Sharp Bounds on Heterogeneous Treatment Responses

Jinhyun Lee*
UCL

Job Market Paper
November 2009

Abstract

This paper aims to identifiy heterogeneous causal effects of an ordered discrete variable chosen by individuals, without relying on parametric assumptions. The proposed nonparametric restrictions have set identifying power of the value of the nonseparable structural function. When applied to an "unordered" binary endogenous variable, our model can be used to answer the question regarding whether a policy/treatment has a positive or negative impact on an individual when his/her ranking in the unobserved characteristic is given. The identification result is applied to examine the effects of Vietnam-era veteran status on earnings. The empirical findings show that when other observable characteristics are the same, military service had positive impacts for those with low earnings potential, while individuals with high earnings potential had negative impacts.

1 Introduction

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 targeted groups by the policy actually benefit from it.

*I am grateful to Andrew Chesher for valuable discussion and advice. Comments by Adam Rosen clarified many points in various parts of the paper. Special thanks are due to Toru Kitagawa for detailed discussion and comments. I also thank Yoichi Arai, Orazio Attanasio, Richard Blundell, Songnian Chen, Silvia Espinosa, Sergio Firpo, Hide Ichimura, Dennis Kristenssen, Thomas Lemieux, Arthur Lewbel, Chuck Manski, Brendon McConnell, Daniel McFadden, Costas Meghir, Tsuna Okumura, Christoph Rothe, Andrew Shephard, Aleksey Tetenov, and other participants in 2008 LAMES in Rio de Janeiro, Brazil, 2009 ENTER Jamboree in London, SLS at UCL, 2009 Royal Economic Society annual meeting, UK, ESRC/cemmap workshop on Quantile Regression in London, UK, 2009 SETA in Kyoto, Japan, and 2009 ESEM in Barcelona, Spain. All errors are my own.
Correspondence : jinhyun.lee@ucl.ac.uk.

By heterogeneous treatment responses we mean idiosyncratic treatment effect even after accounting for observed characteristics[1]. Several studies[2] allowed for individual heterogeneity in response. However, identification is achieved by integrating it out[3] in these studies. By identifying average responses, much information regarding the distributional consequences of a policy - heterogeneity in responses - would be lost. In this paper individual heterogeneity in responses is allowed by use of a non-additive structural relation and our model identifies heterogeneity by identifying partial differences of the structural relation. We demonstrate how "partial" information (the sign and not the size of treatment effect) regarding who benefits (individual heterogeneous response) can be recovered from data by using quantiles rather than averages.

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

$$\begin{aligned} W &= h(Y, X_1, U) \\ Y &= g(Z, X_2, V) \end{aligned}$$

where X_1 and X_2 are vectors 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 unobservable individual characteristics such as ability, preference, level of effort, or patience which are considered to be determinants of the outcome and the choice, and which may be correlated with each other. The structural relations may be derived from some optimization processes such as demand functions. If there is not a well-defined economic theory behind, 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 outcome when different values of X_1, X_2, Z, U and V are given. The key implication of the nonseparable functional form is that partial derivatives or partial differences are themselves stochastic objects

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

[2] The standard linear IV model cannot identify heterogeneus treatment effects. See Heckman and Navarro (2004) and Heckman and Urzua (2009). LATE by Imbens and Angrist (1994) can be used to identify heterogeneous treatment effects for the subpopulation characterized by "observed" covariates.

For identification under heterogeneous response 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), Heckman and Vytlacil (1999), Arellano and Bonhomme (2009), Gautier and Kitamura (2009) etc.

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

that have distributions[4].

Causal effects of a variable indicate the effects of the variable only, separated from other possible influences. Thus, when the outcome is determined by the above relationship, the causal effects of changing the value of Y from y^a to y^b on W other things being equal 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 for $X_1 = x$ and $U = u$. Individuals with different values for X_1 and U may have different values of $\Delta(y^a, y^b, x, u)$, so, the heterogeneity can be both observed and unobserved dimensional. If $\Delta(y^a, y^b, x, u)$ were identified, then individual heterogeneous response could be identified. This partial difference, $\Delta(y^a, y^b, x, u)$, is the major parameter of interest for identification analysis in this paper and we focus on heterogeneity in the unobserved characteristic ignoring the observed characteristics, x.

When Y is binary, the causal effects 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 approach, let W_{1i} denote the hypothetical outcome with a treatment and W_{0i} the hypothetical outcome without a treatment of the individual i whose observed and unobserved characteristics are x and u. If we can assume that W_{1i} and W_{0i} are generated by the structural relation then we can write

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

This way we map the problem in the potential outcomes approach into the structural approach[5]. 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, so the difference of the two for each individual is never observed. The identification problem in the structural approach (identification of the object on the

[4] If the structural function is linear, that is,

$$W = a + bY + cX_1 + U,$$

then partial derivative of this linear function with respect to Y is b. By imposing linearity, the structural feature of interest that captures the causal effect of Y is b, a constant. Thus, assuming a linear structural relation corresponds to assuming a "homogenous" response.

[5] 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 we are 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. A key characterization of the structural approach is that economic interpretation of certain functional of the structural relation is justified. One of the examples is partial derivatives or partial differences which can be interpreted as "ceteris paribus" impacts of a variable.

right) arises because independent variation in each coordinate of the structural relation is hard to achieve from the observed data due to possible endogeneity.[6]

The potential outcomes approach focuses on the distribution of the potential outcomes and does not utilize the information on the economic processes that generate the potential outcomes. Instead of the left hand side object, $W_{1i} - W_{0i}$, this paper focuses on identification of the right hand side object, $h(1, x, u) - h(0, x, u)$, by assuming the existence of economic processes and by imposing restrictions on such decision mechanisms. The proposed model does not assume continuity or differentiability and it is characterized by restrictions on the "modes" by which each variable - observables and unobservaebles - is related. As long as the existence of the relations and the restrictions imposed on them are "believed", our identification results by the structural approach can be informative. See the recent debate between Deaton (2009)[7] and Imbens (2009).

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 interpretations 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 from, inference of casual 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 the restrictions.

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. Chesher (2003,2007a) study 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. Thus, Chesher (2005) cannot be applied to a binary endogenous variable case. Moreover, Chesher (2005)'s

[6]Without selection or endogeneity problem, point identification is achievable. Suppose that there is no endogeneity problem, for example, suppose that data obtained from a randomized trial are available. Then we still cannot observe both counterfactuals, W_0 and W_1, thus, the left hand side object is not observed. However, if we believe that the counterfactuals are generated by a certain structural relation, then we can still "point" identify the right hand side object by applying Matzkin (2003) using quantiles.

[7]Deaton (2009) is based on the Keynes Lecture in Economics given at British Academy in October 2008. He addressed the issues in measuring effectiveness of any development policy. Randomized trials have been widely used and advocated as an alternative to various econometric methods which inherently contain many shortcomings. He emphasized the importance of modelling individual behavior in data analysis and criticized the movement to discard econometrics.

rank condition is hard to be satisfied with weak IV.

The contribution of this paper is to propose a model that relaxes Chesher (2005)'s "strong" rank condition and that imposes one more restriction on the structure such that it interval identifies partial differences of the structural relation. Our weak rank condition can be applied to examples such as regression discontinuity designs, a case with a binary endogenous variable or weak IV.

1.1 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[8]. When parametric assumptions are avoided, point identification is often not possible[9] with a discrete endogenous variable. In such cases one could aim to define a set where the parameter of interest is located. This partial identification idea, which was developed 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, 1994), Balke and Pearl (1997), Manski and Pepper(2000), Cross and Manski (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 inequality 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 (2007b)) etc.

The importance of the information regarding heterogeneous treatment effects was recognized earlier, but identification of them has not been studied until quite recently. Many parameters of interest are functionals of the distribution of individual treatment effects as Heckman, Smith, and Clements (1997) noted. Average treatment effects are found from the marginal distributions of the potential outcomes since the mean

[8]Nonparametric identification under endogeneity has been studied by several authors under different frameworks : one could specify the whole structural system as in Roehrig (1988), and Matzkin (2008), one could impose triangularity into the system as in Newey, Powell, and Vella (1999), Chesher (2003), or Imbens and Newey (2009), or one could use an incomplete single equation IV model as in Newey and Powell (2003), or Chernozhukov and Hansen (2005), or Chesher (2007b,2009), or one could use a conditional independence restriction to deal with endogeneity as in Altonji and Matzkin (2005) and Hoderlein and Mammen (2007).

[9]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.

is a linear operator. Other functionals such as quantiles require the knowledge of the distribution of the individual treatment effects[10].

Some information regarding heterogeneity can be recovered by using quantiles. One particular object that has been the focus of research is the quantile treatment effect (QTE) defined by Lehman (1974) and Doksum (1974). The QTE is measured by the difference of quantiles of the marginal distributions of the potential outcomes. One could recover the marginal distributions of the potential outcomes to find the QTE. Imbens and Rubin (1997) and Abadie (2002) have results on identification of the marginal distributions of potential outcomes. Or one could model the QTE as a linear function. See Abadie, Angrist, and Imbens (2002) under the LATE-type assumption, Firpo (2007) under the matching assumption, and Frolich and Melly (2009) under the regression discontinuity design. Abadie, Angrist, and Imbens (2002) dealt with the selection problem using the LATE type assumption, thus, the estimate from their method can be interpreted as a "causal" effect on a particular point of the distribution, not on the individuals. The QTE may measure the causal effect of a policy on a particular point of a distribution. However, the QTE should not be interpreted as the causal effects of the treatment on the individuals since the individual ranked in the particular point in the marginal distributions of the potential outcomes may not be the same individuals.

Another approach has been taken to recover heterogeneity in treatment effects by identifying the distribution of $W_1 - W_0$ directly[11]. Heckman, Smith and Clements (1997) use the Hoeffding-Frechet bounds, and Fan and Park (2006) 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. In our structural approach we do not have to recover the marginal distribution of the potential outcomes, and the heterogeneity in response can be modeled by a non-additive structural relation. We focus on the partial difference of the structural relation with respect to the binary variable. Our identification results provide a way to recover heterogeneity in responses among the observationally same individuals.

Chernozhukov and Hansen (2005) adopt the structural approach as here. Chernozhukov and Hansen (2005)'s moment condition based on their IV-QR model provides a way to estimate $h(1, x, u)$ and $h(0, x, u)$ separately, as u-quantile functions conditional on the IV under uniform normalization of U. Whether the difference obtained can be interpreted as causal effect on the "individual" depends on whether

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

[11] 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".

the rank similarity restriction holds. More detailed discussion on identification of heterogeneous treatment effects using quantiles can be found in section 4.4.

Jun, Pinkse, and Xu (2009) report tighter bounds when a different rank condition from Chesher (2005)'s is used, while other restrictions in Chesher (2005) are adopted. If there is only one pair of values that satisfies their rank condition, the bounds are essentially the same as Chesher (2005) bounds. When there are more such pairs of instrumental values, then their identification strategy that utilizes the Dynkin system argument lead to tighter bounds than Chesher (2005)'s when their rank condition is stronger than Chesher (2005)'s rank condition. Similar monotonicity restrictions to ours have been adopted by Manski and Pepper (2000) and Bhattacharya, Shaikh, and Vytlacil (2008). More discussion on these papers and how they are different from this paper can be found in section 4.

The remaining part is organized as follows. Section 2 introduces the model for "ordered" discrete endogenous variables and contains the main results on identification. Section 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. We 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. Section 6 concludes.

2 Local Selection and Response Match (LSRM) model - $\mathcal{M}^{LSRM}$

2.1 The quantile - based control function approach

While the control function approach can be described as conditioning on $V = v$,[12] the quantile-based control function approach can be described as conditioning on the v - quantile of Y given Z to correct for endogeneity[13]. We first introduce how the

[12] For the sample selection models - both parametric and non/semi - parametric models - the propensity score function plays the role of control function (see Heckman and Robb (1986), Das, Newey, and Vella (2001), and Heckman and Navarro (2004)). In the parametric nonlinear Tobit model as in Smith and Blundell (1986), the first stage residual plays the role of control function. In Newey, Powell and Vella (1999)'s nonparametric function with additively separable error under mean independence restriction (with (2) being $g(Z, V) = \theta(Z) + V$), the residual obtained from the first stage regression of Y on Z, $v = y - \theta(z)$), plays the role of control function. In Imbens and Newey (2009) with non-additive error under full independence of unobservables and an IV, the distribution of endogenous variable given covariates, $v = F_{Y|Z}(y|z)$, plays the role of control function.

[13] Imbens and Newey (2009) showed that the two control function approaches are equivalent when the endogenous variable is continuous and U is a scalar. 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.

quantile based control function approach by Chesher (2003)[14] works and why it fails to achieve point identification when the endogenous variable is discrete.

Suppose that the endogenous variable, Y, is continuous. In the following triangular[15] system ignoring other covariartes than Z

$$\begin{aligned} W &= h(Y,U) \quad (1) \\ Y &= g(Z,V). \quad (2) \end{aligned}$$

It is impossible to identify the whole structure, $\{h, F_{U|Z}\}$ in (1) even without endogeneity (Lemma in Matzkin (2003)) due to nonseparability : normalization of $F_{U|Z}$ is required for identification of the structural function, h[16]. Instead of recovering the whole structure, we focus on the identification of the structural function evaluated at a specific quantile of the unobservable ($u^* = Q_{U|VZ}(\tau_U|v,z)$) - we need not know what the exact value of the unobservable at that quantile - we call this function the "structural quantile function (SQE)".

To be able to identify a structural function we need to observe independent variation in each coordinate of the function. When the unobserved heterogeneity, U, is not independent of Y, we need to guarantee whether we can separately cause variation in the endogenous variable other things being fixed[17]. Independent variation in Y can be found by fixing U at $u^* = Q_{U|VZ}(\tau_U|v,z)$ and by changing the value of Z for given value of $V = v$. Note that when Y is continuously varying, for given value of Z there is a one-to-one relationship between Y and V following the eq. (2). This implies that conditioning on Y and Z can control the value of V by the invertibility of the function g. When Y is discrete, the function g is not invertible any more, and thus, knowing the values of Y and Z does not fix the value of V.

The value of the structural function evaluated at (y, u^*) can be identified by the quantile of the conditional distribution of W given Y and Z (Chesher (2003)), rather than the quantile of the conditional distribution of W given Y only (Matzkin (2003)) as follows :

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

[14] Chesher (2003) originally considers a recursive triangular equations system advocated by Strotz and Wold (1960). See also Koenker (2005)'s exposition of Chesher (2003)'s approach in chapter 8.

[15] "Triangular" in the sense that Y is the determinant of W, while W is not a determinant of Y. By imposing triangularity we exclude the possibility that the choice is affected by the (expectation of) outcome.

[16] $g(z,v)$ is identified by $Q_{Y|Z}(v|z)$.

[17] Under triangularity the covariation between U and Y is caused by the covariation between U and V. Thus, the identification problem in the triangular system is whether we can generate variation in Y fixing V.

As the value of Z changes from z^a to z^b, the value of Y changes from y^a to y^b for given value of $V = v$ following the structural equation in (2), while the value of u^* is fixed because Z is independent of U and V.

When Y is continuously varying, once we identify the values of the structural function, the partial difference is also found by[18]

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

See <Figure 1>

With discrete Y, the equality in (3) fails to hold because the event of $\{Y = y$ and $Z = z\}$ corresponds to a set of values of V, rather than a point, which implies that u^* would lie in a set, rather than being fixed at a point under endogeneity. Causing variation in Y by varying Z no longer generates an exogenous variation in Y when Y is discrete. However, this still restricts the possible range where the value of the structural function can lie if some more restrictions are imposed as Chesher (2005) and the LSRM model. Note that the loss of point identifying power with discrete regressors is due to the presence of endogeneity. If U and V are not correlated, then we can point identify the structural function by quantiles of $F_{W|Y}$ without the help of Z. See <Figure 2>.

