing in $\overline{v}_m \in (P^{m-1}, P^m]$. However, regardless of the direction of monotonicity, for $\overline{v}_m \in (P^{m-1}, P^m]$, $Y = y^m$. Thus, the value of $\overline{v}_m$ does not affect the value of h_a as long as Y is fixed at $Y = y^m$. That is, for fixed $\overline{\tau}_m$, and Y, $h_a(y, u)$ is constant as u increases due to change in $\overline{v}_m$.

[5] That is, the data distribution that is generated by the structure constructed in part 1 is actually what we observe. Note that this can be shown because we have constructed the structure using the observed distribution.

2. Conditional Quantile Invariance : $u^* \equiv Q^a_{U|VZ}(\tau_U|\tau_V, z)$ is invariant with respect to $z \in \overline{z}_m \equiv \{z'_m, z''_m\}$, for $\tau_V \in [P^m(z'_m), P^m(z''_m)]$. Note that there should exist a true structure, $S^0 = \{h_0, F^0_{U|VZ}\} \in \mathcal{M}^{LDRM} \cap \Omega_0$, that generates the data we observe. The distinction of the true structure, S^0 from the constructed structure, S^a, should be noted in this proof. For $u^* = h_a^{-1}(y^m, w^*)$

$$
\begin{aligned}
\tau_U &\equiv F^a_{U|VZ}(u^*|\tau_V, z'_m) \\
&= F^0_{W|YZ}(w^*|y^m, z'_m) \\
&= \frac{1}{p_m(z'_m)} \int_{P^{m-1}(z'_m)}^{P^m(z'_m)} F^0_{U|VZ}(h_0^{-1}(y^m, w^*)|s, z'_m)ds \\
&= \frac{\Pr(U \leq h_0^{-1}(y^m, w^*) \cap P^{m-1}(z'_m) \leq V \leq P^m(z'_m))}{p_m(z'_m)} \\
&= F^0_{U|V}(h_0^{-1}(y^m, w^*)|V \in (P^{m-1}(z'_m), P^m(z'_m)]) \\
&= F^0_{U|Y}(h_0^{-1}(y^m, w^*)|y^m) \\
&= F^0_{U|Y}(u^*|y^m)
\end{aligned}
$$

the first equality is by construction in (S2$'$), the second equality is due to Lemma 1 in Chesher (2005), and the third equality follows by integration. The fourth equality is by definition of the conditional probability, the fifth equality is due to how the value of Y is determined. Similarly for $Z = z''_m$,

$$
\begin{aligned}
\tau_U &\equiv F^a_{U|VZ}(u^*|\tau_V, z''_m) \\
&= F^0_{W|YZ}(w^*|y^m, z''_m) \\
&= \frac{1}{p_m(z''_m)} \int_{P^{m-1}(z''_m)}^{P^m(z''_m)} F^0_{U|VZ}(h_0^{-1}(y^m, w^*)|s, z''_m)ds \\
&= \frac{\Pr(U \leq h_0^{-1}(y^m, w^*) \cap P^{m-1}(z''_m) \leq V \leq P^m(z''_m))}{p_m(z''_m)} \\
&= F^0_{U|V}(h_0^{-1}(y^m, w^*)|V \in (P^{m-1}(z''_m), P^m(z''_m)]) \\
&= F^0_{U|Y}(h_0^{-1}(y^m, w^*)|y^m) \\
&= F^0_{U|Y}(u^*|y^m)
\end{aligned}
$$

yielding $u^* = Q^a_{U|VZ}(\tau_U|\tau_V, z'_m) = Q^a_{U|VZ}(\tau_U|\tau_V, z''_m) = Q^0_{U|Y}(\tau_U|y^m)$, invariant with respect to $z \in \overline{z}_m$.

3. LDRM :

(1) First, it is noted that $F^a_{U|VZ}(u|v, z)$ is monotonic in v, for $u \in \mathbf{U}, v \in \mathbf{V}$, where $\mathbf{U}$ and $\mathbf{V}$ are defined in Restriction LDRM. This is so since $F^a_{U|VZ}(u|v, z)$ is defined as a step function in v, for the range of $\mathbf{V}$ only two constants ($F^0_{W|YZ}(w^*|y^m, z)$, and $F^0_{W|YZ}(w^{m+1}|y^{m+1}, z)$) should be considered, and with two constants, monotonicity always holds.

