



NONPARAMETRIC IDENTIFICATION UNDER DISCRETE VARIATION

Andrew Chesher

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP19/03

Nonparametric identification under discrete variation

ANDREW CHESHER*

CENTRE FOR MICRODATA METHODS AND PRACTICE
INSTITUTE FOR FISCAL STUDIES AND UNIVERSITY COLLEGE LONDON

December 14th 2003

ABSTRACT. This paper provides weak conditions under which there is nonparametric interval identification of local features of a structural function which depends on a discrete endogenous variable and is nonseparable in a latent variate. The function may deliver values of a discrete or continuous outcome and instruments may be discrete valued. Application of the analog principle leads to quantile regression based interval estimators of values and partial differences of structural functions. The results are used to investigate the nonparametric identifying power of the quarter of birth instruments used by Angrist and Krueger (1991) in their study of the returns to schooling.

1. INTRODUCTION

1.1. Discrete variation. This paper explores the limits to identification of structural features when there is discrete variation. Many of the econometric models involving discrete variables used in practice embody parametric restrictions. As persuasively argued by Roehrig (1988), it is important to know the extent to which parametric restrictions make an essential contribution to a model's identifying power. This paper sheds light on this issue by studying the identifying power of nonparametric econometric models when there is discrete variation.

When there is discrete variation parametric restrictions are powerful. That is a matter for concern because parametric restrictions never flow from economic theory. Discrete variation in covariates places limits on the values of endogenous variables at which features of structural functions can be nonparametrically identified and prohibits nonparametric identification of derivatives of structural functions.¹ Parametric restrictions allow "interpolation" between points at which nonparametric identification is feasible.

*I am grateful to Roger Koenker, Charles Manski, and Joao Santos Silva and to seminar participants at cemmap, Northwestern University, University of Chicago and at the Harvard-MIT econometrics workshop for comments on the related paper Chesher (2002) and to Hidehiko Ichimura, Whitney Newey and Elie Tamer for comments on this paper. The paper was presented at the 14th EC² Meeting held at cemmap, London, December 12th - 13th 2003.

¹See Chesher (2002).

Discrete variation in endogenous variables poses particular problems. It is shown here that point nonparametric identification of values of structural functions is infeasible - at best there is interval identification - unless special circumstances prevail. Consequently, there is generally at best interval nonparametric identification of the effects on structural functions of varying the values of discrete endogenous variables.

The paper provides conditions under which there is interval nonparametric identification of:

1. values of structural functions at interior points of support of a discrete endogenous variable, and,
2. partial differences of structural functions for variation across interior points of support of a discrete endogenous variable.

The exclusion of extreme points of support means that the binary endogenous variable case lies outside the scope of application of the results given here. The demonstration of the identifying power of the proposed model is constructive and points to easily computed analog interval estimators.

1.2. Nonseparable structural functions. A stylised Becker-Chiswick-Mincer model² of the determination of investment in schooling and the wage provides one motivation for studying the problem considered in this paper. Let Y_1 , Y_2 and X denote respectively the log wage, years of schooling and a list of covariates and let ρ_1 and ρ_2 denote continuously jointly distributed scalar random variables. The structural equations:

$$Y_1 = h_1(Y_2, X, \rho_1) \tag{1}$$

$$Y_2 = h_2(X, \rho_2) \tag{2}$$

deliver a single value of Y_1 and Y_2 given a value of X , ρ_1 and ρ_2 . In this context there is interest in the sensitivity of the structural function h_1 to variation in Y_2 when other arguments are held fixed, because this bears on the “returns to schooling”. If ρ_1 and ρ_2 are correlated this is not identified without further restrictions.

The structural functions (1) and (2) are required to be monotonically varying (normalised increasing), but not additively separable in the latent variates, so the returns to schooling may exhibit random variation driven by the variation in ρ_1 which may co-vary with ρ_2 . In this situation one might be content to identify some average value of the returns to schooling, but of much greater interest is identification of the value of the returns to schooling at particular values of Y_2 , X and ρ_1 . This paper focusses on the case in which there is discrete variation in schooling as often arises in practice.

²See Becker and Chiswick (1966), Chiswick (1974), Chiswick and Mincer (1972) and Mincer (1974). Card (2001) considers parametric nonseparable models of wage determination and schooling choice. Imbens and Newey (2003) motivate a nonparametric nonseparable specification.

There are many situations in which an endogenous variable varies discretely and a nonseparable model is of interest. For example, in an analysis of demand, Y_1 could be household expenditure on electricity and Y_2 could measure the number of electrical appliances owned by a household.³ In an analysis of firm behaviour, Y_1 could measure company sales or profits and Y_2 could measure the number of plants a firm operates, or, in an analysis of innovation, the number of patents registered.⁴

This paper focusses on the case in which an endogenous variable varies discretely. Results for the continuous case, already available, emerge on considering the effect of reducing to zero the distance between the points of support of the discrete endogenous variable. The analysis allows the outcome, Y_1 and the covariates, X , to vary discretely or continuously.