2.2 Restrictions of the model $\mathcal{M}^{LSRM}$

We introduce a model[19] 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, $\mathcal{M}^{LSRM}$, is defined as the set of all the structures that satisfy the restrictions

Definition 1 : 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

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

[18] Note that the value of the structural function $h(y, u^*)$ are found by fixing $u^* = Q_{U|VZ}(\tau_U|v, z)$ 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 $v-$ quantile of Y given Z, $Q_{Y|Z}(v|z)$, would cover the whole points in the support of Y by varying Z.

[19] We adopt this definition of a model as a set of structures satisfying the restrictions following Koopmans and Reiersol (1950).

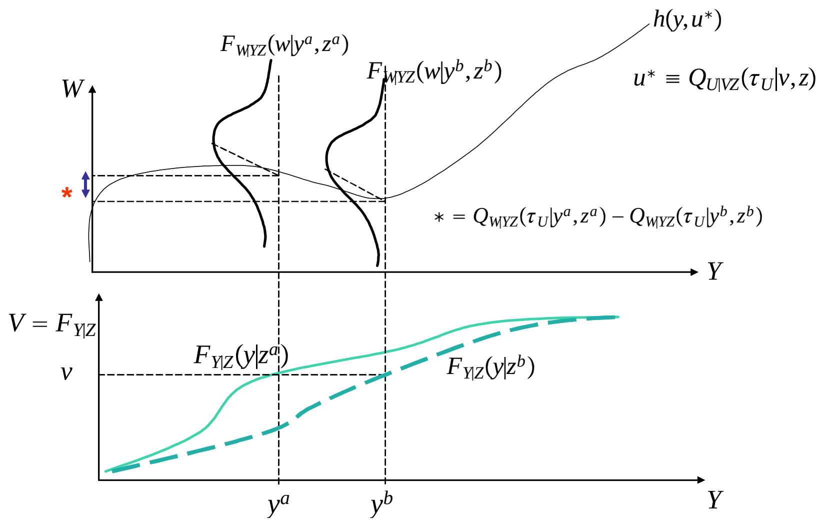


Figure 1: The line, $h(y, u^*)$, is drawn by fixing the value of U at u^*. Thus, the causal effect of changing Y from y^a to y^b should be measured by $*$ since on the line, $h(y, u^*)$, other thing (u^*) is fixed. However, this cannot be identified by Matzkin (2003)'s idea using quantiles of $F_{W|Y}$ since whenever we change the value of Y, 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 (2) under the triangularity allows to control the source of endogeneity V when Y is continuous. Continuity of Y and monotonicity of the strucutral 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 $y^a = Q_{Y|Z}(v|z^a)$ and $y^b = Q_{Y|Z}(v|z^b)$ then conditional distribution of W given Y and Z, $F_{W|YZ}$, rather than $F_{W|Y}$ will deliver the 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 Lee (2009a).

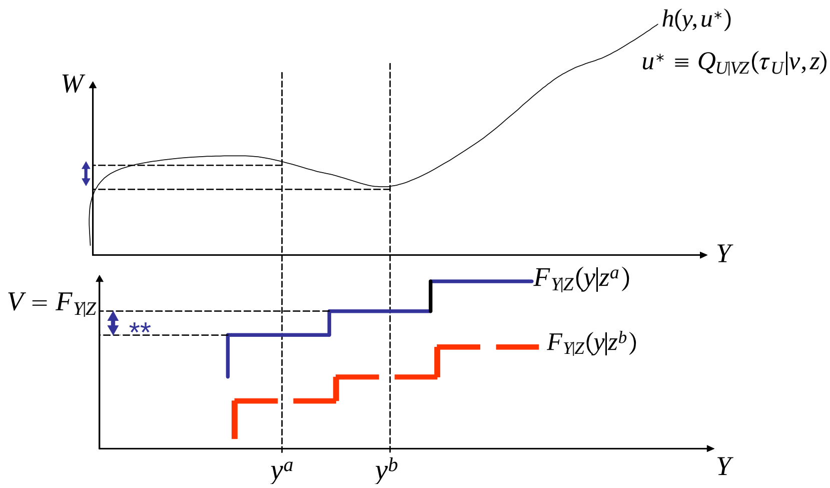


Figure 2: Failure of point identification by the QCFA with discrete Y : knowing the values of Y and Z, for example, $Y = y^a$ and $Z = z^a$, corresponds to a range of values of V of (**). Conditioning on Y and Z does not guarantee that V is fixed at a particular point, which implies that all can be infered by conditioning on Y and Z is that u^* lies in a certain interval if U and V are correlated.

Definition 2 : The latent variates are jointly continuously distributed and V is normalized uniformly distributed on $(0,1)$ independent of Z.

Restriction A-EX imposes conditions on the structural functions that determine W and Y, and will be maintained throughout.

Restriction A-EX (Exclusion)[20]

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

The function h is weakly monotonic[21] with respect to variation in scalar[22] U, normalized caglad, and nondecreasing. The function g evaluated at $Z=z$, $g(z,v)$ is $Q_{Y|Z}(v|z)$, the conditional $v-$quantile function of Y given $Z=z$. See <Figure 3>

The structure we aim to recover from data is defined as $S \equiv \{h, g, F_{UV|Z}\}$. Since the function g is recovered by $g(z,v) = Q_{Y|Z}(v|z)$, we focus on the structural relation h. We also consider the conditional distribution of U given V and Z, $F_{U|VZ}$, rather than joint distribution, $F_{UV|Z}$, since $F_{UV|Z} = F_{U|VZ}F_{V|Z}$ can be recovered once we find $F_{U|VZ}$ as we normalize $F_{V|Z}$ to be uniform over $(0,1)$.

The value $y^m, m \in \{2,...,M-1\}$, is an interior point of support of the distribution of Y. The term u^* is a value of U defined as

$$u^* \equiv Q_{U|VZ}(\tau_U|v,z),$$

[20] Triangularity assumption enables us to avoid the issue of coherency that is possible due to discrete endogenous variables when outcome is discrete (See Gourieroux, Laffont, and Monfort (1980), Blundell and Smith (1994), Tamer (2003), and Lewbel (2007)).

[21] 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. See monotone comparative statics by Milgrom and Shannon (1994) and also monotone comparative statics under uncertainty by Athey (2002).

[22] 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 and the usual monotonicity restriction may not be required (See Altonji and Matzkin (2005), Hoderlein and Mammen (2007), Imbens and Newey (2009), and Chalak, Schennach, and White (2008), and Chernozhukov, Fernandez-Val, and Newey (2009) for identification analysis without monotonicity), but what can be identified is objects with the heterogeneity in response averaged out. On the other hand, the quantile approach can be adopted to recover heterogeneous treatment response if a scalar (index) unobserved heterogeneity is assumed, however, this may be considered to be restrictive since many examples such as models with measurement error cannot be dealt with. See Chesher (2008) for examples where the unobserved elements cannot be collapsed into a scalar index.

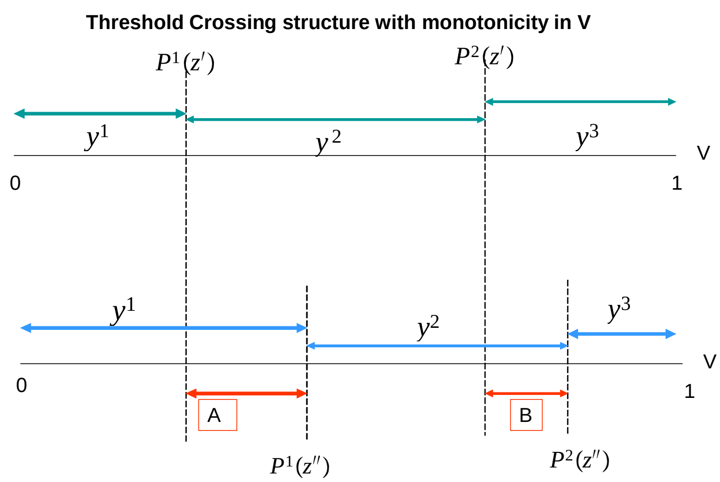


Figure 3: V is normalized to uniform (0,1). Under the threshold crossing structure with monotonicity of g in v, the value of Y is determined by the value of V and the cutoff points are determined by Z. Note that by construction and independence of V and Z, V is not affected by Z but for given value of V, Z will determnine the values of Y by affecting the cutoff points. As we change the value of Z from z' to z'' there are groups of people with certain $V-$ characteristics who shift their choice of Y. Those whose V-ranking is in "A" change the value of Y from y^2 to y^1 as the value of Z changes from z' to z'', while those whose V-ranking is in "B" change the value of Y from y^3 to y^2 as the value of Z changes from z' to z''. The causal interpretation is justified only for these groups.

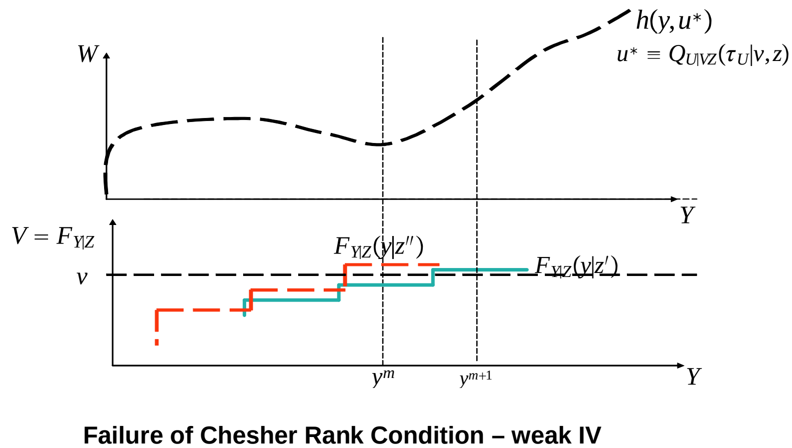


Figure 4: Failure of Chesher (2005) strong rank condition : weak IV

where $\tau_U, v \in (0,1)$ and z lies in a set of instrumental values of Z to be specified shortly.

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

The common support restriction is imposed for sharpness.

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

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

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

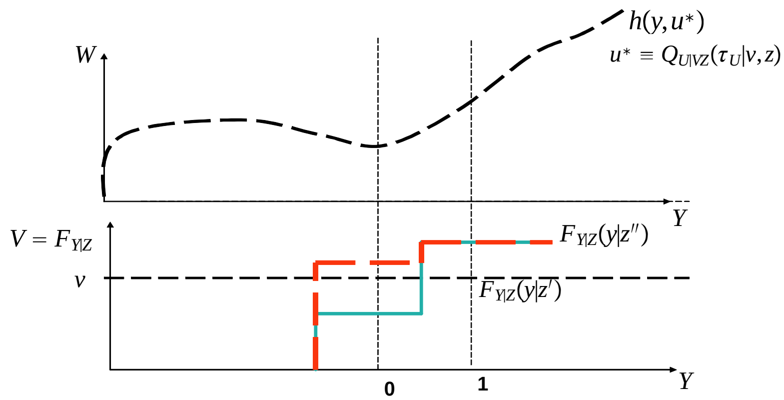
Chesher (2005)'s rank condition is that there exist values of Z, z'_m, and z''_m such that

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

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 with $M = 1$, a binary case or when the instrument is weak. See <Figure 4> and <Figure 5>.

Restriction C-FI (Conditional Full Independence)[23] : $F_{U|VZ}(u|v,z) = F_{U|V}(u|v)$ conditional on v, for all $z \in \{z'_m, z''_m\}$.

[23] Restriction C-FI does not have to hold for all values of z in the support of Z, but it should hold for the values of Z that satisfy Restriction RC. We call those values that satisfy Restriction



Failure of Chesher Rank Condition – Binary endogenous variable

Figure 5: Failure of Chesher (2005) strong rank condition : When the number of points in the support of Y is two, binary endogenous variable, Chesher (2005)'s strong rank condition fails.

Define $V \equiv (P^{m-1}(z), P^{m+1}(z)]$[24].

Restriction LSRM (Local Selection Response Match)[25] **:** The conditional distribution of U given V, $F_{U|VZ}(u^*|v,z)$ is monotonic in $v \in V$. We impose a further restriction on h and $F_{U|VZ}(u|v,z)$ over this range of v : if $F_{U|VZ}(u^*|v,z)$ is weakly nonincreasing in v, then $h(y^{m+1},u^*) \geq h(y^m,u^*)$, and if $F_{U|VZ}(u^*|v,z)$ is weakly nondecreasing in v then $h(y^{m+1},u^*) \leq h(y^m,u^*)$, for $m \in \{1,2,...,M\}$ for $u^* \equiv Q_{U|VZ}(\tau_U|v,z)$. See <Figure 6>

Notation : The case in which $F_{U|VZ}(u^*|v,z)$ is nonincreasing in v is called **PS** (Positive Selection) and the other case, **NS** (Negative Selection) for ease of exposition. 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).

In the returns to schooling example Restriction LSRM allows that ability can be correlated both positively and negatively, but if ability is positively correlated schooling, then the return to schooling should not be negative and vice versa. In the job training example, Restriction LSRM allows that the unobserved heterogeneity (for example, ability) that determines wage can be correlated with both positively

RC as "instrumental values". Although we indicate Z as IV, what is required in identification is the existence of at least two distinct values of Z that satisfy the rank condition. This is the reason why instrumental "values" are more exact expression, rather than instrumental variables (See Chesher (2007a)).

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

[25] Comments by Songnian Chen and Adam Rosen clarified the statement of the restriction and its interpretation.

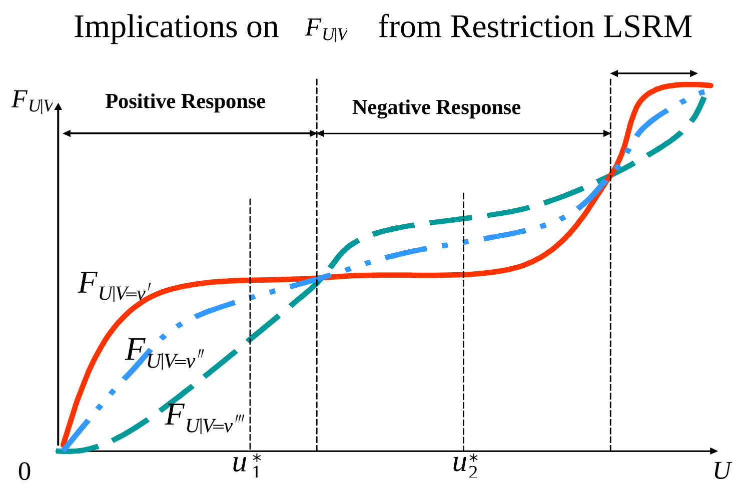


Figure 6: "Local" nature of Restriction LSRM : The information on endogeneity is contained in $F_{U|V}$ - if Y is exogenous, then $F_{U|V}$ is not affected by the values of V. Monotonicity of $F_{U|V}(u^*|v)$ does not have to be global, all is required is monotonicity in some small neighborhood u^*. For $v' \leq v'' \leq v'''$, $F_{U|V}(u_1^*|v)$ is decreasing in v, while $F_{U|V}(u_2^*|v)$ is increasing in $v \in V$.

and negatively with the job training decision (or the unobserved heterogeneity that determines the participation decision such as motivation), but if more capable worker decides to participate, then the treatment effects should not be negative. In the labour supply decision example Restriction LSRM allows the unobserved heterogeneity that determines labour supply decision (such as taste for career) can be positively or negatively correlated with the unobserved heterogeneity that determines the fertility decision(e.g. taste for children), but if females who prefer spending their time for their career are less likely to prefer having more children, then the impact of fertility on labour supply decision should not be positive. In veteran status and earnings example Restriction LSRM allows that the unobserved heterogeneity that determines earnings such as ability can be positively or negatively correlated with the unobserved heterogeneity that determines whether to join the army or not such as risk attitude, but if they are positively correlated, then the impacts of veteran status should not be negative.