(2) Now we check whether the constructed $S^a = \{h_a, F^a_{U|VZ}\}$ satisfies the specified match. Suppose for some $\tau''_m, \tau''_{m+1}, P^m(z''_m)$ and $P^{m+1}(z''_m)$,

$$\begin{aligned} u^* &\equiv Q^a_{U|VZ}(\tau_U|\tau_V, z) \\ &= Q^a_{U|VZ}(\tau''_m|P^m(z''_m), z''_m) = Q^a_{U|VZ}(\tau''_{m+1}|P^{m+1}(z''_m), z''_m), \end{aligned} \tag{3-1}$$

This can be shown by observing the sign of

$$\begin{aligned} & h_a(y^m, u^*) - h_a(y^{m+1}, u^*) \\ = \; & h_a(y^m, Q^a_{U|VZ}(\tau''_m|P^m(z''_m), z''_m)) - h_a(y^{m+1}, Q^a_{U|VZ}(\tau''_{m+1}|P^{m+1}(z''_m), z''_m)) \\ = \; & Q^0_{W|YZ}(\tau''_m|y^m, z''_m) - Q^0_{W|YZ}(\tau''_{m+1}|y^{m+1}, z''_m), \end{aligned} \tag{3-2}$$

where the first equality follows by (3-1), and the second equality is by construction in (S1).

To determine the sign of $h_a(y^m, u^*) - h_a(y^{m+1}, u^*)$, it is required to determine the sign of $Q^0_{W|YZ}(\tau''_m|y^m, z''_m) - Q^0_{W|YZ}(\tau''_{m+1}|y^{m+1}, z''_m)$. We first fix $U = u^*$, and vary the value of V. Then use the monotonicity of $F_{U|VZ}$ in v in a certain range specified in the restriction and see if the match holds.

(3-3)-(3-5) link the distribution of the unobservables with the distribution of the observables, and they are found by expressing u^* using h_a^{-1} and the construction in Part 1.

For $u^* = h_a^{-1}(y^m, w^*)$ and $v = P^m(z''_m)$, let τ''_m be

$$\begin{aligned} \tau''_m &\equiv F^a_{U|VZ}(u^*|P^m(z''_m), z''_m) \\ &= F^a_{U|VZ}(\underbrace{h_a^{-1}(y^m, w^*)}_{\tau''_m - \text{ quantile of } F^a_{U|VZ}} |P^m(z''_m), z''_m) \\ &= F^0_{W|YZ}(\underbrace{w^*}_{\tau''_m - \text{ quantile of } F^0_{W|YZ}} |y^m, z''_m) \end{aligned} \tag{3-3}$$

Note that for $u^* = h_a^{-1}(y^{m+1}, w^{m+1})$ and $v = P^{m+1}(z''_m)$, let τ''_{m+1} be:

$$\begin{aligned} \tau''_{m+1} &\equiv F^a_{U|VZ}(u^*|P^{m+1}(z''_m), z''_m) \\ &= F^a_{U|VZ}(\underbrace{h_a^{-1}(y^{m+1}, w^{m+1})}_{\tau''_{m+1} - \text{ quantile of } F^a_{U|VZ}} |P^{m+1}(z''_m), z''_m) \\ &= F^0_{W|YZ}(\underbrace{w^{m+1}}_{\tau''_{m+1} - \text{ quantile of } F^0_{W|YZ}} |y^{m+1}, z''_m) \end{aligned} \tag{3-4}$$

Also, for $P^m(z'_m) < \overline{v} < P^m(z''_m)$, we have[6]

$$\begin{aligned} \overline{\tau} &\equiv F^a_{U|VZ}(u^*|v, z''_m) \\ &= F^a_{U|VZ}(\underbrace{h_a^{-1}(y^{m+1}, w^{m+1})}_{\overline{\tau}-\text{ quantile of } F^a_{U|VZ}}|\overline{v}, z''_m) \\ &= F^0_{W|YZ}(\underbrace{w^{m+1}}_{\overline{\tau}-\text{ quantile of } F^0_{W|YZ}}|y^m, z''_m) \end{aligned} \tag{3-5}$$

