University of
St Andrews

School of Economics & Finance Discussion Papers

Sharp Bounds on Heterogeneous Individual Treatment Responses

Jinhyun Lee

Sharp Bounds on Heterogeneous *Individual* Treatment Responses

Jinhyun Lee
University of St Andrews

Abstract

This paper discusses how to identify individual-specific causal effects of an ordered discrete endogenous variable. The counterfactual heterogeneous causal information is recovered by identifying the partial differences of a structural relation. The proposed *refutable* nonparametric local restrictions exploit the fact that the pattern of endogeneity may vary across the level of the unobserved variable. The restrictions adopted in this paper impose a sense of order to an unordered binary endogenous variable. This allows for a unified structural approach to studying various treatment effects when self-selection on unobservables is present. The usefulness of the identification results is illustrated using the data on the Vietnam-era veterans. The empirical findings reveal that when other observable characteristics are identical, military service had positive impacts for individuals with low (unobservable) earnings potential, while it had negative impacts for those with high earnings potential. This heterogeneity would *not* be detected by average effects which would underestimate the actual effects because different signs would be cancelled out. This partial identification result can be used to test homogeneity in response. When homogeneity is rejected, many parameters based on averages may deliver misleading information.

1 Introduction

Policies provide individuals with incentives to change their choices. Different people might respond to a policy change differently. If there exists heterogeneity in responses, many econometric methods based on "averages" may fail to provide correct information.[1] Policy evaluation literature typically uses the potential outcomes approach in identifying treatment responses. This paper demonstrates how *additively nonseparable* structural functions are used in recovering heterogeneous causal effect

---

[1]See Angrist (2004) for the potential outcomes approach, and Hahn and Ridder (2011) for the structural approach.

and provides a model that identifies the signs of *individual* treatment effects.[2] The proposed model does not require differentiability of the structural functions nor continuity of observed vairiables. The model does not impose weak separability which would render it impossible to recover individuals' heterogeneous treatment responses.

1.1 Causality, Heterogeneity, and Nonseparable Structural Relations

Suppose we are interested in the impact of a variable (Y) chosen by individuals on their outcome (W) of interest, and suppose the economic process of W can be described by the following relation[3]

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

where X is a vector of characteristics that are exogenously given to individuals such as age, gender, and race, and U is unobserved individual characteristics.[4] Even among the individuals with the same observed characterisics we observe a distribution of the outcome due to the unobserved elements, U. Causal effects of a variable indicate the effects of the variable only, separated from other possible influences. When the outcome is determined by (1), the causal effects of changing the value of Y from y^a to y^b on W other things being equal would be measured by the partial difference of the structural function, h

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

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

---

[2] Under the potential outcomes framework, individuals' treatment effects - difference between the outcomes with and without a treatment - are impossible to *measure* because only either of them is observed, not both.

[3] The structural relation may be derived from some optimization processes such as demand/supply functions. We are agnostic about this. If there is not a well-defined economic theory behind them, then the structural relations can be simply understood as how the outcome and the choice are determined by other relevant (both observable and unobservable) variables. The structural relation delivers the information on "contingent" plans of choice or outcome when different values of X and U are given.

[4] In contrast with (1), switching regression models with a selection equation of the following form have been widely used :

$$
\begin{aligned}
W_0 &= h_0(0, X, U_0) \\
W_1 &= h_1(1, X, U_1)
\end{aligned}
\tag{2}
$$

The counterfactual outcomes are determined by distinct functional relations, h_0 and h_1, and the unobserved heterogeneity for the two counterfactual events, U_0 and U_1, are allowed to be different. The partial difference of h_0 or h_1 would not be interpretaed as causal effects.

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

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

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

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

This way we map the potential outcomes framework into the structural approach[5]. This is the key relation that justifies the *interpretation* of $h(1, x, u) - h(0, x, u)$ as *individual-specific* treatment response. In contrast with the potential outcomes approach, this paper focuses on identification of $h(1, x, u) - h(0, x, u)$, by assuming the existence of economic processes and by imposing restrictions on such decision mechanisms.

The conditional distribution of the outcome, $F_{W|YX}$, is determined by the *interaction*, indicated by the following Hurwicz Relation (HR), between the distribution of the unobserved elements, $F_{U|YX}$ and the structural relation, $h(\cdot, \cdot, \cdot)$

$$
\begin{aligned}
\underbrace{F_{W|YX}(w|y, x)}_{\text{"Data"}} &= \Pr[W \le w | Y = y, X = x] \\
&= \Pr[h(Y, X, U) \le w | Y = y, X = x] \qquad \text{(HR)} \\
&= \underbrace{\int_{\{u : h(y, x, u) \le w\}} dF_{U|YX}(u|y, x)}_{Hurwicz\ Structure}
\end{aligned}
$$

The two components, $h(\cdot, \cdot, \cdot)$ and $F_{U|YX}$, are called the Hurwicz (1950) structure. The identification problem in this paper is to recover $\Delta(y^a, y^b, x, u)$ or $\Delta(1, 0, x, u)$ by imposing restrictions either on the Hurwicz (1950) structure, $\{h(\cdot, \cdot, \cdot), F_{U|YX}\}$ or on the observed distribution, $F_{W|YX}(w|y, x)$, (or Data). A novel restriction is imposed on the mode of the *interaction* between h and $F_{U|YX}$. It exploits the fact that the pattern of endogeneity may vary across the level of the unobserved variable. This model would be particularly informative when the signs of individual effects vary across the population, in which case average effects would underestimate the true effects as different signs will be canceled out.

---

[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.

1.2 Related Studies and Contributions

Since Roehrig (1988)'s recognition of the importance of nonparametric identification, there have been many studies that aim to clarify what can be obtained from data without parametric restrictions (see Matzkin (2007) for a survey on nonparametric identification and the references therein). When parametric assumptions are avoided, point identification is often not possible[6] with a discrete endogenous variable. In such a case one could aim to define a set in which the parameter of interest can be located. This partial identification idea, which was pioneered by C. Manski (see Manski (2003) for a survey of earlier results), has been actively used in many different setups and since it now constitutes a vast literature we only focus on policy evaluation literature.

Many authors[7] emphasize the existence of heterogeneity in individual responses in practice and the importance of the information regarding individual-specific, possibly heterogeneous causal effects of a binary endogenous variable was recognized earlier. Many interesting parameters are functionals of the distribution of individual treatment effects as Heckman, Smith, and Clements (1997) noted[8].

Certain 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)[9]. The QTE can be found from the marginal distributions in principle. To control for possible slection issues, Abadie, Angrist, and Imbens (2002) study the QTE under the LATE-type assumptions using a linear quantile regression model, Firpo (2007) under the matching assumption, and Frandsen, Frolich and Melly (2012) under the regression discontinuity design. Chernozhukov and Hansen (2005)'s moment condition based on their IV-QR model provides a way to estimate QTE controlling for selection or endogeneity problem. However, QTE is not justified to use for individual-specific treatment effects.

One approach to recover individual-specific causal effects is to recover heterogeneity in treatment effects by identifying the distribution of $W_1 - W_0$ directly[10].

---

[6] 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.

[7] See, for example, Heckman (2001).

[8] When the treatment effects are homogeneous the problem is trivial and the distribution of the treatment effects is degenerate. See Firpo and Ridder (2008) for more discussion.

[9] By estimating quantile treatment effects (QTE) using the Connecticut experimental data Bitler, Gelbach, and Hoynes (2006) found that welfare reforms in the ninties had heterogeneous effects on individuals as predicted by labour supply theory. They conclude that "welfare reform's effects are likely both more varied and more extensive". Average effects may miss much information and can be misleading if the signs of individual treatment effects are varying across people. However, when experimental data are not available, QTE does not have causal interpretation on *individuals* because individuals' rankings in the two marginal distributions of the potential outcomes may change.

[10] The quantiles of treatment effects recovered from the distribution of $W_{1i} - W_{0i}$ are examples of

Heckman, Smith and Clements (1997) use the Hoeffding-Frechet bounds, and Fan and Park (2011) 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.

Alternative to these potential outcomes setups, one could use structural approaches. By adopting a triangular structural setup, Chesher (2003,2007) 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. Jun, Pinkse, and Xu (2010) report *tighter* bounds when a different rank condition from Chesher's (2005) is used, while the same restrictions on the structure as in Chesher (2005) are adopted. Jun, Pinkse, and Xu (2010) does not have identifying power for a binary endogenous variable if the IV is binary. Vytalcil and Yildiz (2007) use a triangular system and report a point identification result of the average treatment effect of a dummy endogenous variable under weak separability and an exclusion restriction. Their results rely on certian characteristics of variation in exogenous variables as well as exclusion restrictions to achieve point identification. Vytalcil and Yildiz (2007) does not guarantee identification of partial difference. They focus on identificaiton of the average effect, not the structural function. Manski and Pepper (2000) and Shaikh and Vytlacil (2011) have partial identification results on average effects.

This paper contributes to the nonparametric identification literature by providing new identification results on a non-additive structural function when an endogenous variable is discrete/*binary* by using a control function approach without relying on continuity of exogenous variables. Use of nonseparable relation is not just a generalization.[11] One of the key implications of additively nonseparable functional form is that partial differences are themselves stochastic objects that have distributions[12].

$D\Delta-$ 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".

[11] If there exist different responses among the observationally identical agents, and if there exists endogeneity, then nonseparable structural relation should be used. In this case conditional moment conditions do not have identifying power. See Hahn and Ridder (2011).

[12] If the structural function is linear, that is, $W = a + bY + cX_1 + U$, then the partial derivative of this linear function with respect to Y is b. Thus, assuming a linear structural relation corresponds to assuming "homogenous" responses. If an additively separable structural function, for example, $W = f(Y, X_1) + U$, allows for heterogeneity in responses, but once conditioning on the observables, there are no differences among the people with different unobserved characteristics as the ceteris paribus effect measured by the partial derivative, $\dfrac{\partial f(y,x)}{\partial y}$, is determined by observed characteristics

Thus, heterogeneity in individual causal effects can be found by identifying partial differences of a non-additive structural function. Nonetheless, individual-specific causal effects have not been discussed so far.

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

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

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

---

only.

6

1.3 Organization of the Paper

The remaining part is organized as follows. Section 2 introduces the model for "ordered" discrete endogenous variables and contains the main identification results. Section 3 discusses "unordered" binary endogenous variable as a different case of discrete endogenous variable. Section 4 discusses the restrictions imposed in the model and other related studies in more detail. Section 5 illustrates the usefulness of the identification results by examining the effects of the Vietnam-era veteran status on the civilian earnings using a binary IV. Section 6 concludes.

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

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

In this section a set of restrictions is introduced that interval identifies the value of the structural function evaluated at a certain point in the presence of an endogenous discrete variable. The model, $\mathcal{M}^{LDRM}$, is defined as the set of all the structures that satisfy the restrictions[13].

Restriction QCFA[14] : Scalar Unobservables Index (SIU)/Monotonicity/Exclusion

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

the function h is weakly increasing[15] with respect to variation in scalar U. The conditional distribution of Y given $X = x$ and $Z = z$ is discrete with points of support $y^1 < y^2 < ... < y^M$, invariant with respect to x and z, with positive probability

---

[13]Koopmans and Reiersol's (1950) definition of a model is adopted as a set of structures satisfying the restrictions imposed.

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

[15]Both h and g are restricted to be monotonic. This monotonicity restriction is one of the key restrictions in the QCFA identification strategy. This enables us to use the equivariance property of quantiles and g evaluated at $Z = z$, $X = x$ and $V = \tau_V$, $g(z, x, \tau_V)$ is identified by $Q_{Y|Z}(\tau_V|z, x)$. In many applications this can be justified. See Imbens and Newey (2009) for examples that justify monotonicity.

masses $\{p_m(z,x)\}_{m=1}^M$. Cumulative probabilities $\{P^m(z,x)\}_{m=1}^M$ are defined as

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

The scalar unobserved variables U and V are jointly continuously distributed and their marginal distributions are normalized uniformly distributed on $(0,1)$.

If g is weakly increasing in v, then h needs to be weakly increasing in u and if g is weakly decreasing, h needs to be be weakly decreasing as well. The monotonicity restriction on g is reflected in the threshold crossing structure. As g is assumed to be weakly increasing, h is assumed to be weakly increasing in Restriction QCFA. Because a binary variable is often unordered, Restriction QCFA imposes a sense of order. Whether to assume that h is weakly increasing or weakly decreasing is dependent on how to define the binary variable[16].