If there is no endogeneity the distribution of U given V is independent of V under the triangular structure, thus, $u^* \equiv Q_{U|VZ}(\tau_U|v,z)$ would not vary even though the endogenous variable may be discrete. Therefore, point identification is achieved even if the explanatory variable is discrete.

2.3 The bounds on the values of the structural relation

We have the following interval identification for $h(y^m, u^*)$, where $u^* \equiv Q_{U|VZ}(\tau_U|v,z)$ for $m \in \{1,2,...,M-1\}$. For $m = M$, the bound in Theorem 1 is not applied[26].

Theorem 1 *Under Restriction A-EX,FI,RC,and LSRM, there are the inequalities for* $m \in \{1,2,...,M-1\}$ *and* $\tau \equiv (\tau_U, v)$ *for* $u^* \equiv Q_{U|VZ}(\tau_U|v,z)$

$$\begin{aligned}
q_m^L(\tau, y^m, \overline{z}_m) &\leq h(y^m, u^*) \leq q_m^U(\tau, y^m, \overline{z}_m) \\
\text{where } z &\in \overline{z}_m = \{z'_m, z''_m\},\ \tau \equiv (\tau_U, v), \\
q_m^L(\tau, y^m, \overline{z}_m) &= \min\{Q_{W|YZ}(\tau_U|y^m, z'_m), Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)\}, \\
q_m^U(\tau, y^m, \overline{z}_m) &= \max\{Q_{W|YZ}(\tau_U|y^m, z'_m), Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)\},.
\end{aligned}$$

The interval is not empty.

Proof. See 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 $P^m(z'_m) \leq v \leq P^m(z''_m)$. 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.

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

Corollary 2 *Under Restriction A-EX, if U and V are independent, the value of $h(y^m, u^*)$, $u^* = Q_{U|VZ}(\tau_U|v, z)$ is* point *identified by $Q_{W|Y}(\tau_U|y^m)$ and $Q_{W|YZ}(\tau_U|y^m, z) = Q_{W|Y}(\tau_U|y^m)$, for all $Z \in SUPP(Z)$.*

Proof. See Appendix. ■

2.4 Sharpness

Suppose a set identifies the value of the structural feature. Then all distinct "*admitted*[27]" structures that are "*observationally equivalent*[28]" to the true structure should produce a value of the structural feature that is contained in the identified set. All such structures that generate the point in the set are indistinguishable by data. See <Figure 7 >.

If the identified set is *not* sharp, 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. Sharpness guarantees that every point in the identified set is a possible candidate for the true value of the structural feature of interest. Distinct structures may have produced different points in the identified set, 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[29]. See <Figure 8 >.

Sharpness is assumed in many studies[30] on inference in partially identified models. If sharpness were not guaranteed, then the confidence set of the identified set would not be correct since they would contain some parts of the identified set where the true value of the structural feature never lies in. See <Figure 9 >.

Theorem 3 *Under Restrictions CSUPP, A-EX,C-FI,RC,and LSRM, the bounds $I(\tau, y^m, \overline{z}) \equiv [q_m^L(\tau, y^m, \overline{z}_m), q_m^U(\tau, y^m, \overline{z}_m)]$, specified in Theorem 1 for each $m = 1, 2, .., M-1$ and for some $\tau \equiv (\tau_U, v)$, are sharp.*

[27] That is, those structures thay satisfy all the restrictions.

[28] Observationall equivalent structures are those that are indistinguishable from the distribution we observe by data.

[29] Let S_0 be the true structure that generates the distribution of observables available to us, $F^0_{Y|X}$. Denote $\Omega_0 = \{S : F^S_{Y|X} = F^0_{Y|X}\}$ the set of observationally equivalent structures. The model, $\mathcal{M}$ is defined to be the set of the structures that satisfy the restrictions. The model $\mathcal{M}$ **set identifies** the structural feature, $\theta(S_0)$, if we can find a set defined by the distribution function such that for any structure that is admitted and observationally equivalnet to the true structure, the structural feature of such structure lies in the set.

The identified set is called **sharp** if for any value in the identified set $\mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X})$, there exists an admissible structure by $\mathcal{M}$, which is observationally equivalent to S_0.

[30] For example, Galichon and Henry (2009) assumes "internal consistency" for their test to be constructed. "Internal consistency" corresponds to sharpness.

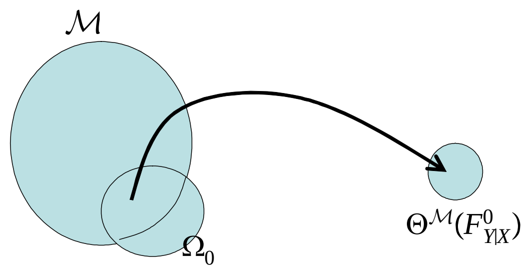


Figure 7: Set identification : the value of the structural feature ($\theta(S)$) generated by any structure that is admitted by the model and observationally equaivalent to the true structure should lie in the set $\mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X})$. Note that there can be some parts of the set, $\mathbf{\Theta}^{\mathcal{M}}(F^0_{Y|X})$, where $\theta(S)$ never lies. Sharpness of the 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.

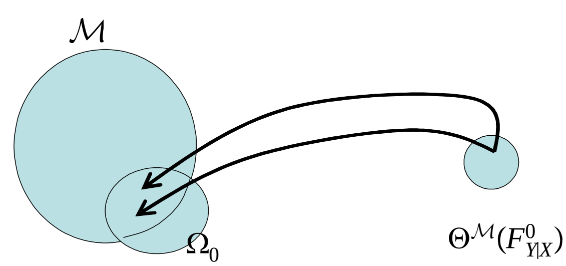


Figure 8: Sharpness : showing sharpness involves showing that for each point in the set there exists *at least* one structure that is admitted and observationally equivalent to the true structure.

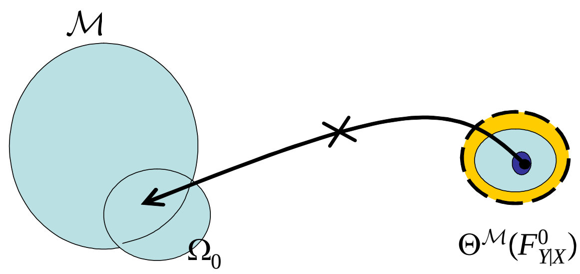


Figure 9: If the identified set is **not** sharp, the identified set may contain some parts (the dark circle inside the identified set) that the true value of the structural feature does not lie in. If such parts are with positive measure, then the confidence set (the dotted circle outside the identified set) constructed would be misleading.

Proof. See Appendix. ■

To show sharpness of the bounds, $I(\tau, y^m, \overline{z}_m)$, for given $\tau \equiv (\tau_U, v)$ and m we need to show that for every point in the interval there exists at least an observationally equivalent structure to S^0 (that generates $F^0_{W|YZ}$), which is admitted by $\mathcal{M}^{LSRM}$ and generates the value of the structural function. To show the existence of such a structure we construct a distribution of the unobservables $(F^a_{U|VZ}(u|v,z))$ and a structural function (h_a) that crosses an arbitrary point in the identified interval[31] $(w^* \in I(\tau, y^m, \overline{z}_m))$ in such a way that they satisfy the restrictions of the model and that the assumed function (h_a) combined with the constructed distribution $(F^a_{U|VZ}(u|v,z)))$ generates the same distribution of observables as $F^0_{W|YZ}$, which is what we observe[32].

[31] The value w^* takes should be an element of the support of W, which can be countable.

[32] In contrast with sharpness proofs in the potential outcomes approach (see for example, Firpo and Ridder (2008), Heckman and Vytlacil (2001)), to show whether the points in the identified set are consistent with the model we need to construct the underlying structural relation and the distribution of the unobservables since the model is characterized by the restrictions imposed on them.

Consider a binary endogenous variable case. In the potential outcomes framework $F_{W_1W_0|X}$ is the hidden data generating process to be recovered from the observed data, $F_{W|X}$, thus, sharpness proof involves the construction of $F_{W_1W_0|X}$ using $F_{W|X}$ that is consistent with the model and data. In the structural approach the underlying economic data generating process is $\{h(1,u^*), h(0,u^*), F_{U|V}\}$

2.4.1 Sharpness of Chesher (2005) bounds

Note that our sharpness proof can be used to show sharpness of Chesher (2005) bounds since once chesher (2005)'s strong rank condition is satisfied, our structure that satisfies Restriction LSRM is a special case of the structures admitted by Chesher (2005) model - we have already shown the existence of at least one admitted structure that is observationally equivalent to the true structure. See Appendix.

2.5 Many instrumental values

If there are many pairs of values of Z that satisfy Restriction RC, then the bounds should be found by taking intersection of the bounds found for each pair. Formally this can be stated as :

Corollary 4 *Under Restriction A-EX,FI,RC,and LSRM, there are the inequalities for* $m \in \{1, 2, ..., M-1\}$ *and* $\tau \equiv (\tau_U, v)$

$$\begin{aligned}
Q_m^L(\tau, y^m, \mathcal{Z}_m) &\leq h(y^m, u^*) \leq Q_m^U(\tau, y^m, \mathcal{Z}_m) \\
\text{where } \tau &\equiv (\tau_U, v) \\
u^* &\equiv Q_{U|VZ}(\tau_U | v, z) \\
\mathcal{Z}_m &= \{\overline{z}_m : P^m(z'_m) \leq v \leq P^m(z''_m), \text{ with } \overline{z}_m = \{z'_m, z''_m\}\} \\
Q_m^L(\tau, y^m, \mathcal{Z}_m) &= \max_{\overline{z}_m} \ q_m^L(\tau, y^m, \overline{z}_m), \ \overline{z}_m \in \mathcal{Z}_m \\
Q_m^U(\tau, y^m, \mathcal{Z}_m) &= \min_{\overline{z}_m} \ q_m^U(\tau, y^m, \overline{z}_m), \ \overline{z}_m \in \mathcal{Z}_m \\
q_m^L(\tau, y^m, \overline{z}_m) &= \min\{Q_{W|YZ}(\tau_U | y^m, z'_m), Q_{W|YZ}(\tau_U | y^{m+1}, z''_m)\} \\
q_m^U(\tau, y^m, \overline{z}_m) &= \max\{Q_{W|YZ}(\tau_U | y^m, z'_m), Q_{W|YZ}(\tau_U | y^{m+1}, z''_m)\}
\end{aligned}$$

This intersection interval can be empty.

Define $Q_m \equiv [Q_m^L(\tau, y^m, \mathcal{Z}_m), Q_m^U(\tau, y^m, \mathcal{Z}_m)]$ and $q_m = [q_m^L(\tau, y^m, \overline{z}_m), q_m^U(\tau, y^m, \overline{z}_m)]$. Then the Corollary can be written as $Q_m = \cap_{\overline{z}_m} q_m$. By taking intersection of each bound defined by each pair of instrumental values satisfying Restriction RC we have smaller bounds when we have more pairs of instrumental values.

2.6 Testability of Restriction LSRM

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

for given u^* in the triangular structure when ignoring other covariates, X. Therefore, the sharpness proof involves the construction of $\{h(1, u^*), h(0, u^*), F_{U|V}\}$ using $F_{W|YX}$ and consistency with the model and the data should be shown.

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. Lee (2009b) generalizes Koopmans and Reiersol (1950)'s principle into the set identifying models and show that the identified set of an observationally more restrictive model should be included by that of a less restrictive model.[33]

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 $\mathcal{M}^{LSRM}$. If the strong rank condition is satisfied, $\mathcal{M}^{LSRM}$ is contained by Chesher (2005) model, thus, $\mathcal{M}^{LSRM}$ is observationally more restrictive. Lee (2009b) implies that LSRM bound should be equal to or smaller than Chesher (2005) bound.

LSRM restriction is "not directly testable[34]", in other words, LSRM restriction does 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[35]", thus, once the strong rank condition is satisfied we can say that the model, $\mathcal{M}^{LSRM}$ is observationally more restrictive than the model in Chesher (2005). In this case, the identified interval by $\mathcal{M}^{LSRM}$ should be included by the identified interval by Chesher (2005) if restriction LSRM is satisfied. Therefore, if the bounds constructed by Chesher (2005) is smaller than the bounds formed by LSRM model, then this implies that the LSRM restriction is not the right description of the true underlying structure that generated the data.

We cannot "confirm[36]" whether the LSRM holds, but we can "refute" the restriction by comparing $Q_{W|YZ}(\tau_U|y^m, z'')$ with $Q_{W|YZ}(\tau_U|y^{m+1}, z'')$.

[33] Suppose that a model, $\mathcal{M}^1$, identifies a structural feature, $\theta(S)$, by a set $\mathbf{\Theta}^1(F^S_{Y|X})$, for $S \in \mathcal{M}^1$, and another model, $\mathcal{M}^2$, identifies the same structural feature, $\theta(S)$, by $\mathbf{\Theta}^2(F^S_{Y|X})$, for $S \in \mathcal{M}^2$.

Theorem (Lee (2009b) If $\mathcal{M}^1 \subseteq \mathcal{M}^2$, then $\mathbf{\Theta}^1(F^S_{Y|X}) \subseteq \mathbf{\Theta}^2(F^S_{Y|X})$, for $\forall S \in \mathcal{M}^1 \cap \Omega_0$.

[34] Note that LSRM is a restriction imposed on the structural relation and the distribution of the unobservables. The restrictions imposed on the structure are not testable unless they have implications on the distribution of the observables.

[35] Data are informative about whether the rank condition is satisfied since the rank condtion is about the conditional distribution of Y on Z.

[36] Lee (2009b) also defines the framework of testability under the general structural setup allowing for partial identification.

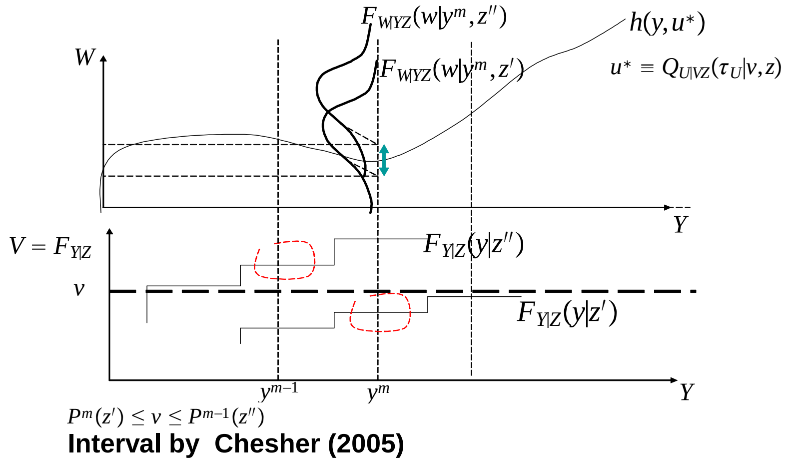


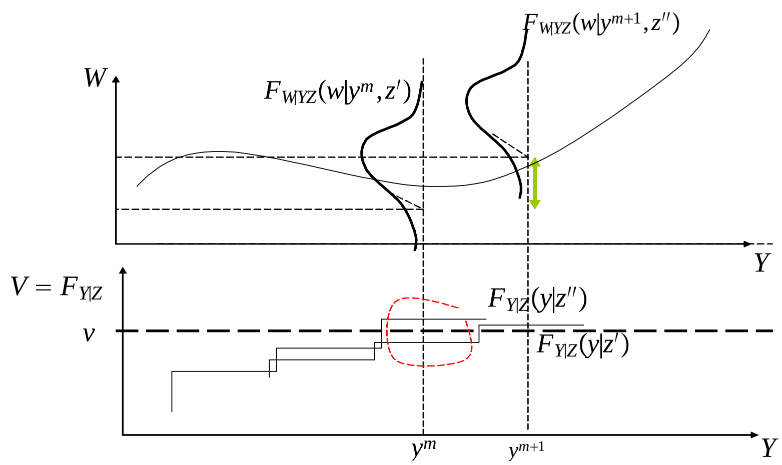
Figure 10: Chesher (2005) strong rank condition is that there exist values of Z, z'_m and z''_m such that $P^m(z'_m) \leq v \leq P^{m-1}(z''_m)$: the arrow indicates the Chesher bound. Note that if Chesher (2005)'s strong rank condition holds our rank condition always holds since $P^m(z') \leq v \leq P^{m-1}(z'') \leq P^m(z'')$. 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.