Step 2 : Order of (3-3)-(3-5) :

Note $P^m(z'_m) \le P^m(z''_m) \le P^{m+1}(z''_m)$. Then PD implies that

$$\tau''_{m+1} \le \tau''_m \le \overline{\tau} \tag{*PD}$$

since we are comparing the values of the three conditional distributions evaluated at the same value u^*. And ND implies that

$$\tau''_{m+1} \ge \tau''_m \ge \overline{\tau} \tag{*ND}$$

Step 3 : Quantile expressions for w and u^*

Now we express u^* and w^* and w^{m+1} as quantiles of the distributions so that we can find the order of the two, $h_a(y^m, u^*)$ and $h_a(y^{m+1}, u^*)$ using (*PD) and (*ND). (4-2)-(4-5) imply (4-6) and (4-7) under continuity of W and U :

$$\begin{aligned} u^* &= Q^a_{U|VZ}(\tau''_m|P^m(z''_m), z''_m) \\ &= Q^a_{U|VZ}(\tau''_{m+1}|P^{m+1}(z''_m), z''_m) \\ &= Q^a_{U|VZ}(\tau'_{m+1}|P^{m+1}(z'_m), z'_m) \\ &= Q^a_{U|VZ}(\overline{\tau}|\overline{v}, z''_m), \quad \text{for } P^m(z'_m) < \overline{v} < P^m(z''_m) \end{aligned} \tag{3-6}$$

$$\begin{aligned} w^* &\overset{(a)}{=} Q^0_{W|YZ}(\tau''_m|y^m, z''_m) = Q^0_{W|YZ}(\tau''_m|y^m, z''_m) \\ w^{m+1} &\overset{(c)}{=} Q^0_{W|YZ}(\tau''_{m+1}|y^{m+1}, z''_m) \end{aligned} \tag{3-7}$$

(a) follows from (3-3), (b) from (3-5) and (c) is by (3-4).

Step 4 : Match?

Finally we use the construction of the structural function using (3-6). Then we can

[6] This is for $P^{m-1}(z'') \le P^m(z')$. Other cases can be shown similarly.

$$\begin{aligned} \overline{\tau} &\equiv F^a_{U|VZ}(r|v, z'') \\ &= F^a_{U|VZ}(h_a^{-1}(y^{m+1}, w^{m+1})|v, z'') \\ &= \begin{pmatrix} F^0_{W|YZ}(w^{m+1}|y^m, z'') & \text{if} \quad P^{m-1}(z'') \le P^m(z') \\ F^0_{W|YZ}(w^{m+1}|y^{m+1}, z') & \text{if} \quad P^m(z'') \le P^{m+1}(z') \end{pmatrix} \end{aligned} \tag{3-5'}$$

determine the direction of the response : we have from (3-2)[7]

$$\begin{aligned}
& h_a(y^m, u^*) - h_a(y^{m+1}, u^*) \\
= \;& h_a(y^m, Q^a_{U|VZ}(\tau''_m | P^m(z''_m), z''_m)) - h_a(y^{m+1}, Q^a_{U|VZ}(\tau''_{m+1} | P^{m+1}(z''_m), z''_m)) \\
= \;& Q^0_{W|YZ}(\tau''_m | y^m, z''_m) - Q^0_{W|YZ}(\tau''_{m+1} | y^{m+1}, z''_m) \\
= \;& Q^0_{W|YZ}(\tau''_m | y^m, z''_m) - Q^0_{W|YZ}(\overline{\tau} | y^m, z''_m) \\
& \begin{pmatrix} \leq 0 \text{ if PD} \\ \geq 0 \text{ if ND} \end{pmatrix}
\end{aligned}$$

the third equality is by (c) in (3-7). Then the inequality follows because $\tau''_m \leq \overline{\tau}$ (*PD) and $\tau''_m \geq \overline{\tau}$ (*ND), and the property of quantiles.