From here on other exogenous variables, X, are ignored. X can be added as conditioning variables in any steps of discussion without changing the results. The variable W is a discrete, continuous, or mixed discrete continuous random variable and all the results apply regardless of whether W is continuous or not. The model admits multiple factors of unobserved heterogeneity as long as they affect the outcome though a scalar index.[17]

Restriction CQ-I (Conditional Quantile Invariance) : $Q_{U|VZ}(\tau_U|v,z)$ is invariant with $z \in \overline{z}_m \equiv \{z'_m, z''_m\}$ for $v \in \mathbf{V}$ for $u \in \mathbf{U}$.

Restriction CQ-I is a weaker form of exclusion restriction imposed on Z. What is required for identificaion is quantile independence locally at certain points.

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

---

[16]For example, in the example of Vietnam-era verans, Y is 1 if joining in the army and it is assumed that individuals with higher V joins the army, the annual labour earnings eqaution, h, needs to be weakly increasing in U.

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

$$P^m(z_m') \le \tau_V \le P^m(z_m'')$$

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

Define $\boldsymbol{V} \equiv (V_L, V_U]$, where $V_L = \min_{z \in \bar{z}_m} P^{m-1}(z)$, and $V_U = \max_{z \in \bar{z}_m} P^{m+1}(z)$.[18] Define also $\boldsymbol{U} \equiv (U_L, U_U]$, where $U_L = \min_{\tau_V \in \boldsymbol{V}} Q_{U|VZ}(\tau_U|\tau_V, z)$, and $U_U = \max_{\tau_V \in \boldsymbol{V}} Q_{U|VZ}(\tau_U|\tau_V, z)$. $\boldsymbol{V}$ is determined by the variation between Y and Z, and $\boldsymbol{U}$ is determined by the "degree of the endogeneity", for example, U and V were highly dependent, $\boldsymbol{U}$ would be large. Any value $u^* \in \boldsymbol{U}$ can be written as a quantile of the conditional distribution of U given V and Z The value, u^*, is not known, but it indicates τ_U- ranked individual's value of U in the conditional distribution of U given V and Z. (See Appendix A.1)

For a given value $u^* \in \boldsymbol{U},$ the case in which $F_{U|VZ}(u^*|v, z)$ is nonincreasing in v is called PD (Positive Dependence) and the other case in which $F_{U|VZ}(u^*|v, z)$ is nonincreasing in v is called ND (Negative Dependence). Also, for a given value $u^* \in \boldsymbol{U},$ if $h(y^{m+1}, u^*) \ge h(y^m, u^*)$, it is called Strong-PR (Strong Positive Response) and if $h(y^{m+1}, u^*) \le h(y^m, u^*)$, it is called Strong-NR (Strong Negative Response). Weaker versions are also used. The case in which $h(y^{m+1}, \bar{u}) \ge h(y^m, u^*)$ for $\bar{u}, u^* \in \boldsymbol{U},$ with $\bar{u} \ge u^*$ is called PR (Positive Response) and the case in which $h(y^{m+1}, \bar{u}) \le h(y^m, u^*)$, $\bar{u}, u^* \in \boldsymbol{U},$ with $\bar{u} \le u^*$, is called NR (Negative Response). There can be distinct patterns of *interaction* between h and $F_{U|VZ}$ locally in $\boldsymbol{U}$ and $\boldsymbol{V}$. The next condition restricts the pattern of the interaction in certain ways.

Restriction LDRM (Local (Quantile) Dependence Response Match) : $F_{U|VZ}(u|v, z)$ is assumed to be weakly monotonic in $v \in \boldsymbol{V}$ for $u \in \boldsymbol{U}.$ If $F_{U|VZ}(u|v, z)$ is weakly decreasing in $v \in \boldsymbol{V}$ for $u \in \boldsymbol{U}$, then $h(y^{m+1}, \bar{u}) \ge h(y^m, u^*)$ for $\bar{u}, u^* \in \boldsymbol{U},$ with $\bar{u} \ge u^*.$ Conversely, $F_{U|VZ}(u|v, z)$ is weakly increasing in $v \in \boldsymbol{V}$ for $u \in \boldsymbol{U},$ then $h(y^{m+1}, \bar{u}) \le h(y^m, u^*)$, $\bar{u}, u^* \in \boldsymbol{U},$ with $\bar{u} \le u^*$ for $m \in \{1, 2, ..., M-1\}$. See <Figure 1>.

2.2 Discussion on Restrictions

2.2.1 Restriction QCFA - Scalar Index Unobservables, U and V

There is a tradeoff between using a vector and a scalar unobserved heterogeneity - allowing for a vector unobserved heterogeneity in the structural relation would be more realistic. Several studies report identification results without monotonicity restrictions. However, what can be identified without monotonicity is objects with the heterogeneity in responses averaged out, while the quantile-based approaches under

---

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

Figure 1: **"Local" nature of Restriction LDRM** : the information on endogeneity is contained in $F_{U|V}$ under triangularity- if Y is exogenous, then $F_{U|V}$ is invariant with V. $F_{U|V}$ is drawn for different values of V by assuming monotonicity in V. The solid line is the distribution of U given $V = v$. Monotonicity of $F_{U|V}(u^*|v)$ does not have to be global in U, all that is required is monotonicity in some region U of u. In this graph, $F_{U|V}(u_1^*|v)$ is decreasing in v, while $F_{U|V}(u_2^*|v)$ is increasing in $v \in V$ in distinct range of U. Notice that the range $\mathbf{U}$ is determined by the "degree of endogeneity", that is, if U and V were highly dependent, $\mathbf{U}$ would be large.

monotonicity can be adopted to recover heterogeneous treatment response only if a scalar (index) unobserved heterogeneity is assumed.

2.2.2 Rank Condition[19]

When the structural relation is linear, weak IVs are considered to cause problems in inference, *not in identification.* Under the nonparametric setup, weakness of IV (how closely the endogenous variable and the IV are related) causes problems not only in inference, but also in identification. In the nonparametric setup point identification fails if certain rank conditions or completeness conditions that specify how IV and the indigenous variable are related, are not satisfied.

As in Chesher (2005) the identification and testability results of this paper require restrictions on how the endogenous variable is related with the IV. The point-identifying power of Restriction QCFA and Restriction CQ-I in Chesher (2003) is lost when the endogenous variable is discrete. The set-identifying power of Chesher (2005) for an ordered discrete endogenous variable comes from the rank condition in addition to Restriction QCFA and Restriction CQ-I. Consider Chesher (2005)'s rank

---

[19]The name "rank condition" comes from the classical linear simultaneous equations model where a rank condition of a matrix indicates how certain variables are correlated. Under the nonparametric setup the rank conditions do not necessarily indicate the rank of a matrix, but they play a similar role - they specify how the endogenous variable and the IVs are related.

condtion of the following :

Restriction RC^C (Rank Condition in Chesher (2005)) There exist instrumental values of Z, $\{z'_m, z''_m\}$, such that

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

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

As is illustrated in Chesher (2005) by using the Angrist and Krueger (1992)'s quarters of birth IV example, if the IV is weak, Chesher (2005)'s rank condtion is not satisfied. If Chesher (2005)'s rank condition holds, our rank condition also holds since $P^{m-1}(z''_m) \leq P^m(z''_m)$. In this sense, Chesher (2005)'s rank condition is stronger than our rank condition, that is, even when Restriction RC^C fails, Restriction RC can be satisfied. Chesher (2005) is not applicable to a binary endogenous variable case as Restriction RC^C is not satisfied. All the rank conditions specified in this paper in principle can be tested once data are available.

2.2.3 Local Dependence and Response Match (LDRM)

Endogeneity is often defined as the dependence between explanatory variables and the unobserved elements in the structural relationship. They can be positively dependent or negatively dependent. "Dependence" is used instead of "correlation" to clarify the *local* information contained in Restriction LDRM. Under triangularity the source of endogeneity is caused by the dependence between U and V and this information is contained in the conditional distribution of $F_{U|V}$. The shape of $F_{U|V}$ would be varying significantly as the value of V changes if U and V were highly dependent.

Restriction LDRM is concerned with how the pattern of dependence varies with the level of the unobserved characteristic and the modes in which the pattern is linked with that of the response function. Restriction LDRM applies to locally each point in the support of the unobserved variable U. As U is normalized to be uniform $(0,1)$ and each point in $(0,1)$ is indicated by expressing it as *quantiles*, thus, "local" implications of Restriction LDRM can be understood in terms of *quantiles*.

To identify the sign of the partial difference with respect to a binary endogeneous variable, a stronger version of LDRM of the following is required.

Restriction S-LDRM (Strong Local (Quantile) Dependence Response Match) : $F_{U|VZ}(u|v,z)$ is assumed to be weakly monotonic in $v \in \boldsymbol{V}$ for $u \in \boldsymbol{U}$. If $F_{U|VZ}(u|v,z)$ is weakly decreasing in $v \in \boldsymbol{V}$ for $u \in \boldsymbol{U}$, then $h(y^{m+1}, u^*) \geq h(y^m, u^*)$, **(Strong PDPR)** and if $F_{U|VZ}(u|v,z)$ is weakly increasing in $v \in \boldsymbol{V}$ for $u \in \boldsymbol{U}$, then $h(y^{m+1}, u^*) \leq h(y^m, u^*)$, **(Strong NDNR)** for any $u^* \in \boldsymbol{U}$ for $m \in \{1, 2, ..., M-1\}$.
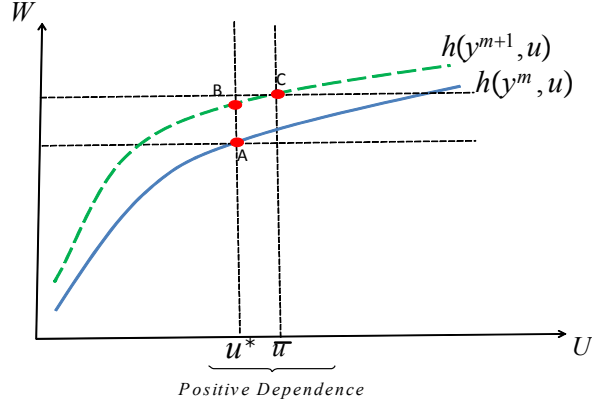
Figure 2: Restriction S-LDRM as well as Restriction LDRM are satisfied (A<B<C) around the region of U exhibiting positive dependence.

For example, college graduates may be different from high school graduates in terms of unobservable ability (U) when other observed characteristics are identical. It may be the case that individuals with very low ability are not allowed to get into college due to low test scores, on the other hand, individuals with extremely high ability may not choose to go to college if they have better options that will lead to higher income. The case in which our model is not applicable is when education is so deterimental that the hypothetical wage with one more year of education is smaller than that without it, among those with "similar" ability. On the other hand, S-LDRM assumes that wage with one more education needs to be larger or equal to than without it among the "same" level of ability if more able individuals choose to get educated more.

<Figure 2>, <Figure 3>, and <Figure 4> are drawn for the case where the unobserved elements are positively dependent in the range specified in $\boldsymbol{U}$ and $\boldsymbol{V}$. The lines need to be increasing in $\boldsymbol{U}$ showing positive dependence under Restriction QCFA. Restriction S-LDRM specified that $h(y^{m+1}, u^*) \geq h(y^m, u^*)$ (A<B), thus, by monotonicity of h with respect to u, $h(y^{m+1}, \overline{u}) \geq h(y^{m+1}, u^*) \geq h(y^m, u^*)$, for $\overline{u} \geq u^*$ (A>B>C). <Figure 2> shows the case in which Restriction S-LDRM is satisfied (A>B>C). <Figure 3> shows the case in which Restriction S-LDRM fails (A>B), but Restriction LDRM holds (A<C), and <Figure 4> shows the case in which Restriction LDRM fails (A$\geq B$ and A>C).

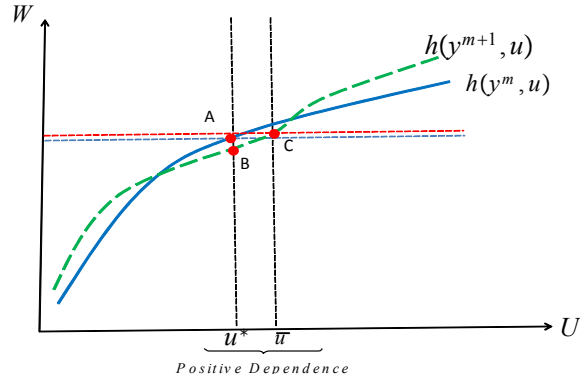Figure 3: Restriction LDRM is satisfied (C>A), although Restriction S-LDRM fails to hold (A>B) around the region of U exhibiting positive dependence.
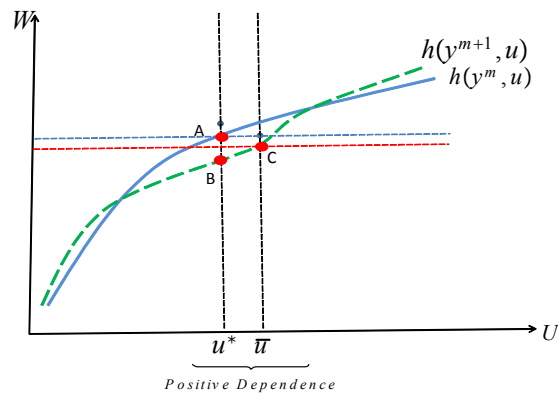

Figure 4: Both Restriction LDRM and Restriction S-LDRM fail to hold (A>C>B) around the region of U exhibiting positive dependence.