3 Binary Endogenous Variable

In this section we apply the LSRM model to a binary endogenous variable and identify the signs of the ceteris paribus impacts of the binary variable - the 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. The values of binary endogenous variable Y, 0 and 1, do not indicate any order. Despite this fact, when we apply the same restrictions as $\mathcal{M}^{LSRM}$ to binary endogenous variables we can bound the values of partial differences. We adopt the same restrictions in $\mathcal{M}^{LSRM}$, but the unordered nature leads to different identification results.

Although in many empirical studies, the distribution of the treatment effects can deliver a 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.[37] However, partial differences of the structural quantile function which are interval identified by the quantile-based control function approach, can provide bounds for quantiles of treatment effects.

[37] 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.



Failure of Chesher Rank Condition – Interval by LSRM

Figure 11: 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 LSRM bound. The graph is drawn for the case with the non-negative response case. Note that unless Chesher (2005) rank condition holds we are not sure whether the quantiles 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 LSRM bounds. See <Figure 12>.

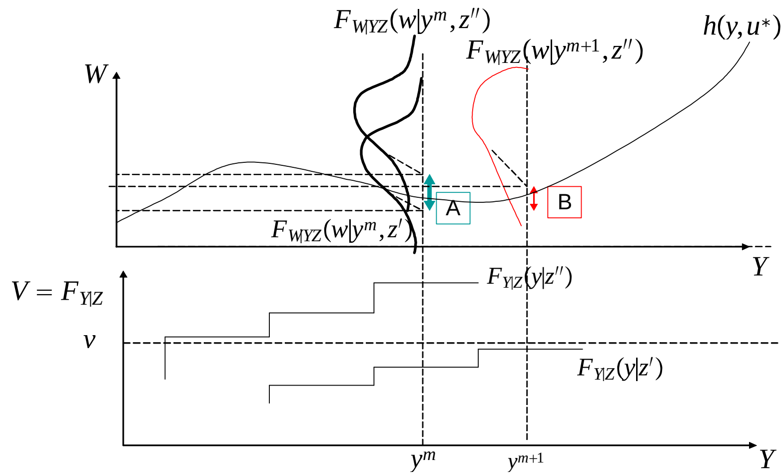


Figure 12: Testability of LSRM : when Chesher (2005) rank condition is satisfied Chesher bound(A) should be larger than or equal to LSRM bound(B) - if not, Restriction LSRM is not satisfied by the true structure.

3.1 The bounds on the values of the structural quantile function

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

Corollary 5 *Under Restriction A-EX,FI,RC,and LSRM there are the inequalities for* $\tau \equiv (\tau_U, v)$

$$\begin{aligned}
q^L(\tau, y, \overline{z}) &\leq h(y, u^*) \leq q^U(\tau, y, \overline{z}) \\
\textit{where } y &\in \{0, 1\} \\
z &\in \overline{z} = \{z', z''\},\ \tau \equiv (\tau_U, v), \\
u^* &\equiv Q_{U|VZ}(\tau_U | v, z) \\
q^L(\tau, y, \overline{z}) &= \min\{Q_{W|YZ}(\tau_U|0, z'), Q_{W|YZ}(\tau_U|1, z'')\}, \\
q^U(\tau, y, \overline{z}) &= \max\{Q_{W|YZ}(\tau_U|0, z'), Q_{W|YZ}(\tau_U|1, z'')\}, .
\end{aligned}$$

The bounds are sharp.

Proof. See 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 under Restriction LSRM we can identify the bounds on partial differences, $h(1, u^*) - h(0, u^*)$ as we can see in the next subsection.

3.2 Identification of heterogeneous treatment response, $W_1 - W_0$

In this subsection we show that we can use Corollary 5 to recover heterogeneous treatment effects.

Following the notation from the potential outcomes framework, we can link the structural equations and the potential outcomes as the following :

$$
\begin{aligned}
W &= YW_1 + (1-Y)W_0, \\
&\quad \text{where } W = h(Y,U), \\
&\quad \textit{and} \;\; Y \in \{0,1\}.
\end{aligned}
$$

where W_1 and W_0 are defined in the introduction.

We assume that the potential outcomes are generated by the following structural relation

$$
\begin{aligned}
W_1 &= h(1,X,U), \\
W_0 &= h(0,X,U).
\end{aligned}
$$

The treatment effects for each individual is $W_1 - W_0$, which may be heterogeneous, varying over individuals even after conditioning on the observables. However, since we do not observe both W_1 and W_0 for each individual, $W_1 - W_0$ can not be directly measured. The identification in the policy evaluation literature has focused on identifying some features of the distribution of $W_1 - W_0$, such as averages or quantiles from the marginal distributions of the counterfactuals, W_1 and W_0 under certain restrictions.

Corollary 5 can be used to derive some information on quantiles of the distribution of $W_1 - W_0$. Note that

$$W_1 - W_0 = h(1,X,U) - h(0,X,U).$$

Thus, the identification of the heterogeneous treatment effect, $W_1 - W_0$ is achieved by identification of $h(1,U) - h(0,U)$, the ceteris paribus impacts for the same U. Theorem 6 states the identification result of (possibly) heterogeneous treatment effects, $W_1 - W_0$.

Theorem 6 *Define $\Delta \equiv h(1,u^*) - h(0,u^*)$, the ceteris paribus impact of $Y \in \{0,1\}$. Under Restriction A-EX,FI,RC,and LSRM, Δ is identified by the following intervals:*

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

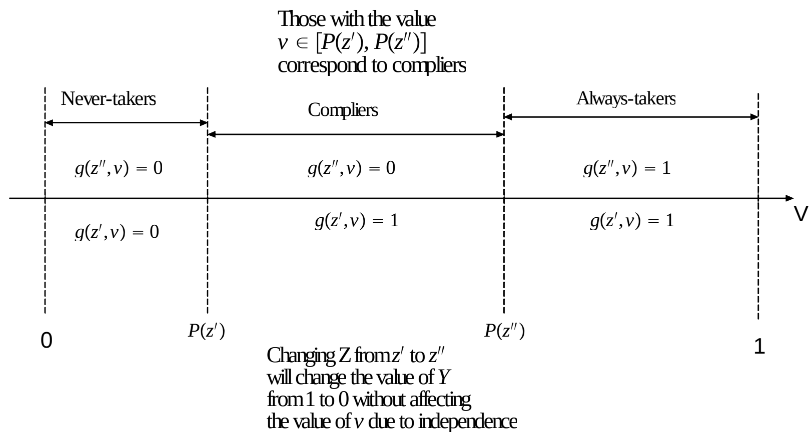


Figure 13: The structural vs. potential outcomes approaches

Proof. The result follow from Corollary 5 and Restriction LSRM. Either upper or lower bound is always 0 due to Restriction LSRM. ■

Discussion

1. The rank condition restricts the group in the whole population that the identification of causal impacts is possible into those who are ranked between $P(z')$ and $P(z'')$, where $P(z) = \Pr(Y = 0|Z = z)$. 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. <Figure 13> shows how we map our framework into the potential outcomes framework. The names never-takers, compliers, defiers and always-takers refer to the setting of a randomized experiment with noncompliance, where the instrument is the random assignment to the treatment and the endogenous regressor is an indicator for the actual receipt of the treatment.

2. In some sense we can only identify the signs of the impacts. These results allow the identification of the heterogeneous treatment effects : $\triangle$ would be understood as the treatment effects of the τ_U-ranked individuals in the sub-population whose V - ranking is $P(z') < v \leq P(z'')$. By varying τ_U, we could obtain the whole distribution of the treatment effects for this sub population.

4 Discussion

4.1 Comparison with Manski and Pepper (2000) and Bhattacharya, Shaikh and Vytlacil (2008)

Manski and Pepper (2000) and Bhattacharya, Shaikh and Vytlacil (2008) adopt certain monotonicity in the structural relations. Under MTS(Monotone Treatment Selection)-MTR (Monotone Treatment Reponse) restriction Manski and Pepper (2000) estimated the upper bounds on the returns to schooling. With monotonicity in response, the lower bound is always zero.[38] The two restrictions together define the bounds on the mean outcome, and the restrictions together can be tested, but whether each is true is not testable[39].

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, as LSRM 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 LSRM allows the direction (either PSPR or NSNR) of the match to vary over the support of U, while MTR-MTS allow the match - either positive response with positive selection or negative response with negative selection - to be determined a priori for the mean, Roughly speaking, LSRM restriction can be described as a local version of MTR-MTS. Another difference is that 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 LSRM model can recover heterogeneity in treatment effects among the observationally same individuals.

Bhattacharya, Shaikh and Vytlacil (2008) compare Shaikh and Vytlacil (2005) bounds with Manski and Pepper (2000)[40] by applying them to binary outcome and binary endogenous variable case. Bhattacharya, Shaikh and Vytlacil (2008)'s bounds are found under 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 ours. Note also that when Y is binary, we can always reorder 0 and 1 due to the "unordered nature" of a binary variable. Thus, the restrictions in Bhattacharya,

[38] In the returns to schooling example, MTR implies that each individual would earn more with more education, thus, on average, counterfactual wage function would be increasing with schooling. On the other hand, MTS implies that when college graduates are compared with high school graduates college graduates' "counterfactual" wage for different hypothetical schooling is on average higher than those for high school graduates.

[39] Okumura and Usui (2009) impose concavity to Manski and Pepper (2000) framework and showed that 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).

[40] 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.

Shaikh and Vytlacil (2008) are equivalent to our restrctions. 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[41]. Data will inform about the direction of the monotonicity, however, the direction of MTR and MTS should be matched into a certain way[42].

The advantage of LSRM (Local Selection Response Match) assumption is that it allows the match to vary with different quantiles unlike MTS-MTR in Manski and Pepper (2000) or Bhattacharya, Shaikh and Vytlacil (2008). However, LSRM may not be very informative when the outcome is binary in practice, since the values that the partial difference difference can take are -1,0, and 1, while it is legitimate to apply to binary outcomes in theory.

4.2 Comparison with Jun, Pinkse, and Xu (2009)

Jun, Pinkse, and Xu (2009, JPX(2009) hereafter) devise a new rank condition that can be applied to the Chesher (2005) setup. They include all the restrictions from Chesher (2005) except for the Chesher (2005)'s strong rank condition and conclude that their bound is tighter than Chesher (2005) bounds and when a continuous IV exists, point identification can be achieved when a continuous IV is available even with a discrete endogenous variable in the presence of endogeneity.

Although their identification stretegy can produce a different identified interval from Chesher (2005) interval, their conclusion that their bound is "tighter" and point identification is achieved seems to be a bit misleading. First of all, their conclusion of tighter bounds seems to be drawn by comparing Chesher (2005) bounds with their bounds applied to a binary endogenous variable. This is not comparable because Chesher (2005) bounds cannot be applicable to binary endogenous variable since

[41] 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.

[42] 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 say $E(y|z=0) \leq E(y|z=1)$, 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.

Chesher (2005) rank condition is not satisfied for binary endogenous variable. The Chesher (2005) bound is not defined for binary endogenous variable. To claim that their bounds are "tighter" than Chesher (2005) bounds, they should consider an ordered discrete endogenous variable with more than three points in the support. However, when applied to an ordered discrete variable, if there are more than one pair of values of IV that satisfy the Chesher (2005) rank condition, there is no clear conclusion can be drawn regarding whether Chesher (2005)'s min-max rank condition is stronger than their rank condition in the sense that whenever either of the rank condition holds, the other holds. This would be determined by data in principle[43] depending on the nature of the relationship between IV and the endogenous variable. If their rank condition is stronger, then the bound found by their model should be smaller as we have demonstrated by our LSRM bounds and Chesher (2005) bounds comparison in section 2.6 Lee (2009b). Secondly, their example 2 to show point identification, does not seem to be enough since in such a case their Lemma 4 is not applicable.

4.3 Applicability to regression discontinuity designs (RDD)

The regression discontinuity design is a quasi-experimental design with the probability of receiving treatment changes discontinuously as a function of one or more underlying variables (See Hahn, Todd, and Van der Klaauw (2001), Lee and Lemieux (2009), and the recent special issue of Journal of Econometrics, 2008). The regression discontinuity methods can be useful since geographic boundaries and eligibility for a program often creates discontinuities in the treatment assignment rule that can be exploited. 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 (See Hahn, Todd and Van der Klaauw (2001) for average treatment effects, and see Frandsen (2009) and Frolich and Melly (2009) for QTE).

However, the continuity condition (Assumption (A1) in Hahn, Todd, and Van der Klaauw (2001)) is not usually satisfied in many applications. When the RDD is available, our rank condition[44] is guaranteed to hold, thus, as long as Restriction LSRM is applicable into the context of interest, our model can be applicable to an RD design even when all other variables are not continuous in the treatment - determining variable at the threshold. For example, age or date of birth are often only available at a monthly, quarterly, or annual frequency level. Studies relying on an age-based cutoff thus typically rely on discrete values of the age variable when implementing an

[43] Rank condition is testable using data in principle. Checking whether the rank conditions are satisfied is a different issue.

[44] 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.

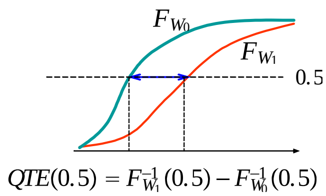


Figure 14: QTE

RD design.

Under the assumptions in Hahn *et al* (2001), the causal interpretation is possible on the sub-group of individuals at the discintinuity threshold, and uninformative about other groups. On the other hand, when the LSRM model is applicable, causal interpretation is plausible on the subgroup of individuals whose V-ranking is between $P(z')$ and $P(z'')$.

4.4 Different approaches to heterogeneous treatment response

We discuss three different approaches to recover heterogeneous treatment effects. The three approaches can answer different policy questions.

Quantile Treatment Effect(QTE) Quantile treatment effect(QTE) defined by Lehman (1974) and Doksum (1974) has been used to recover heterogenous treatment effects by several papers[45]. QTE is defined as the horizontal differences of the marginal distributions of the potential outcomes.

Interpretation of QTE QTE can be used to investigate the impacts of any policy on, for example, median individuals in the distributions with and without a

[45] Imbens and Rubin (1997) proposes a way to identify the marginal distributions of the potential outcomes under the LATE assumptions. Abadie (2002) reports the identification results of the marginal distributions of the potential outcomes to develop the tests of equality, first and second-order stochastic dominance. Abadie, Angrist, and Imbens (2002) specified the QTE as the linear functions rather than recovering QTE from the marginal distributions. What Abadie, Angrist, and Imbens (2002) identify is not the quantiles of $W_1 - W_0$, rather, it is the impact of the treatment on the quantiles of an outcome distribution, the quantile treatment effects (QTE). Firpo (2007) studies the identification and estimation of the marginal distributions of the potential outcomes under the unconfoundedness assumptions. Frolich and Melly (2009) and Frandsen (2009) study QTE under the regression discontinuity design.

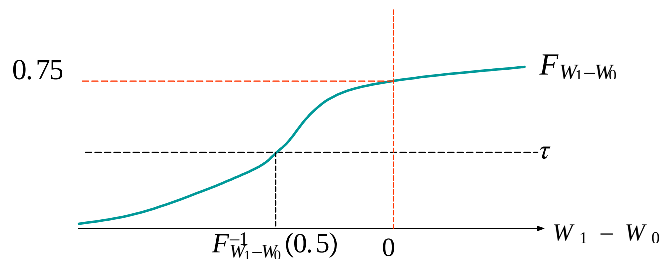


Figure 15: 25% of the population has benefitted from the treatment.

policy, which can be informative in the study of changes in inequality. However, QTE would not be used to derive information on the impacts of the policy on individuals 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 the rank is preserved across the treatment status, the size of the QTE would not necessarily be the same as the quantiles of the treatment effects.