[7]Recall that this is the case for $P^{m-1}(z'') \leq P^m(z')$. The other case can be shown similarly.

Bibliography

[1] Abadie, A. (2002), "Bootstrap tests for distributional treatment effects in instrumental variable models," *Journal of the America Statistcal Association*, 97. 284-292.

[2] Abadie,A., J. Angrist, and G. Imbens (2002), "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earning," *Econometrica*, 70(1), 91-117.

[3] Ahn, H. and J. Powell (1993), "Semiparametric estimation of censored selection models with a nonparametric selection mechanism," *Journal of Econometrics*, 58, 3-29.

[4] Altonji, J. and R. Matzkin (2005), "Cross-section and panel data estimation for nonseparable models with endogenous regressors," *Econometrica*, 73, 1053-1102.

[5] Andrews, D., S. Berry and P. Jia (2004), "Confidence regions for parameters in discrete games with multiple equilibria, with an application to discount chain store," mimeo.

[6] Angrist, J. (1990), "Lifetime earnings and the Vietnam era draft lottery : evidence from the social security administrative records," *American Economic Review*, 80, 313-336.

[7] Angrist, J. (1997) "Conditional independence in sample selection models," *Economics letters*, 54, 103-112.

[8] Angrist, J. and G. Imbens (1995), "Two-stage least squares estimation of average causal effects in models with variable treatment intensity," *Journal of the American Statistical Association*, 90, 431-442.

[9] Amemiya, T. (1985), *Advanced Econometics,* Basil Blackwell.

[10] Athey, S. and P. Haile (2002), "Identification of standard auction models," *Econometrica*, 70, 2107-2140.

[11] Athey, S. and P. Haile (2005), "Nonparametric approaches to auctions," *Handbook of Econometrics.*

[12] Athey, S. (2002), "Monotone comparative statics under uncertainty," *Quarterly Journal of Economics*, 117, 187-223.

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

[14] Benkard, L. and S. Berry, (2006), "On the nonparametric identification of nonlinear simultaneous equations models : comment on Brown (1983) and Roehrig (1988)," *Econometrica*, 74(5), 1429-1440.

[15] Beresteanu, A., I. Molchanov, and F. Molinari (2008), "Sharp identification regions in games," mimeo.

[16] Berry, S. and E. Tamer (2007), "Identification in models of oligopoly entry," *Advances in Economics and Econometrics, Theory and Applications : 9th World Congress of the Econometric Society,* edited by Richard Blundell, Torsten and Whitney Newey. Cambridge.

[17] Bhattacharya, J., A. Shaikh, and E. Vytalcil (2008), "Treatment effects bounds under monotonicity assumptions : an application to Swan - Ganz Catheterization," *American Economic Review*, 98. 351-346.

[18] Bitler, M., J. Gelbach, and H. Hoynes (2006), "What mean impacts miss : distributional effects of welfare reform experiments," *Americam Economic Review*, 96. 988-1012.

[19] Blundell, R. and J. Powell (2003), "Endogeneity in nonparametric and semiparametric regression models," in M. Dewatripont, 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).

[20] Blundell, R. and J. Powell (2004), "Endogeneity in semiparametric binary response models," *Review of Economic Studies*, 71, 655-679.

[21] Blundell, R. and R. Smith (1989), "Estimation in a class of simultaneous equation limited dependent variable models," *Review of Economic Studies*, 56, 37-58.

[22] Blundell, R. and R. Smith (1994), "Coherency and estimation in simultaneous models with censored or qualitative dependent variables," *Journal of Econometrics*, 64, 355-373.

[23] Blundell, R, A. Gosling, H. Ichimura, and C. Meghir (2007), "Changes in the distribution of male and female wages accounting for employment composition using bounds," *Econometrica*, 75 (2), 323-363.

[24] Blundell, R. and J. Horowitz (2007), "A nonparametric test of exogeneity," *Review of Economic Studies,* 74(4), 1035-1058.

[25] Blundell, R., M. Browning, and I. Crawford (2007), "Best nonparametric bounds of consumer responses," cemmap working paper.