3 Main Results

3.1 Ordered Discrete Endogenous Variables

3.1.1 Bound on the Value of the Structural Relation

We first consider the case with an ordered discrete endogenous variable. The following interval identification of the value, $h(y^m, u^*)$ can be established for $m \in \{1, 2, ..., M - 1\}$, where $u^* \equiv Q_{U|VZ}(\tau_U|\tau_V, z)$. For $m = M$, the bound in Theorem 1 is not applied[20].

Theorem 1 Under Restriction QCFA,CQ-I,RC, and LDRM, the following holds for $m \in \{1, 2, ..., M - 1\}$ and $\tau \equiv \{\tau_U, \tau_V\}$

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

The interval is **not** empty.
Proof. See Appendix A2. ∎

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

3.1.2 Sharpness

Suppose that the value of the structural feature is identified by a set. Then *all* distinct "*admitted*" structures that are "*observationally equivalent*" to the true structure produce values of the structural feature that are contained in the identified set. All such structures that generate a point in the set, are indistinguishable by data. A sharp identified set contains *all and only* such values that are generated by admitted and observationally equivalent structures.

Common support restriction is imposed for sharpness.

---

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

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

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

Proof. See Appendix A3. ∎

3.1.3 Testable implications of the Model

Since identification results would be reliable only if the restrictions imposed were satisfied, it would be more credible if there is a way to convince that the restrictions imposed by the model were in fact ture description of the structure and the data. In this section testable implications are derived so that the validity of some of the restrictions is examined.

Lemma 1 reports the implications on the observed distribution, more specifically, on $Q_{W|YZ}(\tau_U|y^m, z_m')$ and $Q_{W|YZ}(\tau_U|y^{m+1}, z_m'')$. It is required to regularize the variation between Y and Z by the following rank condition to derive Lemma 1.

Restriction RC^{L1} (Rank Condition for Lemma 1) There exist instrumental values of Z, $\{z_m', z_m''\}$, such that

$$P^m(z_m') - P^{m-1}(z_m') \geq P^{m+1}(z_m'') - P^m(z_m''), \qquad (\text{RC}^{L1})$$

for $m \in \{1, 2, ..., M - 1\}$.With $P^0(z) = 0$ and $P^M(z) = 1$, for the binary case with $M = 2$, this condition is statated as

$$P^1(z_1') + P^1(z_1'') \geq 1.$$

Restriction RC^{L1} has implications on the conditional probability mass of Y on Z. Restriction RC^{L1} can be equivalantly expressed as $p_m(z_m') \geq p_{m+1}(z_m'')$, for $m \in \{1, 2, ..., M - 1\}$. Lemma 1 states the observable implications of the Model LDRM.

Lemma 1 Under Restriction QCFA,CQ-I,RC^{L1}, and LDRM, we observe locally in **U** and **V**,

$$
\begin{aligned}
Q_{W|YZ}(\tau_U|y^m, z_m') &\leq Q_{W|YZ}(\tau_U|y^{m+1}, z_m'') \text{ iff } PDPR, \text{ and} \\
Q_{W|YZ}(\tau_U|y^m, z_m') &\geq Q_{W|YZ}(\tau_U|y^{m+1}, z_m'') \text{ iff } NDNR, \\
for\ m &\in \{1, 2, ..., M - 1\}.
\end{aligned}
$$

Proof. See Appendix A6. ∎

According to Lemma 1, by comparing $Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)$ with $Q_{W|YZ}(\tau_U|y^m, z'_m)$ for given τ_U, it can be determined whether PDPR or NDNR is implied by the model locally in **U** and **V**. We conclude that if we observe $Q_{W|YZ}(\tau_U|y^m, z'_m) \leq Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)$, PDPR is implied, and if we observe $Q_{W|YZ}(\tau_U|y^m, z'_m) \geq Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)$, NDNR is implied. With this result in hand we can further find a way to investigate data to detrmine whether all the restrictions imposed in the model hold true. Another Rank Condition on the variation between Y and Z is required.

Restriction RC^{L2} (Rank Condition for Lemma 2) There exists a instrumental value of Z, z, such that

$$P^m(z) \geq \frac{1}{2}\left[P^{m-1}(z) + P^{m+1}(z)\right], \tag{RC^{L2}}$$

for $m \in \{1, 2, ..., M-1\}$. With $P^0(z) = 0$ and $P^M(z) = 1$, for the binary case where $M = 2$, this condition is statated as

$$P^1(z) \geq \frac{1}{2}.$$

For example, if the probability mass at $Y = y^m$ is not too small relative to at other points, RC^{L2} holds. Lemma 2 states the observable implications of PDPR and NDNR without relying on Restriction CQ-I. Once Restriction RC^{L2} is verified from data, under Restriction QCFA the implications of LDRM can be derived as follows.

Lemma 2 Suppose Restriction QCFA is satisfied. For any z that satisfies Restriction RC^{L2}, it can be shown that locally **U** and **V** we observe

$$
\begin{aligned}
Q_{W|YZ}(\tau_U|y^m, z) &\leq Q_{W|YZ}(\tau_U|y^{m+1}, z) \text{ iff } PDPR \text{ and} \\
Q_{W|YZ}(\tau_U|y^m, z) &\geq Q_{W|YZ}(\tau_U|y^{m+1}, z) \text{ iff } NDNR, \\
for\ m &\in \{1, 2, ..., M-1\}.
\end{aligned}
$$

Proof. See Appendix A7. ■

From Lemma 1 and Lemma 2 we can state the following testable implications.

Theorem 3 Suppose Restriction QCFA is satisfied and that $\{z'_m, z''_m\}$ satsify RC and RC^{L1}. Then we conclude the following : for any z that satisfies RC^{L2},

$$(i) \text{ if } Q_{W|YZ}(\tau_U|y^m, z'_m) \leq Q_{W|YZ}(\tau_U|y^{m+1}, z''_m),$$
$$\text{then } Q_{W|YZ}(\tau_U|y^m, z) \leq Q_{W|YZ}(\tau_U|y^{m+1}, z)$$
$$\text{and}$$
$$(ii) \text{ if } Q_{W|YZ}(\tau_U|y^m, z'_m) \geq Q_{W|YZ}(\tau_U|y^{m+1}, z''_m),$$
$$\text{then } Q_{W|YZ}(\tau_U|y^m, z) \geq Q_{W|YZ}(\tau_U|y^{m+1}, z),$$
$$for\ m \in \{1, 2, ..., M-1\}.$$

Proof. This is by Lemma 1 and Lemma 2. If we observe $Q_{W|YZ}(\tau_U|y^{m+1}, z''_m) \geq Q_{W|YZ}(\tau_U|y^m, z'_m)$, PDPR is implied by Lemma 1. Then by Lemma 2 we need to observe $Q_{W|YZ}(\tau_U|y^m, z) \leq Q_{W|YZ}(\tau_U|y^{m+1}, z)$ locally in $\mathbf{U}$ and $\mathbf{V}$. The same logic applies to the case in which we observe $Q_{W|YZ}(\tau_U|y^m, z'_m) \geq Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)$. $\blacksquare$

If all the restrictions imposed in Model LDRM were true description of the structure, then Theorem 3 would hold. However, the fact that Theorem 3 holds does not mean that all the restrictions are satisfied since both lemmas assume that Restriction QCFA holds. If Theorem 3 fails to hold then it indicates some of the restrictions in the model are violated locally in $\mathbf{U}$ and $\mathbf{V}$, nevertheless, it is impossible to tell which specific restriction(s) is(are) violated. In other words, Thorem 3 can be used to *refute* the restrictions imposed in the model LDRM, not to *confirm* them.

3.1.4 Many Instrumental Values, Overidentification, and Refutability

If there are many pairs of values of Z that satisfy Restriction RC (overidentification), then each pair defines the causal effect for a *different* subpopulation defined by each pair. Taking intersection of each identified set cannot be a sharp identified set as is discussed Lee (2011). To use all the information available from data and to justify taking intersection of each set defined by distinct pairs of values of Z in producing *a sharp identified set* in this case, a different restriction is imposed.

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

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

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

$$Q_m^L(\tau, y^m, \mathcal{Z}_m) \leq h(y^m, u^*) \leq Q_m^U(\tau, y^m, \mathcal{Z}_m)$$

$$\text{where } u^* = Q_{U|VZ}(\tau_U|\tau_V, z),$$

$$\text{for some } \tau_U \in (0, 1) \text{ and } \tau_V \in \mathbb{V} \equiv \cap_{\overline{z}_m} \mathbf{V}_m(\overline{z}_m)$$

$$Q_m^L(\tau, y^m, \mathcal{Z}_m) = \max_{\overline{z}_m} \ q_m^L(\tau, y^m, \overline{z}_m), \ \overline{z}_m \in \mathcal{Z}_m$$

$$Q_m^U(\tau, y^m, \mathcal{Z}_m) = \min_{\overline{z}_m} \ q_m^U(\tau, y^m, \overline{z}_m), \ \overline{z}_m \in \mathcal{Z}_m$$

$$q_m^L(\tau, y^m, \overline{z}_m) = \min\{Q_{W|YZ}(\tau_U|y^m, z_m'), Q_{W|YZ}(\tau_U|y^{m+1}, z_m'')\}$$

$$q_m^U(\tau, y^m, \overline{z}_m) = \max\{Q_{W|YZ}(\tau_U|y^m, z_m'), Q_{W|YZ}(\tau_U|y^{m+1}, z_m'')\}.$$

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

Proof. Idenitified intervals for each pair $\overline{z}_m \in \mathcal{Z}_m$, are shown in Theorem 1. The bound in this corollary is found by taking intersection of all such identified intervals. This intersection bound is sharp. The same sharpness proof of Thorem 2 applies with some modification in (S2) constructed in the proof in Appendix. When there exist many instrumental values that satisfy the rank condition, RC, the partition, $\{P^l\}_{l=1}^M$ defined in the proof of Theorem 2 can be re-defined as the following :

$$P^l = \begin{cases} \min_{z \in SUPP(Z)}\{P^l(z)\}, & if \quad l < m-1 \\ \max_{z \in SUPP(Z)}\{P^l(z)\}, & if \quad l > m \end{cases}$$

$$P^{m-1} = \min_{z \in \overline{z}_L}\{P^m(z)\}$$

$$P^m = \max_{z \in \overline{z}_U}\{P^m(z)\},$$

$$\text{where } \overline{z}_L \equiv \{z_L : z_L = \min \overline{z}_m, \overline{z}_m \in \mathcal{Z}_m\}$$

$$\overline{z}_U \equiv \{z_U : z_U = \max \overline{z}_m, \overline{z}_m \in \mathcal{Z}_m\}$$

$$\mathcal{Z}_m \equiv \{\overline{z}_m : P^m(z_m') \leq \tau_V \leq P^m(z_m''), \text{with } \overline{z}_m = \{z_m', z_m''\}\}.$$

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

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

Suppose that $\mathbb{V} \neq \varnothing$. Then the bound defined by Corollary 1 should be interpreted as causal effects for the subpopulation defined by $\mathbb{V}$. If $\mathbb{V} = \varnothing$, no causal interpretation would be possible, even though the intersection bound may not be empty since the subpopulation that is affected by the change in the values of Z does not exist. If $\mathbb{V} \neq \varnothing$, but the intersection bound is empty, then this means that some of the restrictions in the model are not satisfied. However, which restrictions are misspecified is not known by the fact that the identified set is empty. This way one can falsify the econometric model, rather than a specific restriction.

3.1.5   Bound on the Partial Difference

Theorem 4 reports the result on the partial difference with respect to the ordered discrete endogenous variable.