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, $F_{W_1-W_0}$. Examples are Heckman, Smith, and Clements (1997), Fan and Park (2009), and Firpo and Ridder (2008). Their object of identification is $F_{W_1-W_0}$, and the identification results are found by the bounds studied in the statistics literature such as Hoeffding bounds, or Makarov bounds. These bounds are found once the marginal distributions of the potential outcomes first. Heckman, Smith, and Clements (1997) assumed that the potential outcomes are normally distributed, and Fan and Park (2009) assume that experimental data are available so that the marginal distributions of the potential outcomes are found. 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.

The information on $F_{W_1-W_0}$ 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.

Quantiles of treatment effects (Q - TE) recovered from partial differences Heterogeneous treatment effects, $W_1 - W_0$, under the potential outcomes framework, can be measured by partial differences under the structural framework, as $W_1 - W_0 = h(1, x, u) - h(0, x, u)$. Note that the heterogeneity in both observable and unobservable dimensions can be recovered by varying the values of x and u of the partial differences $h(1, x, u) - h(0, x, u)$. That is, for fixed values of x, heterogeneity in responses is identified for individuals ranked differently in the outcome. We call this object quantiles of treatment effects.

$h(1, x, u) - h(0, x, u)$ 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 is 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, $W_1 - W_0$.

Comparison of the three

1. QTE vs Q - TE from partial differences : In general, they should be different even with the rank preservation assumption. Chernozhukov and Hansen (2005) identify $h(1, x, u)$ and $h(0, x, u)$ separately using IV-QR. They identify the quantiles of the marginal distribution of the potential outcomes by identifying the structural functions $h(1, x, u)$ and $h(0, x, u)$ without identifying the distribution of the potential outcomes. Thus, as long as the rank of U is preserved (the rank similarity condition in their paper), we can interpret the QTE as Q-TE, which is what they call structural quantile effects. Whether we can interprete what they identify as "causal effect on individual" will depend on whether the rank similarity condition holds. Our object of identification is $h(1, x, u) - h(0, x, u)$. Thus by definition of partial differences, our object can be interpreted as causal effects of Y other things being equal. We do not have to rely on the rank similarity restriction.

2. Q - TE from $F_{W_1 - W_0}$ vs. Q - TE from partial differences : 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 questions of "who benefits" from the treatment by identifying "who" using the observed characteristics and the ranking of the unobserved heterogeneity. That is, we can identify whether the treatment effects are positive or not for the individuals with characteristics $X = x$ and the ranking of the unobservable is $\tau_{U.}$ Our results can then recover the proportion of population whose treatment effects are positive.

4.5 Inference

The inference results under set identification can be categorized into two[46] : 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 studies 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 the studies 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, more relevant to the first line of studies.

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).

When the endogenous variable is binary, the inference problem is somewhat different. The inference problem from the identification results would be (i) estimating the upper bounds or lower bounds as the differences 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 upper bound or the lower bound contains zero since what we identify is the sign of the partial differences and (iii) constructing the confidence intervals for the identified interval. If all the covariates are discrete then the quantiles would be found by the definition of the quantiles of the distribution defined in each cell defined by the values of the covariates. If there are reasons to believe that this distribution is smooth then we would have to impose smoothness in estimating. If some of the covariates are continuous we could use some existing nonparametric methods (see Chaudhuri (1991) or Chaudhuri, Doksum, Samarov (1997)) for the quantiles. (i) and (ii) can be done by existing methods, but (iii) requires extra consideration.

The major inference issue in our identification results 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. This can be done by constructing the confidence intervals of the upper/lower bounds. However, constructing the confidence intervals of the identified set needs more care since either upper or lower bound is always zero, thus, it needs not be estimated. This may lead to a bigger identified interval than the confidence interval of the estimated upper/lower bound. This is where the usual ways of constructing the confidence set fail to apply.

[46] We 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 Empirical Illustration

We illustrate how our results can be used in recovering more heterogeneous information by examining the effects of the Vietnam-era veteran status on the civilian earnings using the data used in Abadie (2002)[47]. We use a sample of 11,637 white men, born in 1950-1953, from March Current Population Surveys of 1979 and 1981-1985. Annual 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, schooling, etc. Whether they are compensated for their service enough has been an important political issue and there has not been any consensus on this matter. Using a linear 2SLS, Angrist (1990) reports negative impacts of veteran status on earnings later in life. Abadie (2002) as well report negative LATE estimates. These results show that on average military service had a negative impact on earnings possibly due to the loss of labour market experience.

The concern about selection of veteran status has led to the use of IV : those who joined the army may be systematically different in unobserved characteristics, thus the causal effects of veteran status on earnings may be biased. As in Angrist (1990) random variation in enrollment induced by the Vietnam era draft lottery is used as the instrument to identify the effects of veteran status on civilian 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[48]. We construct a binary IV based on the lottery number the threshold point being chosen by the government. It would be natural to believe that this IV is not a determinant of earnings, while it will affect the veteran status.

The findings in this section show that when age, gender, and race are controlled, the veteran status had negative impacts for individuals with high earnings potential, while it had positive effects for those with low earnings potential. This information may be useful in evaluating the effects of net benefits. The costs of military service must be larger than the benefits provided by the government for those with high earnings potential, while the benefits may be sufficient for those with low earnings potential. Considering the fact that benefits in terms of insurance, pension, or education opportunity should be targeted at people with less potentials, the findings indicate that the compensation was enough for this group. However, the military service may have higher opportunity costs for individuals with high earnings potential. This findings may be used against conscription.

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

[48] The value of the ceiling varied from 95 to 195. The eligibility is determined by the Department of Defense depending on the needs in the year. We follow Abadie(2002) and the threshold value is 100.

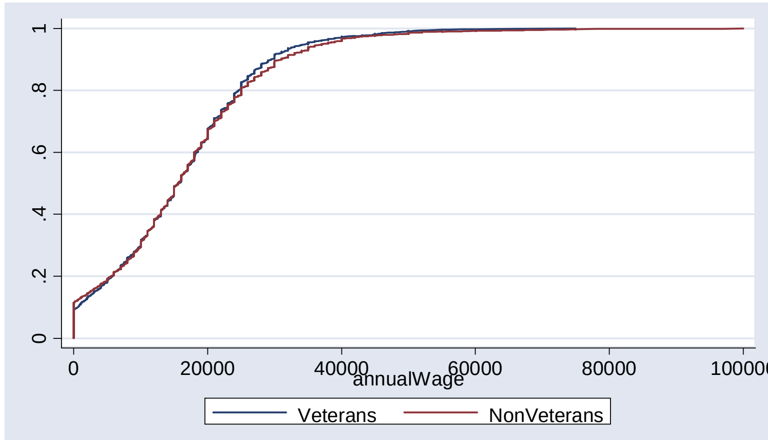


Figure 16: Distributions of annual income for veterans and nonveterans

5.1 The bounds on the 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 the veteran status appears to reduce lower quantiles of the earnings distribution, leaving higher quantiles unaffected. <Figure 16> shows that the cumulative distributions of annual wage of veterans and non-veterans. This is the reproduction of the figure 1 in Abadie (2002). The graph shows that there are differences in the lower quantiles and the upper quantiles, but the average does not seem to be affected. For the lower quantiles veterans's annual earnings are higher than those of non-veterans, and for the upper quantiles the opposite happens. However, as was mentioned, we cannot make any causal conclusion regarding the effects of the veteran status on the earnings based on the simple comparison of the two since the veteran status is likely to be selected by the individuals because it was possible to avoid enrollment for the reasons such as student status, or occupation or family. Moreover, there was a selection process among the volunteers or those drafted based on their health conditions or any felony experience.

Let W be annual earnings, Y be the veteran status, Z be the binary variable determined by the draft lottery. We control age, race, and gender so that the group we are considering is observationally homogenous. The unobserved variables U and V indicate a scalar index for "earnings potential" and "participation preference" 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" single index.

To apply the identification results in Theorem 6 we investigate whether the data satisfy Restriction RC in the model. The participation rate[49] among the draft-non-

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

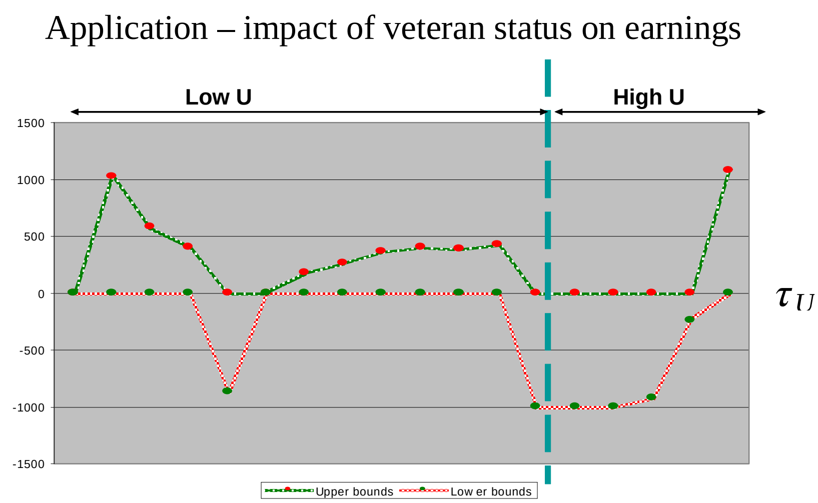


Figure 17: LSRM bounds on heterogeneous treatment effects of Vietnam era veteran status among the observationally similar individuals

eligible ($Z = 0$) is about 0.14 and the participation rate among eligible is 0.22.

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

Thus, $z' = 1$ and $z'' = 0$ in our framework. 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 bounds for the partial differences, $Q_{W|YZ}(\tau_U|1, z'') - Q_{W|YZ}(\tau_U|0, z')$, would be 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.

The LATE can be found by the model in Imbens and Angrist (1994). LATE allows for heterogeneous treatment effects for different subpopulation defined by observed covariates. However, unobserved heterogeneity is integrated out. Without loss of generality we assume that U is uniformly distributed on (0,1). Then LATE can be calculated as

$$LATE = \int_0^1 \left\{ \int_{P(z')}^{P(z'')} \left[h(1, Q_{U|V}(a|b) - h(0, Q_{U|V}(a|b)) \right] dFb \right\} dFa.$$

LATE is found by integrating out the heterogeneity for compliers, therefore, hiding useful information regarding heterogeneity. Abadie (2002) reported that the veteran status seems to have on average negative impact of $1,278, on the annual earnings for compliers when estimated by Imbens and Angrist (1994). Our quantile based analysis

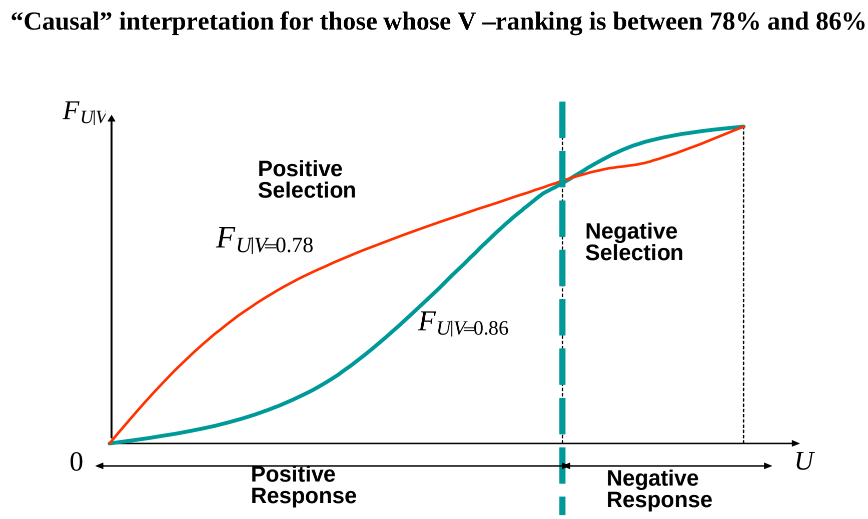


Figure 18: The implications on $F_{U|V}$ from the identification results and Restriction LSRM : Among the "**compliers/draftees**"(people whose V-rank is between 78% and 86%), those who are more likely to join the army (higher V) are more likely to be high ranked in the income distribution (higher U).

reveals that the veteran status had positive impacts for the low-ranked individuals in the income distribution, but negative impacts for the high-ranked individuals (see <Figure 17>). LATE is found without any assumptions on the underlying structure. By imposing the existence of underlying structure and by imposing some restrictions on the structure we could derive more heterogeneous information from data than LATE. How credible the restrictions are, therefore, should be discussed.

The implications of the LSRM restriction on the distribution on the unobservable can be described as in <Figure 18>. The results in <Figure 17> are interpreted as the causal effects for those who change their participation decision as the value of Z changes. The implication from the results should be considered to be true for those group. Among the compliers (the people ranked between 78% and 86%), those who are more likely to join the army (higher V) are more likely to be high ranked in the income distribution. To the extent that we believe this implication on the distribution of the unobservable the bounds would be considered to be informative regarding the population.

In <Figure 19> we report the estimated partial differences of the structural quantile function for given quantiles, and the QTE obtained by estimating the marginal distributions of the potential outcomes for compliers following Abadie(2002). The quantiles are found by conditioning several covariates - the subpopulation of white males, aged 26-29. Over almost all quantiles the sign and magnitude of the two estimates contradict each other. This shows that QTE can be very different from quantiles of treatment effects, and this could be because of failure of rank preservation condition. In such a case focusing on identification of the distributions of the potential outcomes would not be informative in identifying the causal effects on individuals.

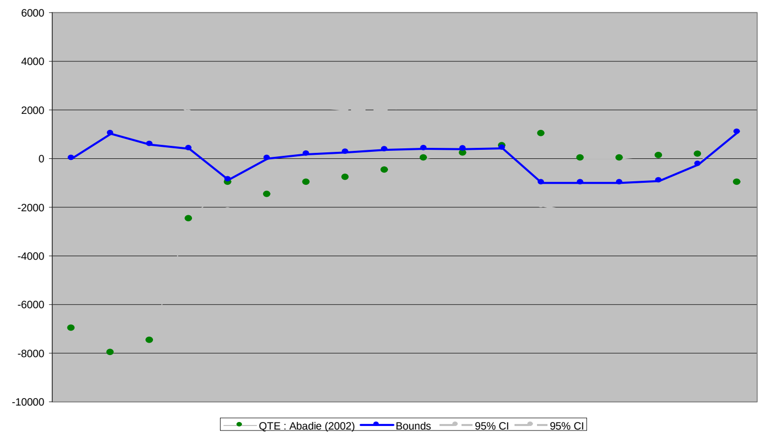


Figure 19: QTE vs. Q - TE : Over almost all quantiles the sign and magnitude of the two estimates contradict each other. This shows that QTE can be very different from quantiles of treatment effects, and this could be because of failure of rank preservation condition. In such a case focusing on identification of the distributions of the potential outcomes would not be informative in identifying the causal effects on individuals.

5.2 Caveats

The set identification result of this paper is applied to recover heterogeneous impacts of the Vietnam-era military service on the earnings later in life. The causal interpretation is justified on the group of compliers. Heterogeneity in responses is recovered for different earnings potentials. If there exists heterogeneity in responses between draftees and volunteers, then our findings cannot be extrapolated into volunteers.

6 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. We propose a model that set identifies the structural features when one of the regressors is ordered discrete. We then apply the model to binary endogenous variable, and found that the "unordered" nature of binary variable restricts the identification results. However, our structural approach turns out to be useful in defining the bounds on the heterogeneous treatment effects, which has not been studied so far under the structural framework without distributional assumptions. The fact that endogeneity causes the loss of identifying power emphasizes the need for testing of exogeneity of a variable. The discussion in this paper suggests one way of doing it.