[26] Blundell, R. and T. Stoker (2005), "Heterogeneity and Aggregation," *Journal of Economic Literature*, 43(2), 347-391

[27] Blundell, R. and T. Stoker (2007), "Models of aggregate economic relationships that account for heterogeneity," *Handbook of Econometrics*, chapter 68, 4610-4666.

[28] Bontemps,C., T. Magnac, and E. Maurin (2008), "Set Identified Linear Models", mimeo.

[29] Breusch, T. (1986), "Hypothesis testing in unidentified models," *Review of Economic Studies*, 635-651.

[30] Brown, B. (1983), "The identification problem in systems of nonlinear in the variables," *Econometrica,* 51(1), 175-196.

[31] Browning, M. and J. Carro (2007), "Heterogeneity and micro modelling," *Advances in Economics and Econometrics, Theory and Applications : 9th World Congress of the Econometric Society,* edited by Richard Blundell , Torsten and Whitney Newey. Cambridge.

[32] Cameron, A. and P. Trivedi (1998), *Regression analysis of count data*, Econometric Society Monograph,Cambriege University Press.

[33] Chalak, K., S. Schennach, and H. White (2009), "Estimating average marginal effects in nonseparable structural systems," mimeo.

[34] Chernozhukov, V., I. Fernandez-Val, J. Hahn, and W. Newey (2009), "Identification and estimation of marginal effects in nonlinear panel models," cemmap working paper.

[35] Chernozhukov, V. and C. Hansen (2005), "An IV model of quantile treatment effects," *Econometrica*, 73(1), 245-261.

[36] Chernozhukov, V., Riggobon, and T. Stoker (2007), "Set identification with Tobit type regressors," mimeo.

[37] Chernozhukov, V., H. Hong, and E. Tamer, (2007), "Estimation and confidence regions for parameter sets in econometric models, " *Econometrica*, 75. 1243-1284.

[38] Chernozhukov, V., S. Lee, and A. Rosen (2008) "Inference on intersection bounds, " mimeo.

[39] Chesher, A. (1985), "Score tests for zero covariances in recursive linear models for grouped or censored data," *Journal of Econometrics*, 28, 291-305.

[40] Chesher, A. (2003), "Identification in nonseparable models," *Econometrica*, 71, 1405-1441.

[41] Chesher, A. (2005), "Nonparametric identification under discrete variation," *Econometrica*, 73(5), 1525-1550.

[42] Chesher, A. (2007), "Instrumental values," *Journal of Econometrics*, 139, 15-34 .

[43] Chesher, A. (2009), "Excess heterogeneity, endogeneity, and index restrictions," *Journal of Econometrics*, 152, 35-47.

[44] Chesher, A. (2010), "Instrumental variable models for discrete outcomes," *Econometrica*, 78 (2), 575-601.

[45] Cross, P. and C. Manski (2002), "Regressions, short and long," *Econometrica*, 70(1), 357-368.

[46] Das, M. (2005), "Instrumental variables estimation of nonparametric models with discrete endogenous regressors," *Journal of Econometrics,* 124. 335-361.

[47] Das, M., W. Newey, and F. Vella (2003), "Nonparametric Estimation of Sample Selection Models," *The Review of Economic Studies*, 70, 33-58.

[48] Dawid, A.P. (1979) "Conditional independence in statistical theory," *Journal of Royal Statistical Society Series B* 41, 1-31.

[49] D'Haultfaeuille, X and P. Fevrier (2007), "Identification and estimation of incentive problems : adverse selection," mimeo.

[50] Doksum, K. (1974) "Empirical probability plots and statistical inference for nonlinear models in the two sample case," *The annals of Statistics*, 2. 267-277.

[51] Fan, Y. and S. Park (2010), "Sharp bounds on the distribution of treatment effects and their statistical inference," *Econometric Theory*, 26, 931-951.

[52] Florens, J., J. Heckman, C. Meghir, and E. Vytlacil (2008), "Identification of treatment effects using control functions in models with continuous endogenous treatment and heterogeneous treatment effects," *Econometrica*, 76. 1191-1206.