Theorem 4 Under Restriction QCFA,C-QI,RC,and LDRM, the following holds for $m \in \{1, 2, ..., M-2\}$ and $\tau \equiv \{\tau_U, \tau_V\}$

$$
\begin{aligned}
\Delta_{m,m+1}^{L} &\leq h(y^{m+1}, u^*) - h(y^m, u^*) \leq \Delta_{m,m+1}^{U} \\
\text{where } u^* &= Q_{U|VZ}(\tau_U | \tau_V, z), \\
\text{for some } \tau_U &\in (0,1) \text{ and} \\
\tau_V &\in V_m(\overline{z}_m) \cap V_{m+1}(\overline{z}_{m+1}) \\
\Delta_{m,m+1}^{L} &= q_{m+1}^{L}(\tau, y^{m+1}, \overline{z}_{m+1}) - q_m^{U}(\tau, y^m, \overline{z}_m) \\
\Delta_{m,m+1}^{U} &= q_{m+1}^{U}(\tau, y^{m+1}, \overline{z}_{m+1}) - q_m^{L}(\tau, y^m, \overline{z}_m) \\
&\quad \text{with } q_k^{L}(\tau, y^k, \overline{z}_k), q_k^{U}(\tau, y^k, \overline{z}_k), \\
\text{and } \overline{z}_k, \ k &= m, m+1 \text{ defined in Theorem 1.}
\end{aligned}
$$

The interval is **not** empty.
Proof. It follows from Theorem 1. ■

If either the upper bound, Δ_{mm+1}^{U}, is negative, or the lower bound, Δ_{mm+1}^{L}, is positive, then the sign of the patial difference, that is, the ceteris paribus effect of changing Y, can be identified. This result does *not* apply to a binary endogeneous variable in which case will be discussed in the next subsection. Stronger version of LDRM restriction is required to identify the sign of partial difference.with respect to a binary endogenous variable

3.2   Binary Endogenous Variable

Although in many empirical studies, the distribution of the treatment effects can deliever 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.[21] In this section we apply the LDRM model to a binary endogenous variable and identify the ceteris paribus impact of the binary variable, or treatment effects. As Chesher (2005) noted, models for an ordered discrete endogenous variable can not directly be applied to binary endogenous variables due to the "unordered" nature of binary variables, however, our model imposes a sense of order to a binary endogenous variable, which enables the model to identify the partial difference.

---

[21]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.

3.2.1 Bound on the Value of the Structural Relation

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

Corollary 2 Under Restriction QCFA,C-QI,RC,and LDRM the following holds for $y^1 = 0$ and $y^2 = 1$, $z \in \overline{z} = \{z', z''\}$, and $\tau \equiv \{\tau_U, \tau_V\}$,

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

The bound is sharp.
Proof. See Appendix A4. ∎

Although the identified intervals for $h(1, u^*)$ is the same as that for $h(0, u^*)$, this is still informative in the sense that the identified interval restricts the possible range that the values $h(1, u^*)$ and $h(0, u^*)$ lie in, and that the sign of $h(1, u^*) - h(0, u^*)$ can be identified as either the upper bound or the lower bound is zero by strengthening Restriction LDRM to Restriction S-LDRM.

3.2.2 Bound on Partial Difference of the Structural Relation

Corollary 2 and Lemma 1 are used to recover heterogeneous treatment responses. Theorem 5 states the partial identification result of heterogeneous treatment effects.To define the bound on partial difference, Restriction S-LDRM is required.

Theorem 5 Under Restriction QCFA,C-QI,RC,RC^{L1} and S-LDRM, $h(1, u^*) - h(0, u^*)$ is identified by the following interval:

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

Proof. Suppose $Q_{W|YZ}(\tau_U | 1, z'') \geq Q_{W|YZ}(\tau_U | 0, z')$. From Corollary 2 we have

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

then we have

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

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

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

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

Whether the treatment effect is positive or negative can be determined by data from the sign of $Q_{\tau_U}^{\Delta}$ based on Theorem 5. If $Q_{\tau_U}^{\Delta} > 0$, then

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

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

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

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

3.3   Discussion

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

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

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

When the endogenous variable is discrete[22], Imbens and Newey (2009)'s approach does not have identifying power.[23] Chesher (2003)'s QCFA fails to produce point identification since the one-to-one mapping between the endogenous variable and the unobserved variable that exists with a continuous endogenous variable does not exist with a discrete endogenous variable. Rather, with a discrete endogenous variable, a specific value of the endogenous vairiable maps into a set of values of the unobservable variable. Without imposing further restrictions, a sharp bound cannot be defined. Chesher (2005) suggested to impose monotonicity of $F_{U|V}(u|v)$ in v and a rank condition to define a bound on the value of the structural funciton. Jun, Pinkse, and Xu (2010) imposed the same monotonicity restriction on $F_{U|V}(u|v)$, but impose a different rank condition.

3.3.2 Nonparametric *Shape* Restrictictions

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

Alternatively, one could adopt restrictions on the structure. Apart from Chesher (2005) and Jun, Pinkse, and Xu (2011)'s monotonicity imposed on the distribution of the unobservables, Manski and Pepper (2000) and Bhattacharya, Shaikh and Vytlacil (2008) adopt certain monotonicity restrictions in the structural relations. Under the 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.

Manski and Pepper (2000) develop their arguments by assuming that both selection and response are increasing, but by assuming that both are decreasing also leads to identification of average effects. In contrast, with Restriction LDRM, 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. Restriction LDRM allows the direction (either PDPR or NDNR) of the match to vary over the support of U, while the MTR-MTS imposed on the mean - either positive response with positive selection or negative response with negative selection. Roughly speaking, the LDRM restriction can be described as a *local (quantile)*[24] version of MTR-MTS.

---

[22]Several studies adopted the potential outcomes approach. See Heckman, Florens, Meghir, and Vytlacil (2008) for average effects of continuous treatment, and Angrist and Imbens (1995), and Nekipelov (2009) for average effects of multi-valued discrete treatment.

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

[24]Restriction MTR-MTS is regarding the *mean*, while Restriction LDRM is regarding each point (locally) in the support of the unobserved variable, U. Every point in the support of U can be

Manski and Pepper (2000) identifies average treatment effects, thus the heterogeneity in treatment effects can be found for the subpopulation defined by the observed characteristics, while LDRM model can recover heterogeneity in treatment effects *even* among observationally identical individuals.

Bhattacharya, Shaikh and Vytlacil (2008) compare Shaikh and Vytlacil (2011) bounds with Manski and Pepper (2000)[25] by applying them to a binary outcome - binary endogenous variable case. Bhattacharya, Shaikh and Vytlacil (2008)'s bounds are found under the restriction that the binary endogenous variable is determined by an IV monotonically. When IV, Z, and Y are binary, their monotonicity is equivalent to the monotonicity here. Note also that when Y is binary, we can always reorder 0 and 1 due to the "unordered nature" of a binary variable. In contrast with their claim, when Manski and Pepper (2000) is applied to a binary case, the direction of the monotonicity of response and selection does not have to be determined a priori[26]. Data will inform about the direction of the monotonicity, however, the direction of MTR and MTS should be matched in a certain way[27].

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

---

expressed as quantiles of the distribution of U.

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

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

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

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

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

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

The counterfactual $E(y(1)|z=0)$ 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.

3.3.3 Heterogeneous Causality Measured by Partial Differences

The major object of interest in this paper is the partial difference of the structural quantile function, $h(1, u^*) - h(0, u^*)$. The value u^* is unknown, but is assumed to be $u^* = Q_{U|VZ}(\tau_U | \tau_V, z)$ for some $\tau_U, \tau_V \in (0, 1)$. $h(1, u^*) - h(0, u^*)$ is interpreted as a ceteris paribus impact of Y. When the value of Y changes from 1 to 0, the value of U would change as well if there exists endogeneity. This is in contrast with other identification results in additively nonseparable models. Other studies identify the values of a *nonadditive* structural function, but their results do not guarantee identification of partial differences.

3.3.4 Rank Condition and Causal Interpretation

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

3.3.5 Applicability to Regression Discontinuity Designs (RDD) and Randomised Trials

Recently, many studies (see Lee and Lemieux (2009), for a survey) adopted regression discontinuity design (RDD) to measure causal effects. Under this design if the continuity condition at the threshold point of the "forcing variable" holds, the causal effects of individuals with the forcing variable just above and below the threshold point are shown to be identified. When the RDD is available, our rank condition[28] is guaranteed to hold, thus, as long as Restriction LDRM is applicable in the context of interest, the proposed model can be applicable to an RD design even when all other variables are not continuous in the treatment - determining variable at the threshold.[29]

3.3.6 Tests of Homogeneous Signs

Homogeneous signs can be tested by adopting existing results on stochastic dominance of order 1 as the null of homogeneity can be expressed as follows, for $Z = z', z''$,

$$H_0 : F_{W|YZ}(w|Y = 1, Z = z'') \geq F_{W|YZ}(w|Y = 0, Z = z'), \forall w.$$

---

[28]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.

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

See, for example, Barret and Donald (2003), for recent developments of test of stochastic dominance of various orders.

3.3.7 Discrete Data

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

4 Empirical Illustration - Heterogeneous *Individual* Treatment Responses

By heterogeneous treatment responses we mean idiosyncratic treatment effects even after accounting for observed characteristics[30]. Several studies allowed for individual heterogeneity in response, yet, identification is achieved by integrating out the heterogeneity in these studies.[31] Average responses may hide heterogeneity in response and information regarding the distributional consequences of a policy would be lost. We demonstrate how the "partial" information, the signs and the bounds of treatment effects, not the exact size of them, regarding who benefits can be recovered from data when "who" is indicated by individual observed characeristics and the ranking in the distribution of the unobserved characteristic. This is illustrated by examining the effects of the Vietnam-era veteran status on the civilian earnings using the data used in Abadie (2002)[32]

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

Let W be annual labour earnings, Y be the veteran status, and Z be the binary variable determined by draft lottery. Age, race, and gender are controlled so that the subgroup considered is observationally homogenous. The unobserved variables

---

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

[31]The standard linear IV model cannot identify heterogeneus treatment effects. See Heckman and Navarro (2004) and Heckman and Urzua (2009). For identification under heterogeneous responses see Heckman, Urzua, and Vytlacil (2006) for binary endogenous variable, and Florens, Heckman, Meghir, and Vytlacil (2008), Imbens and Newey (2009),and 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 (1999) for example. The averaged objects however can exhibit a certain degree of heterogeneity by allowing for treatment heterogeneity.

[32]A sample of 11,637 white men, born in 1950-1953, from the March Current Population Surveys of 1979 and 1981-1985 is used. The data are obtainable in Angrist Data Archive : http://econ-www.mit.edu/faculty/angrist/data1/data

U and V indicate scalar indices for "earnings potential" and "participation preference"/"aptitude for the army" each. There can be many factors that determine these indices, but we assume that these multi-dimensional elements affect the outcome only through a "scalar" index.

4.1.1 Selection on Unobservables

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

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

As in Angrist (1990) the Vietnam era draft lottery is used as an instrument to identify the effects of veteran status on earnings. The lottery was conducted every year between 1970 and 1974. The lottery assigned numbers from 1 to 365 to dates of birth in the cohorts being drafted. Men with the lowest numbers were called to serve up to a ceiling[33] which was unknown in advance. We construct a binary IV based on the lottery number. It is assumed that this IV is not a determinant of earnings, and the unobserved scalar indices are independent of draft eligibility[34].

4.1.3 Rank Conditions - RC, RC^{L1}, and RC^{L2}

To apply the identification results in Theorem 5, we investigate first whether the data satisfy Restriction RC in the model. The participation rate among the draft-non-eligible ($Z = 0$) is about 0.14 and the participation rate among the eligible is

---

[33]See Angrist (1990) for more details.

[34]There has been some discussion on whether individuals' draft lottery numbers caused their behavior, e.g. some men could have volunteered in the hope of serving under better terms and gaining some control over the timing of their service once the lottery numberw were known. If those who change their behavior according to their draft lottery number show certain patterns in their unobserved factors, then the quantile invariance restriction may be violated.

0.22. Thus, Restriction RC is satisfied as

$$
\begin{aligned}
P^1(1) &\equiv P(Y = 0 | Z = 1, X = x) = 0.78 \qquad \text{(RC)}\\
&< P^1(0) \equiv P(Y = 0 | Z = 0, X = x) = 0.86.
\end{aligned}
$$

That is, $z_1' = 1$ and $z_1'' = 0$ in this example and age, gender, and rac are controlled. The compliers[35] (or draftees) are defined as those whose V-ranking is between 78% and 86%.

Both Restriction RC^{L1} and Restriction RC^{L2} are satisfied, which allows us to use Theorem 3 to refute some of the restrictions in Model LDRM. RC^{L1} is satisfied as $P^1(z_1') + P^1(z_1'') = 0.78 + 0.86 \geq 1$. For both values of Z, RC^{L2} is also satisfied as $P^1(z_1') = 0.78 \geq \frac{1}{2}$, and $P^1(z_1'') = 0.86 \geq \frac{1}{2}$

4.1.4 Signs of the Effects - Use of Theorem 5

We use Hansen (2004) in estimating the distribution functions and the quantiles are found based on the estimated distribution function following the definition of a quantile. Smoothing is suggested in Hansen (2004) for efficiency gain in finite samples. Soomthing is also reasonable in this context as it is more likely to believe that treatment effects for individuals in similar ranks would not vary drastically. Thus, the signs of treatment effects are not expected to show a large variation across different ranks. The optimal bandwidth is selected following Hansen (2004).