Appendix I - proofs

A.1 Proof of Theorem 1 : bounds by $\mathcal{M}^{LSRM}$

Proof. We adopt Lemma 2 in Appendix in Chesher (2005).

To show that the value of the structural function $h(y, u^*)$, evaluated at $y = y^m$ and $u^* \equiv Q_{U|VZ}(\tau_U|v, z)$, is set-identified by the model, $\mathcal{M}^{LSRM}$, we need to show every structure admitted by $\mathcal{M}^{LSRM}$ that is observationally equivalent to the true data generating structure, S_0, $h(y^m, u^*)$ lies in the identified set.

Recall that we define $V \equiv (P^{m-1}(z), P^{m+1}(z)]$.

Suppose that $Q_{U|VZ}(\tau_U|v, z)$ is weakly increasing in $v \in V$. Then we have for $Y = y^m$,

$$\begin{aligned} h(y^m, Q_{U|VZ}(\tau_U|P^{m-1}(z'_m), z'_m)) &\leq Q_{W|YZ}(\tau_U|y^m, z'_m) && \text{(A-1)} \\ &\leq h(y^m, Q_{U|VZ}(\tau_U|P^m(z'_m), z'_m)) \\ h(y^m, Q_{U|VZ}(\tau_U|P^{m-1}(z''_m), z''_m)) &\leq Q_{W|YZ}(\tau_U|y^m, z''_m) && \text{(A-2)} \\ &\leq h(y^m, Q_{U|VZ}(\tau_U|P^m(z''_m), z''_m)) \end{aligned}$$

and for $Y = y^{m+1}$

$$\begin{aligned} h(y^{m+1}, Q_{U|VZ}(\tau_U|P^m(z'_m), z'_m)) &\leq Q_{W|YZ}(\tau_U|y^{m+1}, z'_m) && \text{(A-3)} \\ &\leq h(y^{m+1}, Q_{U|VZ}(\tau_U|P^{m+1}(z'_m), z'_m)) \\ h(y^{m+1}, Q_{U|VZ}(\tau_U|P^m(z''_m), z''_m)) &\leq Q_{W|YZ}(\tau_U|y^{m+1}, z''_m) && \text{(A-4)} \\ &\leq h(y^{m+1}, Q_{U|VZ}(\tau_U|P^{m+1}(z''_m), z''_m)) \end{aligned}$$

Under Restriction RC, $P^m(z'_m) \leq v \leq P^m(z''_m)$, when $Q_{U|VZ}(\tau_U|v, z)$ is weakly increasing in v, then :

$$\begin{aligned} Q_{U|VZ}(\tau_U|v, z''_m) &\leq Q_{U|VZ}(\tau_U|P^m(z''_m), z''_m) \\ Q_{U|VZ}(\tau_U|P^m(z'_m), z'_m) &\leq Q_{U|VZ}(\tau_U|v, z'_m) \end{aligned}$$

and because h is weakly increasing in U ,

$$\begin{aligned} h(y^m, Q_{U|VZ}(\tau_U|v, z''_m)) &\leq h(y^m, Q_{U|VZ}(\tau_U|P^m(z''_m), z''_m)) && \text{(B-1)} \\ h(y^m, Q_{U|VZ}(\tau_U|P^m(z'_m), z'_m)) &\leq h(y^m, Q_{U|VZ}(\tau_U|v, z'_m)). && \text{(B-2)} \end{aligned}$$

Combining (A-4) and (B-1) we can find **the upper bound** for $h(y^m, Q_{U|VZ}(\tau_U|v, z''_m))$

$$\begin{aligned} h(y^m, Q_{U|VZ}(\tau_U|v, z''_m)) &\leq h(y^m, Q_{U|VZ}(\tau_U|P^m(z''_m), z''_m)) \\ &\leq h(y^{m+1}, Q_{U|VZ}(\tau_U|P^m(z''_m), z''_m)) \\ &\leq Q_{W|YZ}(\tau_U|y^{m+1}, z''_m) \end{aligned}$$

The first inequality is due to (B-1) and the second inequality is due to Restriction LSRM, and the third inequality is due to (A-4).

The lower bound for $h(y^m, Q_{U|VZ}(\tau_U|P^m(z'_m), z'_m))$ can be found by (A-3) and (B-2) :

$$Q_{W|YZ}(\tau_U|y^m, z'_m) \leq h(y^m, Q_{U|VZ}(\tau_U|P^m(z'_m), z'_m)) \leq h(y^m, Q_{U|VZ}(\tau_U|v, z'_m)).$$

The first inequality is due to (A-3), the second is due to (B-2).

Finally, under **the conditional full independence (C-FI) and exclusion Restrictions (A-EX),** there is for $z \in \{z'_m, z''_m\}$ for $u^* \equiv Q_{U|VZ}(\tau_U|v, z)$,

$$Q_{W|YZ}(\tau_U|y^m, z'_m) \leq h(y^m, u^*) \leq Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)$$

Similarly, when $Q_{U|VX}$ is weakly decreasing in $v \in V$, we have

$$Q_{W|YZ}(\tau_U|y^{m+1}, z''_m) \leq h(y^m, u^*) \leq Q_{W|YZ}(\tau_U|y^m, z'_m)$$

■

A.2 Proof of Theorem 2 : Sharpness

Notation : The case in which $F_{U|VZ}(u^*|v, z)$ is nonincreasing in v is called **PS** (Positive Selection) and the other case, **NS** (Negative Selection) for ease of exposition. 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).

Define

$$h^{-1}(y^m, w) \equiv \sup_u \{u : h(y^m, u) \leq w\}. \tag{*}$$

This implies

$$h(y^m, h^{-1}(y^m, w)) \leq w \tag{**}$$

with equality holding when $h(y^m, u)$ is strictly increasing in u.

Lemma 7 *(Lemma 1 in Chesher (2005)) Under Restriction A-EX and FI, the conditional distribution of W given $Y = y^m$ and $Z = z$ is*

$$\begin{aligned} F_{W|YZ}(w|y^m, z) &= \frac{1}{p_m(z)} \int_{P^{m-1}(z)}^{P^m(z)} F_{U|V}(h^{-1}(y^m, w)|s) ds, \qquad \textit{(Key)} \\ \textit{where } p_m(z) &= \Pr(Y = y_m|Z = z) \\ \textit{for } z &\in \{z'_m, z''_m\}. \end{aligned}$$

This lemma is the key in the construction of the distribution of the unobservables.[50] The left hand side of (Key) is what we observe and this object is generated by the process in the right hand side of (Key). The information regarding endogeneity is contained in the distribution of the unobservables, $F_{U|V}$. There can be many different forms of $F_{U|V}$ that produce the same observed data $F_{W|YZ}$. That is, the shape of the distribution of the observables is not determined by the distribution of the unobserved variables completely in contrast with when the endogenous variable is continuous.

Proof. 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 that (i) the constructed structure is observationally equivalent to the true structure ($F^{S^a}_{W|YZ} = F^0_{W|YZ}$) and (ii) they are admitted by LSRM model ($S^a \in M^{LSRM}$).

Part 1. Construction of a structure :

Let $I(\tau, y^m, \overline{z}_m)$ denote the identified interval, say, $[Q_{W|YZ}(\tau_U|y^m, z'_m), Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)]$. The sharpness of the other case, $I(\tau, y^m, \overline{z}_m) \equiv [Q_{W|YZ}(\tau_U|y^{m+1}, z''_m), Q_{W|YZ}(\tau_U|y^m, z'_m)]$ can be shown similarly. $w^* \in I(\tau, y^m, \overline{z}_m) \equiv [Q_{W|YZ}(\tau_U|y^m, z'_m), Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)]$.

1-A Construction of the distribution of the unobservables.

We construct the conditional distribution of the unobservables as the following based on the key relation described in (Key) for all $m \in \{1, 2, ..., M\}$ with $P^M(z) = 1, P^0(z) = 0$:

$$\begin{aligned}
F^a_{U|VZ}(u^*|v,z) &= F^a_{U|VZ}(h_a^{-1}(y^l, w^l)|v,z) \\
&\equiv \begin{pmatrix}
F^0_{W|YZ}(w^1|y^1,z), & \text{if} \quad 0 < v \le P^1(z) \\
F^0_{W|YZ}(w^2|y^2,z), & \text{if} \quad P^1(z) < v \le P^2(z) \\
\cdots & \cdots \\
F^0_{W|YZ}(w^l|y^{l-1},z), & \text{if} \quad P^{l-2}(z) < v \le P^{l-1}(z) \\
F^0_{W|YZ}(w^l|y^l,z), & \text{if} \quad P^{l-1}(z) < v \le P^l(z) \\
F^0_{W|YZ}(w^{l+1}|y^{l+1},z), & \text{if} \quad P^l(z) < v \le P^{l+1}(z) \\
\cdots & \cdots \\
F^0_{W|YZ}(w^M|y^M,z), & \text{if} \quad P^{M-1}(z) < v \le 1
\end{pmatrix} \qquad \text{(A)} \\
\text{where } u^* &= h_a^{-1}(y^m, w^*) = h_a^{-1}(y^1, w^1) \\
&= h_a^{-1}(y^2, w^2) = \cdots = h_a^{-1}(y^M, w^M)
\end{aligned}$$

[50] See Chesher (2007b, 2009) for the proof of sharpness in the structural approach. Note that in his proofs the key relation was

$$F^{S^a}_{WY|Z}(w^*, y|z) = F^a_{UY|Z}(h_a^{-1}(y, w^*), y|z)$$

since how Y is determined given Z is not specified as it is under triangularity. The proofs in Chesher (2007b, 2009) are by construction of 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 in Chesher (2007b) is concerned with constructing $F_{U|YZ}$, and using $F_{Y|Z}$ the object of interest $F_{UY|Z}$ can be recovered.

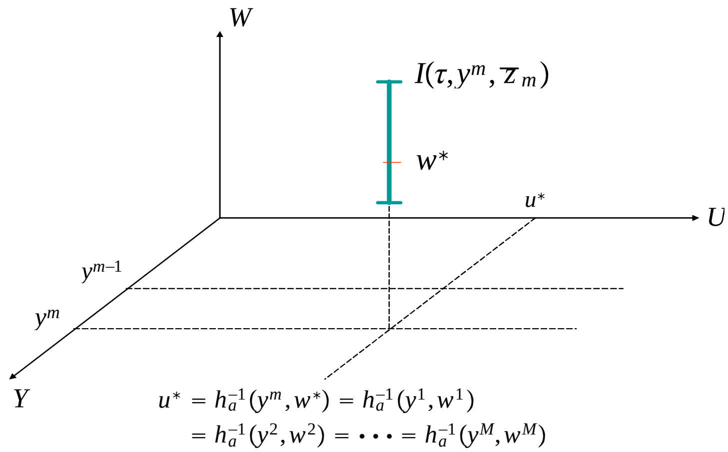


Figure 20: w^* is an arbitrary point in the identified interval. We construct a structural relation and a distribution of the unobservables by whose interaction w^* is generated. Construction of the distribution of the unobservables in (A) - note that u^* can be written in terms of the inverse function of the structural relation evaluated at different points of Y. For given w^* and y^m, w^l, $l = 1, 2, ..., M$, should be found such that $w^l = h_a(y^l, u^*)$.

Note that u^* can be expressed using $h_a^{-1}(y^m, w^*)$ by (B-1) and there are many pairs of (y^l, w^l) that produce the same value u^* (see <Figure 20>) where w^* is an arbitrary point in the identified interval, and $w^1, w^2, ..., w^M$ are values such that $w^1 = h_a(y^1, u^*), w^2 = h_a(y^2, u^*), ..., w^M = h_a(y^M, u^*)$. The value assigned to $F^a_{U|VZ}(u^*|v, z)$ is determined by how u^* is expressed and the value of v. Crucial part is when u^* is expressed as interms of y^l and v lies in the intervals $(P^{l-2}(z), P^{l-1}(z)]$ and $(P^{l-1}(z), P^l(z)]$. If $P^{l-2}(z) < v \leq P^{l-1}(z)$, assign $F^0_{W|YZ}(w^l|y^{l-1}, z)$ into $F^a_{U|VZ}(u^*|v, z)$ and if $P^{l-1}(z) < v \leq P^l(z)$, assign $F^0_{W|YZ}(w^l|y^l, z)$ to $F^a_{U|VZ}(u^*|v, z)$. This way, we can garantee the LSRM to hold. Note that in both intervals of V, the value of the conditional distribution of the observables evaluated at the same value $W = w^l$.

1-B Construction of a structural function[51]. Note that under the common support restriction any point in the identified interval, $w^* \in I(\tau, y^m, \overline{z}_m)$ whose value is in the support of W, can be written as

$$w^* = Q^0_{W|YZ}(\overline{\tau}_m|y^m, z'_m) \text{ for some } \overline{\tau}_m \geq \tau_U$$

That is,

$$\overline{\tau}_m \leq F^0_{W|YZ}(w^*|y^m, z'_m) \text{ for some } \overline{\tau}_m \geq \tau_U$$

[51]Unlike other bounds studies we need to construct the structural relation since we need to show monotonicity of the structural function in the unobservable and restriction LSRM.

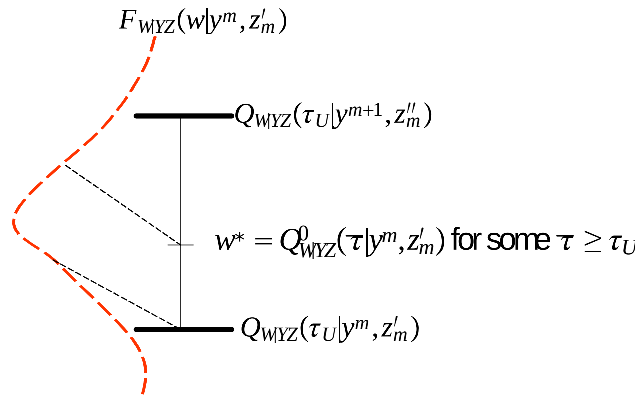


Figure 21: Any point in the interval, $w^* \in I(\tau, m, \overline{z}_m)$, can be expressed using the quantiles of $F_{W|YZ}(w|y^m, z'_m)$ under the common support restriction.

Note also that for $\overline{v}_m \in (P^{m-1}(z'_m), P^m(z'_m)]$ by construction

$$F^a_{U|VZ}(h_a^{-1}(y^m, w^*)|\overline{v}_m, z'_m) = F^0_{W|YZ}(w^*|y^m, z'_m) \geq \overline{\tau}_m$$

thus by definitionof quantiles,

$$h_a^{-1}(y^m, w^*) = Q^a_{U|VZ}(\overline{\tau}_m|\overline{v}_m, z) \text{ for some } \overline{v}_m \in (P^{m-1}(z'_m), P^m(z'_m)]$$

From this we have

$$w^* = h_a(y^m, Q^a_{U|VZ}(\overline{\tau}_m|\overline{v}_m, z)) \text{ for some } \overline{v}_m \in (P^{m-1}(z'_m), P^m(z'_m)]$$

Choose $\overline{v}_m \in (P^{m-1}(z'_m), P^m(z'_m)]$ such that $u^* = Q_{U|VZ}(\tau_U|v, z'_m) = Q_{U|VZ}(\overline{\tau}_m|\overline{v}_m, z'_m)$. Then we have

$$w^* = h_a(y^m, u^*).$$

that is, there exists a structural relation (that satisfies all the restrictions imposed by the model) which crosses an arbitrary point, w^*, in the identified interval.By this logic we construct the structural function as

$$h_a(y^m, u^*) \equiv Q^0_{W|YZ}(\overline{\tau}_m|y^m, z'_m) \text{ for some } \overline{\tau}_m \geq \tau_U \tag{B}$$

The whole structural function for given u^* can be defined as follows

$$h_a(y, u^*) \equiv \sum_{m=1}^{M}[Q^0_{W|YZ}(\overline{\tau}_m|y^m, z)]1(y = y_m) \tag{B-1}$$

The proof uses the fact that u^* can be written in different ways in terms of structural relation as well as quantiles of the distribution of the unobservables as we described in (A) and (B).