[53] Firpo, S. (2007), "Efficient semiparametric estimation of quantile treatment effects," *Econometrica*, 75. 259-276.

[54] Firpo. S. and G. Ridder (2008), "Bounds on functionals of the distribution of treatment effects," IEPR working paper 08.09.

[55] Fisher, R. A. (1922), "On the mathematical foundations of theoretical statistics," *Transactions of the Royal Society*, series A, 222. 309-343.

[56] Galichon, A. and M. Henry (2009), "A test of non-identifying restrictions and confidence regions for partially identified parameters," *Journal of Econometrics*, 152(2), 186-196.

[57] Gourioux, C., J. Laffont, and A. Monfort (1980), "Coherency conditions in simultaneous linear equation models with endogenous switching regimes," *Econometrica*, 48, 675-695.

[58] Graham, B. and J. Powell (2008), "Semiparametric identification and estimation of correlated random coefficient models for panel data," mimeo.

[59] Guerre, E., I. Perrigne, and Q.Vuong (2009), "Nonparametric identification of risk Aversion in first-price auctions under exclusion restrictions," *Econometrica*, forthcoming.

[60] Hahn, J. and G. Ridder (2009), "Conditional moment restrictions and triangular simultaneous equations," mimeo.

[61] Han, A.(1987), "A non-parametric analysis of transformations," *Journal of Econometrics*, 35, 191-209.

[62] Han, A. and J. Hausman (1990), "Flexible parametric estimation of duration and competing risk models," *Journal of Applied Econometrics*, 5, 1-28.

[63] Hausman, J. (1983), "Specification and estimation of simultaneous equation models," chapter 7, *Handbook of Econometrics*, vol 1, 391-448.

[64] Haavelmo, T. (1944), "The Probability Approach in Econometrics," *Econometrica*, 12, Supplement iii-115

[65] Heckman, J. (1978), "Dummy endogenous variables in a simultaneous equation system," *Econometrica*, 46. 931-959.

[66] Heckman. J. (1979), "Sample selection bias as a specification error," *Econometrica*, 47. 153-161.

[67] Heckman, J. (1996), "Identification of causal effects using instrumental variables : Comment," *Journal of the American Statistical Association*, 91, 459-462.

[68] Heckman, J. (2000), "Micro data, heterogeneity, and the evaluation of public policy : Nobel lecture," *Journal of Political Economy*, 109(4), 673-747.

[69] Heckman, J. and R. Robb (1985), "Alternative methods for evaluating the impact of interventions : An overview," *Journal of Econometrics*, 30, 239-267.

[70] Heckman and Singer (1984), "A Model for Minimizing the Impact of Distributional Assumptions in Econometric Models for the Analysis of Duration Data," *Econometrica,* 52(2), 271-321.

[71] Heckman, J., J. Smith, and N. Clements (1997), "Making the most out of program evaluations and social experiments accounting for heterogeneity in program impacts," *Review of Economic Studies*, 64, 487-535.

[72] Heckman, J. and E. Vytlacil (1999), "Local instrumental variables and latent variable models for identifying and bounding treatment,"*Proceedings of the National Academy of Sciences*, 96. 4730-4734.

[73] Heckman, J., S. Urzua, and E. Vytalcil (2006), "Understanding instrumental variables in models with essential heterogeneity," *The Review of Economics and Statistics*, 88, 389-432.

[74] Heckman, J. and E. Vytalcil (2001), "Instrumental variables, selection models, and tight bounds on the average treatment effect," in M. Lechner and F. Pfeiffer (Eds.), *Econometric evaluation of labour market policies*, New York.

[75] Heckman, J. and E. Vytalcil (2005), "Structural equations, treatment effects, and econometric policy evaluation," *Econometrica*, 73, 669-738.

[76] Heckman,J. and S. Urzua (2009), "Comparing IV with structural models : what simple IV can and cannot identify," NBER working paper. no. 14760.

[77] Hoderlein, S. and E. Mammen (2007), "Identification of marginal effects in nonseparable models without monotonicity," *Econometrica*, 75. 1513-1518.