4.1.5 Causal Interpretation

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

Our quantile based analysis reveals that when age, gender, and race are controlled the veteran status had positive causal impacts for individuals with low earnings potential, but negative causal impacts for individuals with high earnings potential (see Figure 5). The results in <Figure 5> show that the sign is positive for those whose U- rank is less than 75%, while negative sign for those ranked higher than that.[36] The lifetime cost of military service may be larger than the benefits provided by the government for those with high earnings potential, while the benefits provided may be sufficient for those with low earnings potential. Considering the fact that benefits

---

[35] Note that the V- ranking is never observed, so we cannot tell whether an individual is a complier or not.

[36] The results in <Figure 5> are interpreted as the causal effects for those who change their participation decision as the value of Z changes.
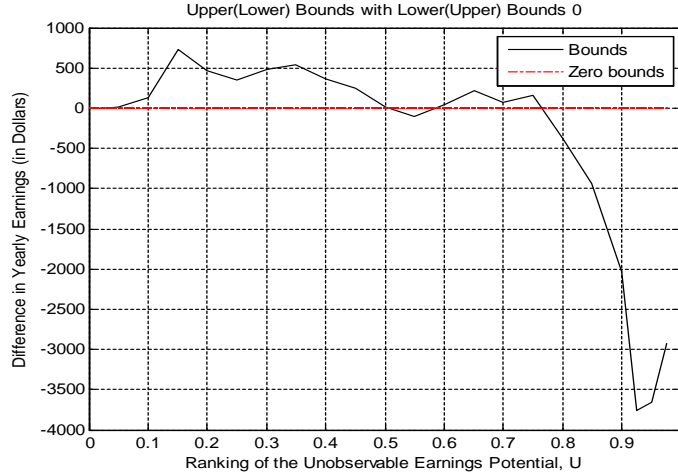
Figure 5: LDRM bounds on individual treatment effects of Vietnam-era veteran status among observationally identical individuals.

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

This can be compared with the results using QTE. By applying his identification results of the marginal distribution of the potential outcomes for compliers, Abadie (2002) reports that military service during the Vietnam era reduces lower quantiles of the earnings distribution, leaving higher quantiles unaffected. The information from the marginal distribution of the potential outcomes (for compliers) may be used to recover QTE, however, it does not reveal information on individual-specific impact on earnings of Vietnam-era veteran experience as individuals'ranking in each marginal distribution can be different.

4.1.6   Refutability of the Model LDRM - Use of Theorem 3

As both Restriction RC^{L1} and Restriction RC^{L2} are satisfied, Thorem 3 is applied to examine whether all the restrictions in the model are satisfied. <Figure 6> shows that some of the model's restrictions are violated for those ranked between 20–45% and 65-75%. Theorem 3 states that in the range in which positive treatment is implied, $Q_{W|YZ}(\tau_U|1,z) \geq Q_{W|YZ}(\tau_U|0,z)$ and in the range where negative treatment is implied, $Q_{W|YZ}(\tau_U|1,z) \leq Q_{W|YZ}(\tau_U|0,z)$, for z that satisfies RC^{L2}. <Figure 6> is drawn for $Z = 1$ as there is small observation when $Z = 0$. As is discussed, Thorem 3 can be used to refute, rather than confirm the model. That is, even those whose U-rank is other than 20-45% or 65-75%, it is still possible that Restriction QCFA is violated. Restriction QCFA needs to be assumed to derive Lemma 1 and Lemma 2, therefore, it is not refutable.
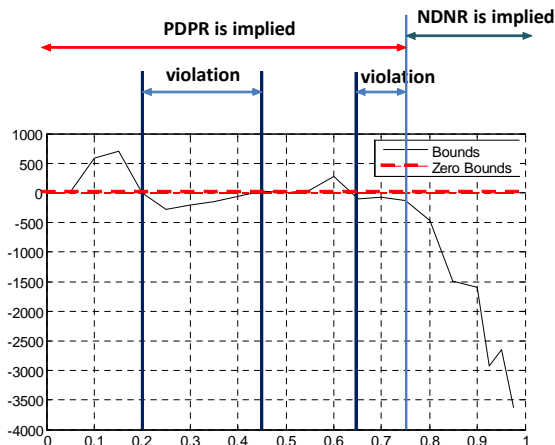
Figure 6: Refutability of the Model : $Q_{W|YZ}(\tau_U|1,1) - Q_{W|YZ}(\tau_U|0,1)$ is indicated by the black line.

5   Conclusion

The presence of endogeneity and discreteness of the endogenous variable cause 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. A refutable model that set identifies certain structural features is proposed when one of the regressors is ordered discrete. The model is applied to a binary endogenous variable. This structural approach turns out to be useful in defining the bounds on heterogeneous individual treatment effects, which have not been studied so far under the structural framework without parametric assumptions.

The set identification result of this paper is applied to recover heterogeneous impacts of the Vietnam-era military service on earnings later in life. As we can see in this example, average effects may miss much information. Even though the proposed model can give only partial information on the individual causal effect, this may be useful in some economic contexts, especially when the sign of the effects may be varying across individuals with different characteristics. The causal interpretation is justified on the group of compliers defined by the pair of instrumental values that satisfy the rank condition. The information on the signs of individual treatment effects is crucial if they vary across the population, since in such a case the average effects would be smaller as different effects with different signs will be canceled out leading to a misleading conclusion. The model can also be used for robustness checks for whether there exists any heterogeneity in causal responses.

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

[4] Angrist, J. and A. Krueger (1991), "Does Compulsory School Attendance Affect Schooling and Earnings?," *Quarterly Journal of Economics,* 106, 979-1014.

[5] Angrist, J. and G. Imbens (1995), "Two-stage least squares estimation of average causal effects in models wth variable treatment intennsity," *Journal of the American Statistical Association*, 90 (430), 431-442.

[6] Angrist, J. (2004), "Treatment effect heterogeneity in theory and practice," *The Economics Journal*, 114, C52-C83.

[7] Barret, G. and S. Donald (2003), "Consistent tests for stochastic dominance," *Econometrica*, 71(1), 71–104.

[8] 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.

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

[10] 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).

[11] Card, D. (1999), "The causal effect of education on earnings," *Handbook of Labour Economics*, 1801–1863.

[12] Chernozhukov, V. and C. Hansen (2005), "An IV Model of quantile treatment effects," *Econometrica*, Vol.73, No. 1, 245-261.

[13] Chernozhukov, V., I. Fernandez-Val, and B. Melly (2010), "Inference on counterfactual distributions in nonseparable Models," mimeo.

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

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

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

[17] Chesher, A. (2009), "Excess heterogeneity, endogeneity, and index restrictions," *Journal of Econometrics*, 152, 35-47.

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

[19] Fan, Y. and S. Park (2010), "Sharp bounds on the distribution of treatment effects and their statistical inference," *Econometric Theory*, 26, 931-951.

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

[21] 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.

[22] Brandson, B, M. Frolich, and B. Melly (2008), "Quantile treatment effects in the regression discontinuity design," *Journal of Econometrics*, 168, 382-395.

[23] Hansen, B (2004), "Bandwidth selection for nonparametric distribution estimation," mimeo.

[24] Hahn, J. and G. Ridder (2011), "Conditional moment restrictions and triangular simulatenous equations," *Review of Economics and Statistics*, forthcoming.

[25] Heckman, J., J. (2001), "Micro data, heterogeneity, and the evaluation of public policy: Nobel lecture," *Journal of Political Economics*,

[26] 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.

[27] 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.

[28] 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.

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

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

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

[32] Imbens, G. and D. Rubin (1997), "Estimating outcome distributions for compliers in instrumental variable models," *Review of Economic Studies*, 64, 555-574.

[33] Imbens, G. and W. Newey (2009), "Identification and estimation of triangular simultaneous equations models without additivity," *Econometrica*, 77(5),1481 - 1512.

[34] Jun, S., J. Pinkse, and H. Xu (2011), "Tighter bounds in triangular system," *Journal of Economcetrics*, 161(2),122–128.

[35] Kitagawa, T. (2009), "Identification region of the potential outcome distributions under instrument independence," cemmap working paper.

[36] Lee, D. and T. Lemieux (2009), "Regression discontinuity designs in economics," NBER working paper.

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

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

[39] Manski, C. and J. Pepper (2000), "Monotone instrumental variables : with an application to the returns to schooling," *Econometrica*, 68(4), 997-1010.

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

[41] Matzkin, R. (2007), "Nonparametric identification," *Handbook of Econometrics*, 5307-5368.

[42] Matzkin, R. (2008), "Identification of nonparametric simultaneous equations," *Econometrica*, 76, 945-978.

[43] Nekipelov, D. (2009), "Endogenous multi-valued treatment effect model under monotonicity," mimeo.

[44] Roehrig, C. (1988), "Conditions for identification in nonparametric and parametric models," *Econometrica*, Vol. 56, No. 2, pp. 433-447.

[45] Shaikh, A. and E. Vytlacil (2011), "Partial identification in triangular systems of equations with binary dependent variables," *Econometrica*, 79, 949-955.

[46] Vytlacil, E. and N. Yildiz (2007), "Dummy endogenous variables in weakly separable models," *Econometrica,* 75(3), 757-779.

Appendix - proofs

A.1 A note on the proofs

$\mathbf{U}$ and $\mathbf{V}$ are defined in Section 2.1. For given $\tau_V \in \mathbf{V}$, any point $u \in \mathbf{U}$ can be expressed as a quantile of the conditional distribution of U given V and Z,

$$u = Q_{U|VZ}(\tau_U|\tau_V, z)$$

for some $\tau_U \in (0,1)$. The following observations, **(O1),(O2),** and **(O3)**, as well as definition (*), will be used in the proofs.

(O1) For given $\tau_V \in \mathbf{V}$, for any $u \in \mathbf{U}$ there can be many pairs of (a,b) s.t

$$u = Q_{U|VZ}(\tau_U|\tau_V, z) = Q_{U|VZ}(a|b, z) \qquad (O1)$$
$$for \text{ some } \tau_U, a, b \in (0,1).$$

Note that any $u \in \mathbf{U}$ can be expressed in terms of the structural function by defining the inverse function of h. We use the definition of Chesher (2005) of the following

$$h^{-1}(y^m, w) \equiv \sup_u \{u : h(y^m, u) \le w\} \qquad (*)$$

with equality holding when $h(y^m, u)$ is strictly increasing in u as in Chesher (2005).

(O2) For a given structural function, h, and a given value of $Y = y^m$, an arbitrary $u \in (0,1)$ can be written as

$$u = h^{-1}(y^m, w)$$

for *some* $w \in SUPP(W)$ where h^{-1} is defined by (*)[37]. Let $w^1, w^2, ..., w^M$ be the values such that for given $u = h^{-1}(y^m, w)$

---

[37]If W is discrete, $h_c^{-1}(y, w)$ indicates only jumping points in U, not all the points of U.

$$w^l = h_c(y^l, h^{-1}(y^m, w)), \text{ for } l = 1, 2, ..., M. \tag{O2-1}$$

Note that $w^m = h_c(y^m, h_c^{-1}(y^m, w)) = w$. Then we can find w^l such that

$$u = h^{-1}(y^m, w) = h^{-1}(y^l, w^l) \text{ for some } l. \tag{O2}$$

(O3) Let $S^0 = \{h_0, F_{U|VZ}^0\}$ be the *true unknown* structure that generates the data we have, $F_{W|YZ}^0(w|y, z)$. The model LDRM is proposed to identify certain features of the true structure, S^0, by using the data, $F_{W|YZ}^0(w|y, z)$. Suppose that $w^* = h_0(y^m, u^*)$, where $u^* \equiv Q_{U|VZ}^0(\tau_U|\tau_V, z)$. Then note that u^* can be expressed in terms of the structural function by (*)

$$u^* \equiv Q_{U|VZ}^0(\tau_U|\tau_V, z) = h_0^{-1}(y^m, w^*). \tag{O3}$$

The proofs involve establishing the order of the values of the strucual function evaluated at different points. To establish inequalities between them, we use (O3). The properties of a quantile are used and they are expressed in terms of the structural function by using the definition (*) when necessary.