Part 2

The construction in Part 1 has been chosen to be admitted by M^{LSRM}, and to be observationally equivalent to S^0. In Part 2 we show how this is the case.

Part 2 - A : Observational equivalence[52] ($F^{S^a}_{W|YZ} = F^0_{W|YZ}$)

We need to show that $F^{S^a}_{W|YZ} = F^0_{W|YZ}$, for $S^a = \{h_a, F^a_{U|VZ}\}$ constructed as in Part 1 : for $p_m(z) = \Pr(Y = y_m|Z = z)$, for all $m \in \{1, 2, ..., M\}$, and for each $z \in \{z'_m, z''_m\}$,

$$\begin{aligned} F^{S^a}_{W|YZ}(w|y^m, z) &= \frac{1}{p_m(z)} \int_{P^{m-1}(z)}^{P^m(z)} F^a_{U|VZ}(h_a^{-1}(y^m, w)|s) ds \\ &= \frac{1}{p_m(z)} \int_{P^{m-1}(z)}^{P^m(z)} F^0_{W|YZ}(w|y^m, z) ds \\ &= F^0_{W|YZ}(w|y^m, z) \end{aligned}$$

the first equality is due to lemma 1 in Chesher (2005), the second equality is due to construction in (B) in Part 1, that is, $F^a_{U|VZ}(h_a^{-1}(y^m, w)|v, z) = F^0_{W|YX}(w|y^m, z)$, for $v \in (P^{m-1}(z), P^m(z)]$ and the last equality is due to integration over the constant and the definition of $p_m(z)$.

Part 2 - B : $S^a \in M^{LSRM}$

0. Rank condition : this can be shown using the data. We suppose this restriction is satisfied.

1. Monotonicity of $h_a(y^m, u^*)$ in u^*

Since we normalize $h_a(y^m, u^*)$ to be nondecreasing in u^* we consider whether $h_a(y, u^*)$ is nondecreasing in u^*. Recall that $h_a(y^m, u^*) = h_a(y^m, Q_{U|VZ}(\overline{\tau}_m|\overline{v}_m, z)) \equiv Q^0_{W|YZ}(\overline{\tau}_m|y^m, z)$, where $u^* \equiv Q_{U|V}(\tau_U|v) = Q_{U|VZ}(\overline{\tau}_m|\overline{v}_m, z)$.

First, fix $\overline{v}_m$, then $h_a(y^m, u^*)$ is weakly increasing in u^* since higher $\overline{\tau}_m$ implies higher $u^* \equiv Q_{U|V}(\overline{\tau}_m|\overline{v}_m)$, as well as higher $Q^0_{W|YZ}(\overline{\tau}_m||y^m, z)$. Next fix $\overline{\tau}_m$, if we observe higher u^*, then it is because of higher $\overline{v}_m$ if $F_{U|V}(u|\overline{v}_m)$ is nonincreasing in $\overline{v}_m$ and lower $\overline{v}_m$ if $F_{U|V}(u|\overline{v}_m)$ is nondecreasing in $\overline{v}_m$,for $(P^{m-1}(z), P^m(z)]$. However, regardless of the direction of the monotonicity, for $\overline{v}_m \in (P^{m-1}(z), P^m(z)]$, $Y = y^m$.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 $\overline{v}_m$.

2. Proper distribution

Now we need to check whether the constructed distribution is proper : since each $F^0_{W|YZ}(w|y^m, z)$, for all $m \in \{1, 2, ..., M\}$ is a proper distribution, $F^0_{W|YZ}(w|y^m, z)$ lies between zero and one, and weakly increasing in w. Thus, the constructed distribution

[52] 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 becasue we have constructed the structure using the observed distribution.

$F^a_{U|VZ}(u^*|v,z)$ lies between zero and one, but to guarantee nondecreasing property of $F^a_{U|VZ}(u^*|v,z)$ in u^*, we need to show that as w increases, $u^* \equiv h_a^{-1}(y,w)$ increases.

Let $w_r \equiv h_a(y,r) \le w$ and $w_{r'} \equiv h_a(y,r') \le w'$ for $r \equiv h_a^{-1}(y^m,w)$ and $r' \equiv h_a^{-1}(y^m,w')$. Suppose on the contrary that $w < w'$ as $r \ge r'$. Then since h_a is weakly increasing in r :

$$h_a(y,r') \le h_a(y,r) \le w < w'$$

The first inequality is due to weak monotonicity in u (nondecreasing by normalization) and the second inequality is due to $w_r = h_a(y,r) < w$ by (**) and the last inequality is due to the assumption made to derive contradiction. This implies that there exists a larger r than r' satisfying $h_a(y,r') \le w'$, which causes a contradiction due to the fact that $r' \ne h_a^{-1}(y,w')$ and $r = h_a^{-1}(y,w')$ since h_a^{-1} is defined as the largest value of u that satisfies $h_a(y,u) \le w'$. Thus we conclude that as $r \ge r'$, $w \ge w'$, in other words, as $F^0_{W|YZ}(w|y^m,z)$ is weakly increasing as w increases for given $Y = y^m$, we have weakly increasing $F^a_{U|VZ}(u|v,z)$. Therefore, $F^a_{U|VZ}(u|v,z)$ is a proper distribution.

3. Conditional Full Independence of $F_{U|VZ}(u^*|v,z)$ with respect to $z \in \{z'_m, z''_m\}$, where $u^* = h_0^{-1}(y^l,w^l)$, for fixed value of $v \in (P^{m-1}(z), P^m(z)]$. Let $(y^l,w^l),(y^m,w^m)$ be the pairs that produce the same value, u^*, that is, $u^* = h_a^{-1}(y^m,w^m) = h_a^{-1}(y^l,w^l)$, $l,m \in \{1,2,...,M-1\}$ for $l \ne m$ or $m-1$.

Recall that $p_m(z) \equiv \Pr(Y = y^m|Z = z)$. Then we have for $v \in (P^{m-1}(z), P^m(z)]$

$$\begin{aligned}
F^a_{U|VZ}(u^*|v,z) &= F^a_{U|VZ}(h_0^{-1}(y^l,w^l)|v,z) \\
&= F^0_{W|YZ}(w^m|y^m,z) \\
&= \frac{1}{p_m(z)} \int_{P^{m-1}(z)}^{P^m(z)} F^0_{U|V}(h_0^{-1}(y^m,w^m)|s)ds \\
&= \frac{\Pr(U \le h_0^{-1}(y^m,w^m) \cap P^{m-1}(z) \le V \le P^m(z))}{p_m(z)} \\
&= F^0_{U|V}(h_0^{-1}(y^m,w^m)|v) \\
&= F^0_{U|V}(u^*|v)
\end{aligned}$$

the first equality is by definition, the second equality is by construction in (A) in part 1, the third equality is due to lemma 1 in Chesher (2005), and the fourth equality follows by definition.

4. LSRM : Now we check whether the constructed $S^a = \{h_a, F^a_{U|VZ}\}$ satisfy the specified match.

For fixed u^* we can express u^* as the following :

$$u^* = h_a^{-1}(y^m,w^*) = h_a^{-1}(y^{m+1},w^{m+1}) \quad (4\text{-}1)$$

Note that $F^a_{U|VZ}(u^*|v,z)$ is monotonic in $v \in V$ since we constructed $F^a_{U|VZ}(u^*|v,z)$ as piecewise constant.

Step 1 : linking the distributions of the unobservables and the observables

(4-2)-(4-5) link the distribution of the unobservables with the distribution of the observables, and they are found by using the definition of u^* and the construction Part 1-A.

Using the construction in Part 1, for $u^* = h_a^{-1}(y^{m+1}, w^{m+1})$ and $v = P^m(z')$ define τ'_m as

$$\begin{aligned} \tau'_m &\equiv F^a_{U|VZ}(u^*|P^m(z'_m), z'_m) \\ &= F^a_{U|VZ}(h_a^{-1}(y^{m+1}, w^{m+1})|P^m(z'_m), z'_m) \\ &= F^0_{W|YZ}(w^{m+1}|y^m, z'_m) \end{aligned} \tag{4-2}$$

the first equality is due to $u^* = h_a^{-1}(y^{m+1}, w^{m+1})$ and the second equality is by the construction of of the conditional distribution of the unobservables in (A). This is key to exclude the possibility of NR

For $u^* = h_a^{-1}(y^m, w^*)$ and $v = P^m(z''_m)$ define τ''_m as

$$\begin{aligned} \tau''_m &\equiv F^a_{U|VZ}(u^*|P^m(z''_m), z''_m) \\ &= F^a_{U|VZ}(h_a^{-1}(y^m, w^*)|P^m(z''_m), z''_m) \\ &= F^0_{W|YZ}(w^*|y^m, z''_m) \end{aligned} \tag{4-3}$$

Note that for $u^* = h_a^{-1}(y^{m+1}, w^{m+1})$ and $v = P^{m+1}(z''_m)$:

$$\begin{aligned} \tau''_{m+1} &\equiv F^a_{U|VZ}(u^*|P^{m+1}(z''_m), z''_m) \\ &= F^a_{U|VZ}(h_a^{-1}(y^{m+1}, w^{m+1})|P^{m+1}(z''_m), z''_m) \\ &= F^0_{W|YZ}(w^{m+1}|y^{m+1}, z''_m) \end{aligned} \tag{4-4}$$

Also, for $P^m(z'_m) < \overline{v} < P^m(z''_m)$, we have[53]

$$\begin{aligned} \overline{\tau} &\equiv F^a_{U|VZ}(u^*|v, z''_m) \\ &= F^a_{U|VZ}(h_a^{-1}(y^{m+1}, w^{m+1})|\overline{v}, z''_m) \\ &= F^0_{W|YZ}(w^{m+1}|y^m, z''_m) \end{aligned} \tag{4-5}$$

Step 2 : Order of (4-2)-(4-5) :

Note $P^m(z'_m) \leq P^m(z''_m) \leq P^{m+1}(z''_m)$. Then PS implies that

$$\tau''_{m+1} \leq \tau''_m \leq \overline{\tau} \leq \tau'_m, \tag{*PS}$$

[53] This is for $P^{m-1}(z'') \leq P^m(z')$. Other cases can be shown similarly.

$$\begin{aligned} \overline{\tau} &\equiv F^a_{U|VZ}(r|v, z'') \\ &= F^a_{U|VZ}(h_a^{-1}(y^{m+1}, w^{m+1})|v, z'') \\ &= \begin{pmatrix} F^0_{W|YZ}(w^{m+1}|y^m, z'') & \text{if} \quad P^{m-1}(z'') \leq P^m(z') \\ F^0_{W|YZ}(w^{m+1}|y^{m+1}, z') & \text{if} \quad P^m(z'') \leq P^{m+1}(z') \end{pmatrix} \end{aligned} \tag{4-5'}$$

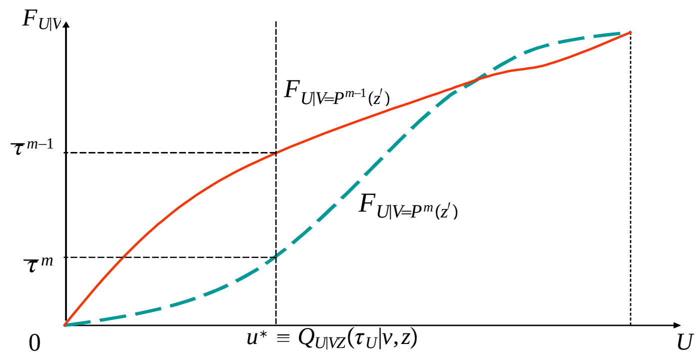


Figure 22: That conditioning on Y and Z corresponds to an interval, V, is the cause of loss of point idenifying power Note that $u^* \equiv Q_{U|V}(\tau_U|v) = Q_{U|VZ}(\overline{\tau}_m|P^m(z'), z') = Q_{U|VZ}(\overline{\tau}_{m-1}|P^{m-1}(z'), z')$

since we are comparing the values of the three conditional distributions evaluated at the same value $u^* = h_a^{-1}(y^m, w^*) = h_a^{-1}(y^{m+1}, w^{m+1})$. And NS implies that

$$\tau''_{m+1} \geq \tau''_m \geq \overline{\tau} \geq \tau'_m \tag{*NS}$$

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 (PS) and (NS). (4-2)-(4-5) imply (4-6) and (4-7) under continuity of W and U :

$$\begin{aligned} u^* &= Q^a_{U|VZ}(\tau'_m|P^m(z'_m), z'_m) \\ &= Q^a_{U|VZ}(\tau''_m|P^m(z''_m), z''_m) \\ &= Q^a_{U|VZ}(\tau''_{m+1}|P^{m+1}(z''_m), z''_m) \\ &= Q^a_{U|VZ}(\overline{\tau}|\overline{v}, z''_m), \quad \text{for } P^m(z'_m) < \overline{v} < P^m(z''_m) \end{aligned} \tag{4-6}$$

$$\begin{aligned} w^* &\overset{(a)}{=} Q^0_{W|YZ}(\tau''_m|y^m, z''_m) \\ w^{m+1} &\overset{(b)}{=} Q^0_{W|YZ}(\overline{\tau}|y^m, z''_m) \overset{(c)}{=} Q^0_{W|YZ}(\tau''_{m+1}|y^{m+1}, z''_m) \end{aligned} \tag{4-7}$$

(a) follows from (4-3), (b) from (4-5) and (c) is by (4-4).

Step 4 : Match?

Finally we use the construction of the structural function using (4-6). Then we can determine the direction of the response : we have [54]

$$\begin{aligned} & h_a(y^m, u^*) - h_a(y^{m+1}, u^*) \\ = \; & h_a(y^m, Q_{U|VZ}(\tau''_m | P^m(z''_m), z''_m) - h_a(y^{m+1}, Q_{U|VZ}(\tau''_{m+1} | P^{m+1}(z''_m), z''_m)) \\ = \; & Q^0_{W|YZ}(\tau''_m | y^m, z''_m) - Q^0_{W|YZ}(\tau''_{m+1} | y^{m+1}, z''_m) \\ = \; & Q^0_{W|YZ}(\tau''_m | y^m, z''_m) - Q^0_{W|YZ}(\overline{\tau} | y^m, z''_m) \\ & \begin{pmatrix} \leq 0 \text{ if PS} \\ \geq 0 \text{ if NS} \end{pmatrix} \end{aligned}$$

the first equality is by (4-6) : $u^* = Q_{U|VZ}(\tau''_m | P^m(z''_m), z''_m) = Q_{U|VZ}(\tau''_{m+1} | P^{m+1}(z''_m), z''_m)$, the second equality is by construction of the distribution of the unobservables (see <Figure 22>), and the third equality is by (c) in (4-7). Then the inequality follows because $\tau''_m \leq \overline{\tau}$ (*PS) and $\tau''_m \geq \overline{\tau}$ (*NS), and the property of quantiles. ■

A.3 Proof of Theorem 4 in section 3 : bounds by M^B.