[78] Horowitz, J. and C. Manski (1998), "Censoring of outcomes and regressors due to survey nonresponse: Identification and estimation using weights and imputations," *Journal of Econometrics*, 84, 37-58.

[79] Horowitz, J. and C. Manski (2000), "Nonparametric analysis of randomized experiments with missing covariate and outcome data," *Journal of the American Statistical Association,* 95. 77-84.

[80] Hurwicz, L. (1950a) "Generalization of the concept of identification," in Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph 10. Wiley, New York.

[81] Hurwicz, L. (1950b) "Systems with Nonadditive Disturbances," in Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph 10. Wiley, New York.

[82] Imbens, G. and J. Angrist (1994), "Identification and estimation of local average treatment effects," *Econometrica*, 62, 467-476.

[83] Imbens, G. and C. Manski (2004), "Confidence intervals for partially identified parameters," *Econometrica*, 72. 1845-1857.

[84] Imbens, G. and W. Newey (2009), "Identification and estimation of triangular simultaneous equations models without additivity," *Econometrica*, 77(5),1481 - 1512.

[85] Galichon, A. and M. Henry (2009), "A test of non-identifying restrictions and confidence regions for partially identified parameters," *Journal of Econometrics*, forthcoming.

[86] Jovanovic, B. (1989), "Observable implications of models with multiple equilibria," *Econometrica*, 57. 1431-1437.

[87] Jun, S., J. Pinkse, and H. Xu (2010), "Tighter bounds in triangular system," mimeo.

[88] Kadane , J. and T. Anderson (1977), "A comment on the test of overidentifying restrictions," *Econometrica*, 45. 1027-1031.

[89] Kitagawa, T. (2009), "Identification region of the potential outcome distributions under instrument independence," mimeo.

[90] Kitagawa, T. (2010), "Testing for instrument independence in the selection model," mimeo.

[91] Khan, S, and E Tamer (2007), "Partial rank estimation of duration models with general forms of censoring ," *Journal of Econometrics,* 136(1), 251-280.

[92] Koenker, R. (2005), *Quantile regression*, Econometric Society Monographs, Cambriege University Press.

[93] Koopmans, T. (1949), "Identification Problems in Economic Model Construction," *Econometrica*, 17(2), 125-144

[94] Koopmans, T. and O. Reiersol (1950), "The identification of structural characteristics," *Annals of Mathematical Statistics*, 21, 165- 181.

[95] Lancaster, T. (1990), *The Econometric Analysis of Transition Data*, Econometric Society Monograph, Cambriege University Press.

[96] Lehman, E. (1974), *Nonparametric statistical methods based on ranks, San Francisco*, Holden-Day.

[97] Lewbel, A. (2000), "Semiparametric qualitative response model estimation with unknown heteroskedasticity or instrumental variables," *Journal of Econometrics*, 97, 145-177.

[98] Lewbel, A. (2007a), "Modeling heterogeneity," *Advances in Economics and Econometrics, Theory and Applications : 9th World Congress of the Econometric Society,* edited by Richard Blundell , Torsten and Whitney Newey. Cambridge.

[99] Lewbel, A. (2007b), "Coherency and completeness of structural models containing a dummy endogenous variable," *International Economic Review*, 48.1379-1391.

[100] Ma, L. and R. Koenker (2006), "Quantile regression methods for recursive structural equation models," *Journal of Econometrics*, 134(2), 471-506.

[101] Maddala, G. (1983), *Limited-dependent and qualitative variables in econometrics*, Cambridge : Cambridge University Press.

[102] Magnac, T, and E. Maurin (2007c), "Set identified linear models" working paper.

[103] Manski, C. (1990), "Nonparametric bounds on treatment effects," *American Economic Review*, 65(6), 1311-1334.

[104] Manski, C. (1995), *Identification problems in the social sciences*, Harvard Press.

[105] Manski, C. (1997), "Monotone treatment response," *Econometrica*, 65(6), 1311-1334.

[106] Manski, C. (2003), *Partial identification of probability distribution*, Springer-Verlag.