A.2 Proof of Theorem 1

Proof. We show the case with PDPR. The other case, NDNR, can be shown similarly. Suppose that $Q_{U|VZ}(\tau_U|v, z)$ is weakly increasing in $v \in \mathbf{V}$, then by Restriction LDRM, PDPR is assumed. By Lemma 2 in Chesher (2005) we have that the quantile of the distribution of the observed variables is bounded by values of the structural function as follows

$$
\begin{align}
h(y^m, Q_{U|VZ}(\tau_U|V_L, z'_m)) &\leq Q_{W|YZ}(\tau_U|y^m, z'_m) \tag{A2-1} \\
&\leq h(y^m, Q_{U|VZ}(\tau_U|P^m(z'_m), z'_m)) \\
h(y^{m+1}, Q_{U|VZ}(\tau_U|P^m(z''_m), z''_m)) &\leq Q_{W|YZ}(\tau_U|y^{m+1}, z''_m) \tag{A2-2} \\
&\leq h(y^{m+1}, Q_{U|VZ}(\tau_U|V_U, z''_m)).
\end{align}
$$

Let $u^* \equiv Q_{U|VZ}(\tau_U|\tau_V, z''_m)$, $\overline{u} \equiv Q_{U|VZ}(\tau_U|P^m(z''_m), z''_m)$, and $\underline{u} \equiv Q_{U|VZ}(\tau_U|P^m(z'_m), z'_m)$. Then $u^*, \overline{u}, \underline{u} \in \mathbf{U}$. As $Q_{U|VZ}(\tau_U|v, z)$ is weakly increasing in $v \in \mathbf{V}$, by Restriction RC we have

$$\underline{u} \leq u^* \leq \overline{u}.$$

Then because h is weakly increasing in u, we have

$$
\begin{align}
h(y^m, Q_{U|VZ}(\tau_U|\tau_V, z''_m)) &\leq h(y^m, \overline{u}) \tag{A2-3} \\
h(y^m, \underline{u}) &\leq h(y^m, Q_{U|VZ}(\tau_U|\tau_V, z'_m)).
\end{align}
$$

Note that $u^* \equiv Q_{U|VZ}(\tau_U|\tau_V, z''_m) = Q_{U|VZ}(\tau_U|\tau_V, z'_m)$ by Restriction C-QI. Therefore, from (A2-3) we have

$$h(y^m, \underline{u}) \leq h(y^m, u^*) \leq h(y^m, \overline{u}). \tag{A2-4}$$

34

By (A2-2) and Restriction LDRM we can find the upper bound on $h(y^m, u^*)$,

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

where the first inequality is due to Restriction LDRM (PDPR in this case) for $u^*, \overline{u} \in \mathbf{U}$ with $u^* \leq \overline{u}$, and the second inequality is due to (A2-2).

The lower bound on $h(y^m, u^*)$ can be found by (A2-1) and (A2-4) :

$$Q_{W|YZ}(\tau_U | y^m, z_m') \leq h(y^m, \underline{u}) \leq h(y^m, u^*),$$

where the first inequality is due to (A2-1), the second is due to (A2-4).

Similarly, when $Q_{U|VZ}(\tau_U | v, z)$ 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.3 Proof of Theorem 2 : Sharpness

Let SUPP(A) indicate the support of a random variable A and let $I(\tau, y^m, \overline{z}_m)$ denote the identified interval. What is required to show sharpness is to construct a structure $(S^c \equiv \{h_c, F_{U|VZ}^c(u|v, z)\})$ such that (a) for any value in the identified interval, $w^* \in I(\tau, y^m, \overline{z}_m)$, for $u^* \equiv Q_{U|VZ}^c(\tau_U | \tau_V, z)$, the structural function crosses w^*, that is, $w^* = h_c(y^m, u^*)$, (b) that the constructed structure is observationally equivalent to the true structure and (c) that the constructed structure is admitted by the LDRM model. In Part 1 we propose a structure $\{h_c, F_{U|VZ}^c(u|v, z)\}$ and in Part 2 we show (a),(b), and (c).The results hold for both continuous and discrete W.

Part 1 - Construction of an admitted and observationally equivalent structure

1-A Construction of a structural function

We consider the case in which $Q_{W|YZ}(\tau_U | y^m, z_m') \leq Q_{W|YZ}(\tau_U | y^{m+1}, z_m'')$. The structural function is constructed as follows : for *some* $v \in (P^{m-1}(z), P^m(z)]$

$$h_c(y^m, u) \equiv Q_{W|YZ}^0(\overline{\tau}_m | y^m, z), \text{ where } u = Q_{U|VZ}^c(\tau_U | \tau_V, z) = Q_{U|VZ}^c(\overline{\tau}_m | v, z), \quad \text{(S1)}$$

for $m = 1, 2, ..., M$. That is, $Q_{W|YZ}^0(\overline{\tau}_m | y^m, z)$ is assigned as the value of $h_c(y^m, u)$ by choosing $\overline{\tau}_m$ such that $u = Q_{U|VZ}^c(\tau_U | \tau_V, z) = Q_{U|VZ}^c(\overline{\tau}_m | v, z)$ for given v. Then $h_c(y^m, u)$ is a well-defined *function*.

1-B Construction of the conditional distribution of the unobservables

We partition $SUPP(V)$ by[38]

$$P^l = \begin{pmatrix} \max_{z \in SUPP(Z)}\{P^l(z)\} & \text{if} & l \neq m-1, m \\ \min_{z \in \overline{z}_m}\{P^m(z)\} & \text{if} & l = m-1 \\ \max_{z \in \overline{z}_m}\{P^m(z)\} & \text{if} & l = m \end{pmatrix}$$

$$l = 1, 2, ..., M, \text{ for } \overline{z}_m \text{ defined in Section 2.}$$

The conditioning value of Y is determined by the value of V as follows : for $l = 1, 2, ..., M$

$$Y = y^l \quad \text{if} \quad P^{l-1} < v \leq P^l$$
$$\text{with } P^0 = 0 \text{ and } P^M = 1.$$

Note that by (O2) any point $u \in (0, 1)$ can be expressed as $u = h_c^{-1}(y^k, \widetilde{w})$ for some y^k and $\widetilde{w}$. $F_{U|VZ}^c(u|v, z)$ is constructed as follows[39] : for given $z \in SUPP(Z)$,

$$F_{U|VZ}^c(u|v, z), \text{ for } u = h_c^{-1}(y^k, \widetilde{w}) \text{ and } P^{l-1} < v \leq P^l \qquad \text{(S2)}$$
$$\equiv \begin{cases} F_{W|YZ}^0(\widetilde{w}|y^l, z) & \text{if} & k = l, l+1 \\ F_{W|YZ}^0(w^l|y^l, z) & & o.w. \end{cases}$$

where $w^1, w^2, ..., w^M$ are found such that $w^l = h_c(y^l, h_c^{-1}(y^k, \widetilde{w}))$, for $l = 1, 2, ..., M$ and $w^k = h_c(y^k, h_c^{-1}(y^k, \widetilde{w})) = \widetilde{w}$ (as is specified in Appendix A1).

1-C Weakly increasing h in u and proper distribution, $F_{U|VZ}^c(u|v, z)$

Then h_c is weakly increasing in u. We show that as u increases, h weakly increases. Note that any point u can be expressed as a conditional quantile (as is noted in Appendix A1). The value of a conditional quantile can increase for two different reasons :

- First, fix $\overline{v}_m$, then $h_c(y^m, u)$ is weakly increasing in u since higher $\overline{\tau}_m$ implies higher $u = Q_{U|VZ}(\overline{\tau}_m|\overline{v}_m, z)$, as well as higher $Q_{U|YZ}(\overline{\tau}_m|y^m, z)$, which is the value assigned to $h_c(y^m, u)$. That is, as u increases, h weakly increases.

- Next, fix $\overline{\tau}_m$, if we observe higher u, then it is because of higher $\overline{v}_m$ if $F_{U|V}(u|\overline{v}_m, z)$ is non-increasing in $\overline{v}_m$ and lower $\overline{v}_m$ if $F_{U|VZ}(u|\overline{v}_m, z)$ is nondecreasing in $\overline{v}_m \in (P^{m-1}, P^m]$. However, regardless of the direction of monotonicity, for $\overline{v}_m \in (P^{m-1}, P^m]$, $Y = y^m$. Thus, the value of $\overline{v}_m$ does not affect the value of h_c. That is, for fixed $\overline{\tau}_m$, and Y, $h_c(y, u)$ is constant as u increases due to change in $\overline{v}_m$. That is, as u increases, h weakly increases.

Moreover, the proposed $F_{U|VZ}^c(u|v, z)$ is a proper distribution by the following arguments : since each $F_{W|YZ}^0(w|y^m, z)$, for all $m \in \{1, 2, ..., M\}$ is a proper distribution, $F_{W|YZ}^0(w|y^m, z)$

---

[38]This partition is chosen to be fixed irrespective of the value of Z.

[39]If W is discrete, h_c^{-1} indicates jumping points only. To define the values of $F_{U|VZ}$, other than in jumping points, we need to partition $(0, 1)$ by jumping points and assign the same value to the points in between jumping points (to all the points in each partition) as that which is assigned to the lower jumping point among the two jumping points defining the partition.

lies between zero and one, and is weakly increasing in w. Thus, the constructed distribution, $F_{U|VZ}^c(u|v,z)$, lies between zero and one, but to guarantee nondecreasing property of $F_{U|VZ}^c(u|v,z)$ in u, we need to show that as w increases, $u = h_c^{-1}(y,w)$ increases for given v and z. This follows from the definition of h_c^{-1} in (*) in Appendix A1 and the fact that h_c is weakly increasing in u.

Part 2

We now show (a),(b), and (c).

Part 2 - (a)[40] :

To show (a) note that under Restriction CSUPP, *any* point in the identified interval, $w^* \in I(\tau, y^m, \overline{z}_m)$ can be written as (see <Figure 7>)[41]

$$w^* = Q_{W|YZ}^0(\overline{\tau}_m|y^m, z_m') \text{ for some } \overline{\tau}_m \geq \tau_U. \tag{A3-3}$$

That is,

$$\overline{\tau}_m \leq F_{W|YZ}^0(w^*|y^m, z_m') \text{ for some } \overline{\tau}_m \geq \tau_U \tag{A3-4}$$

with equality holds when W is continuous.

Now consider $u = h_c^{-1}(y^m, w^*)$ and $v \in (P^{m-1}, P^m]$. From (S2)

$$F_{U|VZ}^c(h_c^{-1}(y^m, w^*)|v, z_m') \stackrel{(S2)}{=} F_{W|YZ}^0(w^*|y^m, z_m') \geq \overline{\tau}_m,$$

due to (A3-4), which implies, by definition of a quantile,

$$h_c^{-1}(y^m, w^*) = Q_{U|VZ}^c(\overline{\tau}_m|v, z_m') \text{ for some } v \in (P^{m-1}, P^m]. \tag{A3-5}$$

Note that $Q_{U|VZ}^c(\overline{\tau}_m|v, z_m')$ is not varying with $v \in (P^{m-1}, P^m]$ since $F_{U|VZ}^c$ is constant over the interval $v \in (P^{m-1}, P^m]$. This then implies by construction in (S1) the value of $Q_{W|YZ}^0(\overline{\tau}_m|y^m, z)$ is assigned as the value of $h_c(y^m, u)$, by choosing $\tau_U \in (0,1)$ and $\tau_V \in (P^{m-1}, P^m]$, s.t. $u = Q_{U|VZ}^c(\overline{\tau}_m|v, z_m') = Q_{U|VZ}^c(\tau_U|\tau_V, z)$, that is, for given v,

$$h_c(y^m, Q_{U|VZ}^c(\overline{\tau}_m|v, z_m')) = Q_{W|YZ}^0(\overline{\tau}_m|y^m, z) \tag{A3-6}$$

$$\text{and } u = h_c^{-1}(y^m, w^*) = u^* \tag{A3-7}$$

as $u = Q_{U|VZ}^c(\overline{\tau}_m|v, z_m') = Q_{U|VZ}^c(\tau_U|\tau_V, z) \equiv u^*$. Finally, from (A3-3) and (A3-6) we conclude that

$$w^* = h_c(y^m, u^*), \text{ for } u = Q_{U|VZ}^c(\overline{\tau}_m|v, z_m') = Q_{U|VZ}^c(\tau_U|\tau_V, z).$$

In other words, h_c passes through an arbitrary point, $w^* \in I(\tau, y^m, \overline{z}_m)$.[42]

---

[40]for an arbitrary point in the identified interval, $w^* \in I(\tau, y^m, \overline{z}_m)$, such that $w^* = h_c(y^m, u^*)$, $u^* \equiv Q_{U|VZ}^c(\tau_U|\tau_V, z) \in \mathbf{U}$, we define $w^l = h_c(y^l, h_c^{-1}(y^m, w^*))$, with w^m
$= w^* \ l = 1, 2, ..., M$ in Part 1-B.