Proof. We adopt the lemma above 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

$$\begin{aligned} h(0, Q_{U|VZ}(\tau_U | 0, z')) &\leq Q_{W|YZ}(\tau_U | 0, z') && \text{(A-1)} \\ &\leq h(0, Q_{U|VZ}(\tau_U | P(z'), z')) \\ h(0, Q_{U|VZ}(\tau_U | 0, z'')) &\leq Q_{W|YZ}(\tau_U | 0, z'') && \text{(A-2)} \\ &\leq h(0, Q_{U|VZ}(\tau_U | P(z''), z'')) \\ h(1, Q_{U|VZ}(\tau_U | P(z'), z')) &\leq Q_{W|YZ}(\tau_U | 1, z') && \text{(A-3)} \\ &\leq h(1, Q_{U|VZ}(\tau_U | 1, z')) \\ h(1, Q_{U|VZ}(\tau_U | P(z''), z'')) &\leq Q_{W|YZ}(\tau_U | 1, z'') && \text{(A-4)} \\ &\leq h(1, Q_{U|VZ}(\tau_U | 1, z'')) \end{aligned}$$

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, $P(z') \leq v \leq P(z'')$, when $Q_{U|VZ}(\tau_U | v, z)$ is weakly increasing in v, then :

$$\begin{aligned} Q_{U|VZ}(\tau_U | v, z'') &\leq Q_{U|VZ}(\tau_U | P(z''), z'') \\ Q_{U|VZ}(\tau_U | P(z'), z') &\leq Q_{U|VZ}(\tau_U | v, z') \end{aligned}$$

[54] Recall that this is the case for $P^{m-1}(z'') \leq P^m(z')$. The other case can be shown similarly.

and because h is monotonic in u and normalized nondecreasing,

$$\begin{aligned} h(1, Q_{U|VZ}(\tau_U|v, z'')) &\leq h(1, Q_{U|VZ}(\tau_U|P(z''), z'')) \quad &\text{(B-1)} \\ h(1, Q_{U|VZ}(\tau_U|P(z'), z')) &\leq h(1, Q_{U|VZ}(\tau_U|v, z')). \quad &\text{(B-2)} \end{aligned}$$

Combining (A-4) and (B-1) we can find the upper bound for $h(1, Q_{U|VZ}(\tau_U|v, z''))$

$$h(1, Q_{U|VZ}(\tau_U|v, z'')) \leq h(1, Q_{U|VZ}(\tau_U|P(z''), z'')) \leq Q_{W|YZ}(\tau_U|1, z'')$$

Use the Restriction LSRM : $h(1, u) \geq h(0, u)$, for all values of z and u in the support of Z and U. Applying Restriction LSRM to (B-2)

$$h(0, Q_{U|VZ}(\tau_U|P(z'), z')) \leq h(1, Q_{U|VZ}(\tau_U|P(z'), z')) \leq h(1, Q_{U|VZ}(\tau_U|v, z')). \quad \text{(C)}$$

Applying (A-1) to (C), we have the lower bound for $h(1, Q_{U|VZ}(\tau_U|v, z'))$

$$Q_{W|YZ}(\tau_U|0, z') \leq h(1, Q_{U|VZ}(\tau_U|v, z')).$$

Finally, under the conditional independence restriction and exclusion Restriction C-FI and AB-EX, there is for $z \in \{z', z''\}$ for $u^* \equiv Q_{U|VZ}(\tau_U|v, z)$

$$Q_{W|YZ}(\tau_U|0, z') \leq h(1, u^*) \leq Q_{W|YZ}(\tau_U|1, z'') \quad \text{(D-1)}$$

Consider next the identification of $h(0, u^*)$.

Under Restriction RC, $P(z') \leq v \leq P(z'')$, when $Q_{U|VZ}(\tau_U|v, z)$ is weakly increasing in v, then :

$$\begin{aligned} Q_{U|VZ}(\tau_U|v, z'') &\leq Q_{U|VZ}(\tau_U|P(z''), z'') \\ Q_{U|VZ}(\tau_U|P(z'), z') &\leq Q_{U|VZ}(\tau_U|v, z') \end{aligned}$$

and because h is monotonic in U and normalized nondecreasing,

$$\begin{aligned} h(0, Q_{U|VZ}(\tau_U|v, z'')) &\leq h(0, Q_{U|VZ}(\tau_U|P(z''), z'')) \quad &\text{(B-3)} \\ h(0, Q_{U|VZ}(\tau_U|P(z'), z')) &\leq h(0, Q_{U|VZ}(\tau_U|v, z')). \quad &\text{(B-4)} \end{aligned}$$

using (A-4) and (B-3), and Restriction LSRM we can find the upper bound for $h(0, Q_{U|VZ}(\tau_U|v, z''))$

$$\begin{aligned} & h(0, Q_{U|VZ}(\tau_U|v, z'')) \overset{(a)}{\leq} h(0, Q_{U|VZ}(\tau_U|P(z''), z'')) \\ & \overset{(b)}{\leq} h(1, Q_{U|VZ}(\tau_U|P(z''), z'')) \\ & \overset{(c)}{\leq} Q_{W|YZ}(\tau_U|1, z'') \end{aligned}$$

(a) is due to (B-3), (b) follows from Restriction LSRM, and (c) is from (A-4).

Applying (A-1) to (B-4) we have

$$Q_{W|YZ}(\tau_U|0,z') \overset{(a)}{\leq} h(0, Q_{U|VZ}(\tau_U|P(z'), z')) \overset{(b)}{\leq} h(0, Q_{U|VZ}(\tau_U|v, z')).$$

(a) follows from (A-4) and (b) is from (B-4). Thus, the lower bound for $h(0, Q_{U|VZ}(\tau_U|v, z'))$

$$Q_{W|YZ}(\tau_U|0,z') \leq h(0, Q_{U|VZ}(\tau_U|v, z')).$$

Finally, under the conditional independence restriction and exclusion Restriction C-FI and A-EX, there is for $z \in \{z', z''\}$

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

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). ■

References

[1] Abadie, A. (2002), "Bootstrap tests for distributional treatment effects in instrumental variable models," Journal of the America Statistical 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] Altonji, J. and R. Matzkin (2005), "Cross section and panel data estimators for nonseparable models with endogenous regressors," Econometrica, 73. 1053-1102.

[4] Angrist, J. (1990), "Lifetime earnings and the Vietnam era draft lottery : evidence from the social security administrative records," American Economic Review, 80, 313-336.

[5] Angrist, J. (2004), "Treatment effect heterogeneity in theory and practice," The Economic Journal, 114.

[6] Athey, S. (2002), "Monotone comparative statics under uncertainty, " Quarterly Journal of Economics, 117, 187-223.

[7] Athey, S. and G. Imbens (2006), "Identification and inference in nonlinear difference in differences models," Econometrica, 74. 431-497.

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

[9] 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.

[10] Beresteanu,A., I. Molchanov, and F. Molinari (2009), "Sharp identification regions in game," mimeo.

[11] Bhattacharya, J., A. Shaikh, and E. Vytlacil (2008), "Treatment effects bounds under monotonicity assumptions : an application to Swan - Ganz Catheterization, " American Economic Review, 98. 351-346.

[12] Bitler, M., J. Gelbach, and H. Hoynes (2006), " What mean impacts miss : distributional effects of welfare reform experiments," American Economic Review, 96. 988-1012.

[13] 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, Vol. 75 (2), 323-363.

[14] Blundell, R. and J. Powell (2003), "Endogeneity in Nonparametric and semiparametric regression models, " in M. Drewatripont, 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).

[15] Blundell, R. and J. Powell (2004), "Endogeneity in semiparametric binary response models, " Review of Economic Studies, 71, 655-679.

[16] Blundell, R. and R. Smith (1994), "Coherency and estimation in simultaneous models with censored or qualitative dependent variables, " Journal of Econometrics, 64, 355-373.

[17] Blundell, R., M. Browning, and I. Crawford (2007), "Best nonparametric bounds of consumer responses, " cemmap working paper.

[18] Brown, B. (1983), "The identification problem in systems of nonlinear in the variables, " Econometrica, 51(1), 175-196.

[19] Browning, M. and J. Carro (2007), "Heterogeneity and microeconometrics 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.

[20] Chaudhuri, P. (1991), "Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation, The Anals of Statistics, 19, 760-777.

[21] Chaudhuri, P., K Doksum, and A. Samarov (1997), "On Average Derivative Quantile Regression," The Anals of Statistics, 25, 715-744

[22] Chalak, K., S. Schennach, and H. White, "Estimating average marginal effects in nonseparable structural systems," mimeo.

[23] Chernozhukov, V. and C. Hansen (2005), "An IV Model of Quantile Treatment Effects," Econometrica, Vol.73, No. 1, 245-261.

[24] Chernozhukov, V., H. Hong, and E. Tamer, (2007), "Estimation and confidence regions for parameter sets in econometric models, " Econometrica, 75. 1243-1284.

[25] Chernozhukov, V., I. Fernadez-Val, and W. Newey (2009), "Quantile and average effects in nonseparable panel model," mimeo.

[26] Chrnozhukov, V., S. Lee, and A. Rosen (2009) "Inference on intersection bounds, " mimeo.

[27] Chernozhukov, V., Riggobon, and T. Stoker (2009), " Set identification with Tobit type regressors, " mimeo.

[28] Chesher, A. (2003), "Identification in nonseparable models," Econometrica, 71, 1405-1441.

[29] Chesher, A. (2004), "Identification in additive error models with discrete endogenous variables, " cemmap working paper.

[30] Chesher, A. (2005), "Nonparametric identification under discrete variation," Econometrica, 73(5), 1525-1550.

[31] Chesher, A. (2007a), "Instrumental values," Journal of Econometrics, 139, 15-34 .

[32] Chesher, A. (2007b), "Endogeneity and discrete outcomes," cemmap working paper.

[33] Chesher, A. (2008), "Excess heterogeneity, endogeneity, and index restrictions," Journal of Econometrics, forthcoming.

[34] Cross, P. and C. Manski (2002), "Regressions, Short and Long, " Econometrica, 70(1), 357-368.

[35] Das, M. (2005), " Instrumental variables estimation of nonparametric models with discrete endogenous regressors," Journal of Econometrics, 124. 335-361.

[36] Das, M., W. Newey, and F. Vella (2001), "Nonparametric estimation of sample selection models," Review of Economic Studies, 70. 33-58.

[37] Deaton, A. (2009), "Instruments of development : randomization in the Tropics, and the search for the elusive keys to economic development," NBER working paper no. 14690.

[38] Doksum, K. (1974) "Empirical probability plots and statistical inference for non-linear models in the two sample case," The annals of Statistics, 2. 267-277.

[39] Fan, Y. and S. Park (2009), "Sharp bounds on the distribution of treatment effects and their statistical inderence," mimeo.

[40] Firpo, S. (2007), " Efficient semiparametric estimation of quantile treatment effects," Econometrica, 75. 259-276.

[41] Firpo. S. and G. Ridder (2008), "Bounds on functionals of the distribution of treatment effects," IEPR working paper 08.09.

[42] 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.

[43] Frandsen, B (2009), "Nonparametric identification and estimation of regression discontinuity quantile treatment effects," mimeo.

[44] Frolich, M. and B. Melly (2008), "Quantile treatment effects in the regression discontinuity design," mimeo.

[45] Galichon, A. and M. Henry (2009), "A test of non-identifying restrictions and confidence regions for partially identified parameters," Journal of Econometrics, forthcoming.

[46] Gourioux, C., J. Laffont, and A. Monfort (1980), "Coherency conditions in simultaneous linear equation models with endogenous switching regimes, Econometrica, 48, 675-695.

[47] Hahn, J., P. Todd, and W. Van der Klaauw (2001), "Identification and estimation of treatment effects with a regression-discontinuity design," Econometrica, 69, 201-209.

[48] Hausman, J. (1978), "Specification tests in econometrics," Econometrics, 46. 1251-1272.

[49] Heckman. J. (1979), "Sample selection bias as a specification error, " Econometrica, 47. 153-161.

[50] 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.

[51] Heckman, J. and S. Navarro-Lozano (2004), "Using matching, instrumental variables, and control functions to estimate economic choice models," The Review of Economics and Statistics, 86. 30-57.

[52] 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.

[53] 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.

[54] Heckman, J. and E. Vytalcil (2005), "Structural equations, treatment effects, and econometric policy evaluation," Econometrica, 73, 669-738.

[55] Heckman,J. and S. Urzua (2009), "Comparing IV with structural models : what simple IV can and cannot identify," NBER working paper. no. 14760.

[56] Hoderlein, S. and E. Mammen (2007), "Identification of marginal effects in non-separable models without monotonicity," Econometrica, 75. 1513-1518.

[57] 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.

[58] 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.

[59] Hurwicz, L. (1950) "Generalization of the concept of identification," in Statistical Inference in Dynamic Economic Models, Cowles Commision Monograph 10. Wiley, New York.

[60] Imbens, G. (2009), "Better LATE than Nothing : some comments on Deaton (2009) and Heckman and Urzua (2009), " mimeo.

[61] Imbens, G. and J. Angrist (1994), "Identification and estimation of Local average treatment effects, " Econometrica, 62, 467-476.

[62] Imbens, G. and C. Manski (2004), "Confidence intervals for partially identified parameters," Econometrica, 72. 1845-1857.

[63] Imbens, G. and W. Newey (2009), "Identification and Estimation of Triangular Simultaneous Equations Models without additivity," Econometrica, forthcoming.

[64] Jun, S., J. Pinkse, and H. Xu (2009), "Tighter bounds in triangular system," mimeo.

[65] Koenker, R. (2005), Quantile Regression, Econometric Society Monographs, Cambridge University Press.

[66] Koopmans, T. and O. Reiersol (1950), " The identification of structural characteristics, " Annals of Mathematical Statistics, 21, 165- 181.

[67] Lee, J. (2009a), "A consistent nonparametric bootstrap test of exogeneity," mimeo.

[68] Lee, J. (2009b), "A note on testability under set identification," mimeo.

[69] Lee, D. and T. Lemieux (2009), "Regression Discontinuity Designs in Economics," NBER working paper.

[70] Lehman, E. (1974), Nonparametric Statistical Methods Based on Ranks, San Francisco, Holden-Day.

[71] Lewbel, A. (2007), " Coherency and completeness of structural models containing a dummy endogenous variable, " International Economic Review, 48.1379-1391.

[72] Manski, C. (1990), "Nonparametric bounds on treatment effects," American Economic Reveiw, 65(6), 1311-1334.

[73] Manski, C. (1995), Identification problems in the Social Sciences, Harvard University Press.

[74] Manski, C. (1997), "Monotone Treatment Response," Econometrica, 65(6), 1311-1334.

[75] Manski, C. (2003), Partial identification of probability distribution, Springer-Verlag.

[76] Manski, C. (2005), Social Choice with Partial Knowledge of Treatment Response, The Econometric Institute Lectures, Princeton University Press.

[77] Manski, C. and J. Pepper (2000), "Monotone Instrumental Variables : With an Application to the Returns to Schooling," Econometrica, 68(4), 997-1010.

[78] Manski, C. and E. Tamer (2003), "Inference on Regressions with Interval Data on a Regressor or Outcome," Econometrica, 70(2), 519-546.

[79] Matzkin, R. (2003), " Nonparametric estimation of nonadditive random functions, " Econometrica, 71, 1332-1375.

[80] Milgrom, P. and C. Shannon (1994), "Monotone comparative statics, " Econometrica, 62. 157-180.

[81] Nekipelov, D (2008), "Endogenous Multi-Valued Treatment Effect Model under Monotonicity" mimeo.

[82] Okumura, T. and E. Usui (2009), "Concave-monotone treatment response and monotone treatment selection : with returns to schooling application," mimeo.

[83] Roehrig, C (1988), " Conditions for Identification in Nonparametric and Parametric Models," Econometrica, Vol. 56, No. 2, pp. 433-447.

[84] Shaikh, A. and E. Vytlacil (2005), "Threshold crossing models and bounds on treatment effects : a nonparametric analysis," mimeo.

[85] Smith, R. and R. Blundell, (1986), "An exogeneity test for a simultaneous tobit model, " Econometrica, 54, 679-685.

[86] Stoye, J. (2009), "More on confidence intervals," Econometrica, forthcoming.

[87] 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.

[88] Tamer, E. (2003), "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria, " Reveiw of Economic Studies, 70, 147 - 165.

[89] Vytlacil,E. and N. Yildiz (2007), "Dummy endogenous variables in weakly separable models," Econometrica, vol. 75, 757-779.