[107] Manski, C. (2005), *Social Choice with Partial Knowledge of Treatment Response*, The Econometric Institute Lectures, Princeton University Press.

[108] Manski, C. and D. McFadden (1983), *Structural analysis of discrete data with econometric applications*, Elsevier.

[109] Manski, C. and J. Pepper (2000), "Monotone instrumental variables : With an Application to the Returns to Schooling," *Econometrica*, 68(4), 997-1010.

[110] Manski, C. and E. Tamer (2003), "Inference on regressions with interval data on a regressor or outcome," *Econometrica*, 70(2), 519-546.

[111] Matzkin, R. (1991), "A nonparametric maximum rank correlation estimator," in Barnett, J. Powell, and G. Tauchen (eds.) Nonparametric and Semiparametric Methods in Econometrics and Statistics, Cambridge: Cambridge University Press

[112] Matzkin, R. (1992), "Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models," *Econometrica*, 60(2), 239-270.

[113] Matzkin, R. (1994), "Restrictions of Economic Theory in Nonparametric Methods," *Handbook of Econometrics.*

[114] Matzkin, R. (2003), "Nonparametric estimation of nonadditive random functions," *Econometrica*, 71, 1332-1375.

[115] Matzkin, R. (2004), "Unobservable instruments," mimeo.

[116] Matzkin, R. (2007a), "Nonparametric identification," *Handbook of Econometrics*, 5307-5368.

[117] Matzkin, R. (2007b), "Heterogeneous choice," *Advances in Economics and Econometrics, Theory and Applications : 9th World Congress of the Econometric Society,* edited by Richard Blundell , Torsten and Whitney Newey. Cambridge.

[118] Matzkin, R. (2008), " Identification of nonparametric simultaneous equations," *Econometrica*, 76, 945-978.

[119] Milgrom, P. and C. Shannon (1994), "Monotone comparative statics," *Econometrica*, 62. 157-180.

[120] Moffitt, R.A. (1996), "Identification of causal effects using instrumental variables : comment," *Journal of the American Statistical Association*, 91, 462-465.

[121] Mullahy, J. (1997), " Instrumental variables estimation of Poisson regression models : application to models of cigarette smoking behavior," *Review of Economics and Statistics.*

[122] Newey, W. and J. Powell (2003), "Instrumental variable estimation of nonparametric models," *Econometrica*, 71(5), 1565-1578.

[123] Newey, W., J. Powell and F. Vella (2003), "Nonparametric estimation of triangular simultaneous equations models," *Econometrica*, 67(3), 565-603.

[124] Pakes,A., J. Porter, K. Ho and J. Ishii (2006), "Moment inequalities and their application," mimeo.

[125] Perrigne, I. and Q. Vuong (2007), "Nonparametric identification of incentive regulation models," mimeo.

[126] Ridder, G. (1990), "The non-parametric identification of generalized accelerated failure-time models," *Review of Economic Studies,57,* 167-181 .

[127] Roehrig, C (1988), "Conditions for identification in nonparametric and parametric models," *Econometrica*, 56(2), 433-447.

[128] Romano, J. and A. Shaikh (2006), " Inference for identifiable parameters in partially identified econometric modes," mimeo.

[129] Rosen, A. (2006), "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities," cemmap working paper.

[130] Shaikh, A. and E. Vytlacil (2005), "Threshold crossing models and bounds on treatment effects : a nonparametric analysis," mimeo.

[131] Stoye, J. (2009), "More on confidence intervals," *Econometrica*, forthcoming

[132] Strotz, R. and H. Wold (1960), "Recursive vs. nonrecursive systems: an attempt at synthesis (part I of a triptych on causal chain systems)," *Econometrica*, 28, 417-427

[133] Tamer, E. (2003), "Incomplete simultaneous discrete response model with multiple equilibria," *Review of Economic Studies*, 70, 147 - 165.

[134] van den Berg, J. (2002), "Duration models: specification, identification, and multiple durations," *Handbook of Econometrics.*

[135] Vytlacil,E. and N. Yildiz (2007), "Dummy endogenous variables in weakly separable models," *Econometrica*, 75, 757-779.