[41]Alternatively, one can find $\overline{\tau}_m$ such that $w^* = Q_{W|YZ}^0(\overline{\tau}_m|y^{m+1}, z_m'')$ for some $\overline{\tau}_m \leq \tau_U$

[42]By (O1) in Appendix A1, we can find τ_U and $\tau_V \in (P^{m-1}, P^m]$ such that

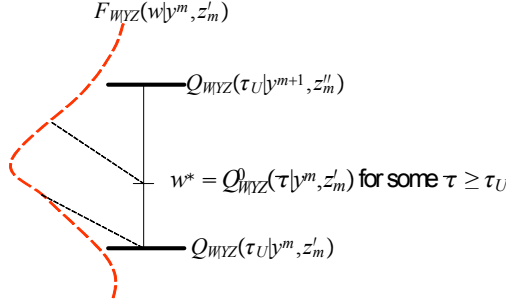$$u = Q_{U|VZ}^c(\overline{\tau}_m|v, z_m') = Q_{U|VZ}^c(\tau_U|\tau_V, z_m').$$

Figure 7: 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.

Part 2 - (b) : Observational equivalence[43] $(F^c_{W|YZ} = F^0_{W|YZ})$

To show (b) we need to show that the data generated by the constructed structure in Part 1, $(S^c = \{h_c, F^c_{U|VZ}\})$, is actually the data we observe, in other words, $F^c_{W|YZ} = F^0_{W|YZ}$: for $p^c_m = P^m - P^{m-1}$, for all $m \in \{1, 2, ..., M\}$,

$$
\begin{aligned}
F^c_{W|YZ}(w|y^m, z) &= \frac{1}{p^c_m} \int_{P^{m-1}}^{P^m} F^c_{U|VZ}(h_c^{-1}(y^m, w)|s, z) ds \\
&= \frac{1}{p^c_m} \int_{P^{m-1}}^{P^m} F^0_{W|YZ}(w|y^m, z) ds \\
&= F^0_{W|YZ}(w|y^m, z)
\end{aligned}
$$

the first equality is due to Lemma 1 in Chesher (2005), the second equality is by construction in (S2), that is, $F^c_{U|VZ}(h_c^{-1}(y^m, w)|v, z) = F^0_{W|YZ}(w|y^m, z)$, for $v \in (P^{m-1}, P^m]$ and the last equality is due to integration over constant and the definition of p^c_m.

Part 2 - (c) : Admissibility by the model

We next show (c). To show sharpness it is required to show that *any* point in the identified set is generated by a structure that satisfies all the restrictions imposed by the model. To show exsitence of such a structure, we constructed a structure in part I, thus, it is required next to show that the structure suggested in part I actually satisfy all the restriction.

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

1. Monotonicity of $h_c(y^m, u)$ **in** u **:** This is shown in Part 1-A.

2. Conditional Quantile Invariance : The distinction of the true structure, S^0, from the constructed structure, S^c, needs to be made in this proof. Recall that $w^* \in I(\tau, y^m, \overline{z}_m)$, and

---

[43] That is, the data distribution that is generated by the structure, constructed in part 1, is actually what we observe.

38

$u^* \equiv Q^c_{U|VZ}(\tau_U | \tau_V, z)$. Note that $h_c^{-1}(y^m, w^*) = u^*$, where $u^* \equiv Q^c_{U|VZ}(\tau_U | \tau_V, z)$ (from (A3-7)). For $u^* = h_c^{-1}(y^m, w^*)$

$$
\begin{aligned}
\tau_U &\equiv F^c_{U|VZ}(u^* | \tau_V, z'_m) \\
&= F^c_{U|VZ}(h_c^{-1}(y^m, w^*) | \tau_V, z'_m) \\
&= F^0_{W|YZ}(w^* | y^m, z'_m) \\
&= \frac{1}{p_m(z'_m)} \int_{P^{m-1}(z'_m)}^{P^m(z'_m)} F^0_{U|VZ}(h_0^{-1}(y^m, w^*) | s, z'_m) ds \\
&= \frac{\Pr(U \le h_0^{-1}(y^m, w^*) \cap P^{m-1}(z'_m) \le V \le P^m(z'_m))}{p_m(z'_m)} \\
&= F^0_{U|VZ}(h_0^{-1}(y^m, w^*) | V \in (P^{m-1}(z'_m), P^m(z'_m)]) \\
&= F^0_{U|Y}(h_0^{-1}(y^m, w^*) | y^m)
\end{aligned}
$$

the second equality is by construction in (S2), the third equality is due to Lemma 1 in Chesher (2005), and the fourth equality follows by integration. The fifth equality is by definition of the conditional probability, the sixth equality is due to how the value of Y is determined. Similarly for $Z = z''_m$,

$$
\begin{aligned}
\tau_U &\equiv F^c_{U|VZ}(u^* | \tau_V, z''_m) \\
&= F^c_{U|VZ}(h_c^{-1}(y^m, w^*) | \tau_V, z'_m) \\
&= F^0_{W|YZ}(w^* | y^m, z''_m) \\
&= \frac{1}{p_m(z''_m)} \int_{P^{m-1}(z''_m)}^{P^m(z''_m)} F^0_{U|VZ}(h_0^{-1}(y^m, w^*) | s, z''_m) ds \\
&= \frac{\Pr(U \le h_0^{-1}(y^m, w^*) \cap P^{m-1}(z''_m) \le V \le P^m(z''_m))}{p_m(z''_m)} \\
&= F^0_{U|VZ}(h_0^{-1}(y^m, w^*) | V \in (P^{m-1}(z''_m), P^m(z''_m)]) \\
&= F^0_{U|Y}(h_0^{-1}(y^m, w^*) | y^m).
\end{aligned}
$$

This yields that $u^* = Q^c_{U|VZ}(\tau_U | \tau_V, z'_m) = Q^c_{U|VZ}(\tau_U | \tau_V, z''_m) = Q^0_{U|Y}(\tau_U | y^m)$, invariant with respect to $z \in \overline{z}_m$.

3. S-LDRM :

Proof. Note that if Restriction S-LDRM is satisfied, Restriction LDRM is guaranteed to be satisfied, thus, it is shown whether the constructed structure, $S^c \equiv \{h_c, F^c_{U|VZ}(u|v, z)\}$, satisfies Restriction S-LDRM, omitting Restriction LDRM.

(1) First, it is noted that $F^c_{U|VZ}(u|v, z)$ is monotonic in v, for $u \in \mathbf{U}$ and $v \in \mathbf{V}$. This is so since $F^c_{U|VZ}(u|v, z)$ is defined as a step function. In the range of $\mathbf{V}$ only two constants

$(F^0_{W|YZ}(w^*|y^m, z)$, and $F^0_{W|YZ}(w^{m+1}|y^{m+1}, z))$ should be considered, and with two constants, monotonicity always holds.

(2) It needs to be shown that S-LDRM holds locally in $\mathbf{U}$ and $\mathbf{V}$. That is, the sign of $h_c(y^{m+1}, u^*) - h_c(y^m, u^*)$, where $u^* \equiv Q^c_{U|VZ}(\tau_U|\tau_V, z)$, needs to be established. We use (O3) for this purpose. Note that by (A3-7) $u^* = h_c^{-1}(y^m, w^*)$, for some $w^* \in I(\tau, y^m, \overline{z}_m)$, and that by (O2) we can express $u^* = h_c^{-1}(y^m, w^*) = h_c^{-1}(y^{m+1}, w^{m+1})$ for some w^{m+1}, where $w^{m+1} = h_c(y^{m+1}, h_c^{-1}(y^m, w^*))$.

Consider $u^* = h_c^{-1}(y^m, w^*) \in \mathbf{U}$ and $v = P^m(z''_m)$[44]. Let τ''_m be

$$
\begin{aligned}
\tau''_m &\equiv F^c_{U|VZ}(u^*|P^m(z''_m), z''_m) \\
&= F^c_{U|VZ}(h_c^{-1}(y^m, w^*)|P^m(z''_m), z''_m) \\
&\overset{(S2)}{=} F^0_{W|YZ}(w^m|y^m, z''_m) \text{ where } w^m = h_c(y^m, h_c^{-1}(y^m, w^*))(= w^*)
\end{aligned}
\tag{A3--6}
$$

as $k = m = l$ in (S2). Next consider $u^* = h_c^{-1}(y^{m+1}, w^{m+1}) \in \mathbf{U}$ and $v = P^{m+1}(z''_m)$. Let τ''_{m+1} be:

$$
\begin{aligned}
\tau''_{m+1} &\equiv F^c_{U|VZ}(u^*|P^{m+1}(z''_m), z''_m) \\
&= F^c_{U|VZ}(h_c^{-1}(y^{m+1}, w^{m+1})|P^{m+1}(z''_m), z''_m) \\
&\overset{(S2)}{=} F^0_{W|YZ}(w^{m+1}|y^{m+1}, z''_m), \text{where } w^{m+1} = h_c(y^{m+1}, h_c^{-1}(y^{m+1}, w^{m+1})),
\end{aligned}
\tag{A3-7}
$$

as $k = l = m+1$ in (S2). Then consider $u^* = h_c^{-1}(y^{m+1}, w^{m+1}) \in \mathbf{U}$ and $P^m(z'_m) < \overline{v} < P^m(z''_m)$, we have

$$
\begin{aligned}
\overline{\tau} &\equiv F^c_{U|VZ}(u^*|v, z''_m) \\
&= F^c_{U|VZ}(h_c^{-1}(y^{m+1}, w^{m+1})|\overline{v}, z''_m) \\
&\overset{(S2)}{=} F^0_{W|YZ}(w^{m+1}|y^m, z''_m), \text{ where } w^{m+1} = h_c(y^{m+1}, h_c^{-1}(y^{m+1}, w^{m+1})),
\end{aligned}
\tag{A3-8}
$$

as $k = m + 1$ and $l = m$ in (S2). Note $P^m(z'_m) \leq P^m(z''_m) \leq P^{m+1}(z''_m)$. Then PD implies that

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

since we are comparing the values of the three conditional distributions evaluated at the same value u^*. ND implies that

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

Note that (*PD) and (*ND) hold regardless of whether W is continuous or discrete.

---

[44]Note that if W is continuous, $u^* = h_c^{-1}(y^m, w^*) = h_c^{-1}(y^{m+1}, w^{m+1})$, where $w^{m+1} = h_c(y^{m+1}, h_c^{-1}(y^m, w^*))$, that is, $u^* = \tilde{u}$. If W is discrete, $h_c^{-1}(y^m, w^*) < h_c^{-1}(y^{m+1}, w^{m+1})$, by definition of h_c^{-1} in (*) in Part 1-A as $h_c^{-1}(y^m, w^*)$ indicates jumping points.

We then express u^*, $w^*(= w^m)$ and w^{m+1} as quantiles of the distributions so that we can find the order of the two, $h_c(y^m, u^*)$ and $h_c(y^{m+1}, u^*)$ by utilizing (*PD) and (*ND). (A3-6)-(A3-8) imply (A3-9) and (A3-10) under continuity of W and U :

$$
\begin{aligned}
u^* &\equiv h_c^{-1}(y^{m+1}, w^{m+1}) = h_c^{-1}(y^m, w^*) \qquad &\text{(A3-9)} \\
&= Q_{U|VZ}^c(\tau_m''|P^m(z_m''), z_m'') \\
&= Q_{U|VZ}^c(\tau_{m+1}''|P^{m+1}(z_m''), z_m'') \\
&= Q_{U|VZ}^c(\bar{\tau}|\bar{v}, z_m''), \quad \text{for } P^m(z_m') < \bar{v} < P^m(z_m'')
\end{aligned}
$$

$$
\begin{aligned}
w^m(= w^*) &\overset{(a)}{=} Q_{W|YZ}^0(\tau_m''|y^m, z_m'') \qquad &\text{(A3-10)} \\
w^{m+1} &\overset{(b)}{=} Q_{W|YZ}^0(\tau_{m+1}''|y^{m+1}, z_m'') \overset{(c)}{=} Q_{W|YZ}^0(\bar{\tau}|y^m, z_m'')
\end{aligned}
$$

(a) follows from (A3-6), (b) from (A3-7) and (c) is by (A3-8).

Finally we can determine the direction of the response : we have [45]

$$
\begin{aligned}
&h_c(y^m, u^*) - h_c(y^{m+1}, u^*), \text{ for } u^* \in \mathbf{U}, \\
=\ &h_c(y^m, Q_{U|VZ}^c(\tau_m''|P^m(z_m''), z_m'')) - h_c(y^{m+1}, Q_{U|VZ}^c(\tau_{m+1}''|P^{m+1}(z_m''), z_m'')) \\
=\ &Q_{W|YZ}^0(\tau_m''|y^m, z_m'') - Q_{W|YZ}^0(\tau_{m+1}''|y^{m+1}, z_m'') \\
=\ &Q_{W|YZ}^0(\tau_m''|y^m, z_m'') - Q_{W|YZ}^0(\bar{\tau}|y^m, z_m'') \\
&\begin{pmatrix} \leq 0 \text{ if PD} \\ \geq 0 \text{ if ND} \end{pmatrix}
\end{aligned}
$$

the first equality is by (A3-9), the second equality is due to by (S1) (or (A3-6)), and the third equality is by (c) in (A3-10). Then the inequality follows because $\tau_m'' \leq \bar{\tau}$ (*PD), $\tau_m'' \geq \bar{\tau}$ (*ND), and the property of quantiles. $\blacksquare$

A.4 Proof of Corollary 2

Proof. We show the case with PDPR. The other case, NDNR, can be shown similarly. Suppose that $Q_{U|VZ}(\tau_U|v, z)$ is weakly increasing in $v \in \mathbf{V}$, then by Restriction LDRM, PDPR is assumed. We adopt Lemma 2 in Chesher (2005) when $m = 1$ with $P^0(z) = 0$ and $P^1(z) = P(z)$, where $P(z) = \Pr(Y = 1|Z = z)$ and when $m = 2$ with $P^2(z) = 1$ and $P^1(z) = P(z)$. We denote $z' \equiv z_1'$ and $z'' \equiv z_1''$ that satisfy Restriction RC. Then we have

$$
\begin{aligned}
h(0, Q_{U|VZ}(\tau_U|0, z')) &\leq Q_{W|YZ}(\tau_U|0, z') \qquad &\text{(A4-1)} \\
&\leq h(0, Q_{U|VZ}(\tau_U|P(z'), z')) \\
h(1, Q_{U|VZ}(\tau_U|P(z''), z'')) &\leq Q_{W|YZ}(\tau_U|1, z'') \qquad &\text{(A4-2)} \\
&\leq h(1, Q_{U|VZ}(\tau_U|1, z''))
\end{aligned}
$$

<hr />

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

Let $u^* \equiv Q_{U|VZ}(\tau_U|\tau_V, z_m'')$, $\overline{u} \equiv Q_{U|VZ}(\tau_U|P(z''), z'')$, and $\underline{u} \equiv Q_{U|VZ}(\tau_U|P(z'), z')$. Then $u^*, \overline{u}, \underline{u} \in \mathbf{U}$, with $\underline{u} \leq u^* \leq \overline{u}$ by Restriction RC. Note that $u^* \equiv Q_{U|VZ}(\tau_U|\tau_V, z') = Q_{U|VZ}(\tau_U|\tau_V, z'')$ by Restriction CQ-I. Then by the same logic as Appendix A2, the upper bound on $h(1, u^*)$ can be found as

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

where the first inequality is due to monotonicity of h in u and the second inequality is due to (A4-2). Applying (A4-1) and Restriction LDRM yields the lower bound as follows

$$Q_{W|YZ}(u|0, z') \leq h(0, \underline{u}) \leq h(1, u^*). \tag{A4-3}$$

where the first inequality is by (A4-1) and the second inequality is by Restriction LDRM.

Consider next the identification of $h(0, u^*)$. Under Restriction RC, using Restriction LDRM and (A4-2) we can find the *upper* bound on $h(0, u^*)$

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

where the first inequality is due to Restriction LDRM and the second inequality is by (A4-2). Similarly, the *lower* bound on $h(0, u^*)$ can be found by

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

where the first inequality is by (A4-1) and the second is by monotonicity of h in u. Thus, we have

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

Note that the identified intervals for $h(0, u^*)$ and $h(1, u^*)$ are the same. ∎

A.5 Lamma A1

Lemma A1 is used in proving Lemma 1 and Lemma 2.

Lemma A1. Let $Q_{yz}^\tau \equiv Q_{W|YZ}(\tau|y, z)$. Then by definition of a quantile we have

$$
\begin{aligned}
\tau &= F_{W|YZ}(Q_{yz}^\tau|Y = y, Z = z) \\
&= \Pr\left(W \leq Q_{yz}^\tau|Y = y, Z = z\right) \\
&= \Pr\left(h(y, U) \leq Q_{yz}^\tau|Y = y, Z = z\right) \\
&= \Pr\left(U \leq h^{-1}(y, Q_{yz}^\tau)|Y = y, Z = z\right).
\end{aligned}
$$

A.6 Proof of Lemma 1

Proof. To show that PDPR iff $Q_{W|YZ}(\tau_U|y^m, z_m') \leq Q_{W|YZ}(\tau_U|y^{m+1}, z_m'')$, let Q_{m+1}'' and Q_m' indicate the values of τ_U- quantiles, $Q_m' \equiv Q_{W|YZ}(\tau_U|y^m, z_m')$ and $Q_{m+1}'' \equiv Q_{W|YZ}(\tau_U|y^{m+1}, z_m'')$. By Lemma A1 we have

$$\tau_U = \Pr\left(U \leq h^{-1}(y^m, Q_m')|Y = y^m, Z = z_m'\right) \tag{A6-1}$$

and

$$\tau_U = \Pr(U \leq h^{-1}(y^{m+1}, Q''_{m+1})|Y = y^{m+1}, Z = z''_m). \tag{A6-2}$$

Suppose PD. Then for[46]

$$P^m(z'_m) - P^{m-1}(z'_m) \geq P^m(z''_m) - P^{m+1}(z''_m), \tag{RC^{L1}}$$

we have

$$
\begin{aligned}
\tau_U &= \Pr(U \leq h^{-1}(y^{m+1}, Q''_{m+1})|Y = y^{m+1}, Z = z''_m) \\
&= \Pr(U \leq h^{-1}(y^{m+1}, Q''_{m+1})|V \in (P^m(z''_m), P^{m+1}(z''_m)]) \\
&= \int_{P^m(z''_m)}^{P^{m+1}(z''_m)} F_{U|VZ}(h^{-1}(y^{m+1}, Q''_{m+1})|s, z''_m) ds \\
&\leq \int_{P^{m-1}(z'_m)}^{P^m(z'_m)} F_{U|VZ}(h^{-1}(y^{m+1}, Q''_{m+1})|s, z'_m) ds, \\
&= \Pr(U \leq h^{-1}(y^{m+1}, Q''_{m+1})|V \in (P^{m-1}(z'_m), P^m(z'_m)]) \\
&= \Pr(U \leq h^{-1}(y^{m+1}, Q''_{m+1})|Y = y^m, Z = z'_m) \equiv \overline{\tau_U}
\end{aligned}
$$

where the first equality is due to Lemma A1, and the second equality follows from the fact that the event $\{V \in (P^m(z''_m), P^{m+1}(z''_m)]\}$ is equivalent to the event $\{Y = y^{m+1}, Z = z''_m\}$. The third equality is by expressing the probability using the conditional distribution, $F_{U|VZ}$, and the last two equalities result from the same logic. The inequality is due to PD, Restriction RC^{L1} and C-QI (this follows by comparing the areas under an weakly decreasing function over two distinct intervals whose size is specified in Restriction RC^{L1}). Then

$$\tau_U \leq \overline{\tau_U} . \tag{A6-3}$$

From (A6-1) - (A6-3), we have

$$
\begin{aligned}
\tau_U &\overset{(A6-1)}{=} \Pr\left(U \leq h^{-1}(y^m, Q'_m)|Y = y^m, Z = z'_m\right) \\
&\overset{(A6-2)}{=} \Pr(U \leq h^{-1}(y^{m+1}, Q''_{m+1})|Y = y^{m+1}, Z = z''_m) \\
&\overset{(A6-3)}{\leq} \Pr\left(U \leq h^{-1}(y^{m+1}, Q''_{m+1})|Y = y^m, Z = z'_m\right) \equiv \overline{\tau_U},
\end{aligned}
$$

the inequality follows since $\tau_U \leq \overline{\tau_U}$. This implies that

$$h^{-1}(y^m, Q'_m) \leq h^{-1}(y^{m+1}, Q''_{m+1})$$

---

[46]For a binary case, this holds if

$$P(z') + P(z'') \geq 1$$

by the nondecreasing property of distribution function, i.e., $a \leq a'$ iff $F_{A|B}(a|b) \leq F_{A|B}(a'|b)$. Let $u^* \equiv h^{-1}(y^m, Q'_m)$ and $u^{**} \equiv h^{-1}(y^{m+1}, Q''_{m+1})$. Then inverting h^{-1} yields

$$
\begin{aligned}
Q'_m &= h(y^m, u^*) \\
Q''_{m+1} &= h(y^{m+1}, u^{**}).
\end{aligned}
$$

By PDPR, we have

$$
h(y^m, u^*) \leq h(y^{m+1}, u^{**}),
$$

which results in $Q'_m \equiv Q_{W|YZ}(\tau_U|y^m, z'_m) \leq Q''_{m+1} \equiv Q_{W|YZ}(\tau_U|y^{m+1}, z''_m)$. That is, in the case of PD if we restrict the interaction such that PD is matched with PR, we need to have

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

The other case, NDNR, can be shown similarly. Thus, we conclude that if we observe

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

PDPR is implied, and if we observe

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

NDNR is implied. ∎

A.7 Proof of Lemma 2 : Testable Implications of the Model

Proof. Let Q^z_m and Q^z_{m+1} indicate the value of τ_U- quantile, $Q^z_m \equiv Q_{W|YZ}(\tau_U|y^m, z)$, and $Q^z_{m+1} \equiv Q_{W|YZ}(\tau_U|y^{m+1}, z)$. Then by Lemma A1

$$
\tau_U = \Pr\left(U \leq h^{-1}(y^m, Q^z_m)|Y = y^m, Z = z\right) \tag{A7-1}
$$

$$
\tau_U = \Pr\left(U \leq h^{-1}(y^{m+1}, Q^z_{m+1})|Y = y^{m+1}, Z = z\right). \tag{A7-2}
$$

Consider PD. Then for

$$
P^m(z) \geq \frac{1}{2}\left[P^{m-1}(z) + P^{m+1}(z)\right], \tag{RC^{L2}}
$$

we have

$$
\begin{aligned}
\tau_U &= \Pr\left(U \leq h^{-1}(y^m, Q^z_m)|Y = y^m, Z = z\right) \tag{A7-3}\\
&= \Pr\left(U \leq h^{-1}(y^m, Q^z_m)|V \in (P^{m-1}(z), P^m(z)]\right)\\
&= \int_{P^{m-1}(z)}^{P^m(z)} F_{U|VZ}(h^{-1}(y^m, Q^z_m)|s, z)ds\\
&\geq \int_{P^m(z)}^{P^{m+1}(z)} F_{U|VZ}(h^{-1}(y^m, Q^z_m)|s, z)ds\\
&= \Pr\left(U \leq h^{-1}(y^m, Q^z_m)|V \in (P^m(z), P^{m+1}(z)]\right)\\
&= \Pr\left(U \leq h^{-1}(y^m, Q^z_m)|Y = y^{m+1}, Z = z\right) \equiv \underline{\tau_U}. \tag{A7-4}
\end{aligned}
$$

44

by the same logic as in Appendix A6. Note that Restriction CQ-I is *not* used in deriving inequality.

Then from (A7-2) and (A7-4), we have

$$
\begin{aligned}
\tau_U &= \; \Pr\left(U \leq h^{-1}(y^{m+1}, Q_{m+1}^z)|Y = y^{m+1}, Z = z\right) \\
&\geq \; \Pr\left(U \leq h^{-1}(y^m, Q_m^z)|Y = y^{m+1}, Z = z\right) \equiv \underline{\tau_U}
\end{aligned}
$$

the inequality follows since $\tau_U \leq \overline{\tau_U}$. This implies that

$$
h^{-1}(y^m, Q_m^z) \leq h^{-1}(y^{m+1}, Q_{m+1}^z)
$$

by the nondecreasing property of a distribution function, i.e. $a \leq a'$ iff $F_{A|B}(a|b) \leq F_{A|B}(a'|b)$. Let $u^* \equiv h^{-1}(y^m, Q_m^z)$ and $u^{**} \equiv h^{-1}(y^{m+1}, Q_{m+1}^z)$. Then inverting h^{-1} yields

$$
\begin{aligned}
Q_m^z &= \; h(y^m, u^*) \\
Q_{m+1}^z &= \; h(y^{m+1}, u^{**}).
\end{aligned}
$$

If PR is matched with PD, that is, if $h(y^m, u^*) \leq h(y^{m+1}, u^{**})$ for $u^*, u^{**} \in \mathbf{U}$ with $u^* \leq u^{**}$, then we have

$$
Q_m^z = h(y^m, u^*) \leq h(y^{m+1}, u^{**}) = Q_{m+1}^z.
$$

The cases for ND can be derived in a similar manner. ∎