1.3. Related results. Chesher (2002, 2003) and Imbens and Newey (2003)⁵ study identification in nonseparable models of this sort when endogenous variables are *continuously* distributed. Their results do apply in one case in which endogenous variables vary discretely but this is a case of no econometric interest.⁶

Matzkin (2003) studies identification and estimation of nonseparable structural functions when arguments are exogenous. Roehrig (1988) gives results on global identification in smooth nonseparable models with endogenous variables under continuous variation. Imbens and Newey (2003) relax Roehrig's smoothness restriction and provide estimators and their asymptotic distribution theory for the continuous endogenous variable case. They also study identification and estimation of various average structural functions when there are many sources of stochastic variation.

Chesher (2002, 2003) develops local conditions under which there is identification of values, derivatives and partial differences of nonseparable structural functions at specific values of their arguments for the continuous endogenous variable case. That local focus allows relaxation of the restriction requiring latent variates and covariates to be independently distributed, a restriction essential in a global nonparametric identification analysis such as that of Roehrig (1988), Imbens and Newey (2003) and Matzkin (2003).

There is a large literature on nonparametric identification when latent variates are *additively separable*. See for example Newey, Powell and Vella (1999), Darolles, Florens and Renault (200), Pinkse (2000), Newey and Powell (2003) and Severini and Tripathi (2003) and the references therein.

Nonparametric identification when there are discrete endogenous variables and the structural function is *additively separable* in the latent variate is studied in Das (2003) and Florens and Malavolti (2003).

³See Dubin and McFadden (1984) which employs a parametric specification.

⁴See for example Cockburn and Griliches (1988) and the work surveyed in Griliches (1990).

⁵Imbens and Newey (2003) is a revised version of a 2001 manuscript with the same title.

⁶In Chesher (2002, 2003) and Imbens and Newey (2003) discrete variation in Y_2 is only permitted if variation in X affects the values of the points of support of the distribution of Y_2 but not the probabilities on the points of support.

This paper provides conditions under which there is identification of features of *nonseparable* structural functions when there are discrete endogenous variables. The outcome and the covariates may be discrete or continuous. The attack on the problem follows the approach taken in Chesher (2003). It is helpful to outline that approach now, as a prelude to describing the method of this paper and presenting its main results.

1.4. Identification with continuous endogenous variables. Extensive use is made of conditional quantile functions for which there is the following notation.

$Q_{A|B}(p|b)$ denotes the conditional p -quantile of scalar random variable A given vector random variable $B = b$. With

$$F_{A|B}(a|b) \equiv P[A \leq a|B = b]$$

there is the following definition of $Q_{A|B}(p|b)$.

$$Q_{A|B}(p|b) \equiv \inf \{q : F_{A|B}(q|b) \geq p\}$$

When Y_2 is continuously distributed, conditions sufficient to identify the value of the structural function h_1 and its sensitivity to variation in Y_2 are developed as follows. Substituting for Y_2 in (1) gives

$$Y_1 = h_1(h_2(X, \rho_2), X, \rho_1)$$

and because h_1 is increasing in ρ_1 there is

$$Q_{Y_1|\rho_2, X}(\tau_1|\tau_2, x) = h_1(h_2(x, \tau_2), x, Q_{\rho_1|\rho_2, X}(\tau_1|\tau_2, x))$$

because of the equivariance of quantiles under monotone transformation.

When Y_2 is *continuously* distributed conditioning on $\rho_2 = \tau_2$ and $X = x$ is identical to conditioning on $Y_2 = y_2 \equiv h_2(x, \tau_2)$ and $X = x$, and so there is

$$Q_{Y_1|Y_2, X}(\tau_1|y_2, x) = h_1(y_2, x, Q_{\rho_1|\rho_2, X}(\tau_1|\tau_2, x)). \quad (3)$$

It is convenient to normalise ρ_2 to be uniformly distributed on $(0, 1)$ independent of X so that $y_2 \equiv h_2(x, \tau_2)$ is the conditional τ_2 -quantile of Y_2 given $X = x$, that is $h_2(x, \tau_2) = Q_{Y_2|X}(\tau_2|x)$. With $r_1^* \equiv Q_{\rho_1|\rho_2, X}(\tau_1|\tau_2, x)$ equation (3) can then be written as follows.

$$Q_{Y_1|Y_2, X}(\tau_1|Q_{Y_2|X}(\tau_2|x), x) = h_1(Q_{Y_2|X}(\tau_2|x), x, r_1^*) \quad (4)$$

The value of the conditional quantile on the left hand side of the equation identifies the value of the structural function on the right hand side.

Suppose that, at some point of interest, h_1 and r_1^* are insensitive to some particular variation, Δx in x .⁷ If h_2 is sensitive to this variation there is a nonzero change in

⁷Both x and Δx may be vectors.

$Q_{Y_2|X}(\tau_2|x)$. Then the value of Y_2 at which h_1 is evaluated on the right hand side of (4) is altered as x is moved by Δx , but r_1^* remains constant and there is no effect through the x argument of h_1 . The resulting change in the value of the structural function, entirely attributable to variation in the Y_2 argument, is identified by the effect that moving x by Δx has on the conditional quantile on the left hand side of (4).⁸

If X can be continuously varied then this argument leads to conditions under which there is identification of partial derivatives of structural function, as described in Chesher (2003). But if X can only vary discretely one must be content with identifying partial differences of structural functions as described in Chesher (2002).

1.5. Discrete endogenous variables. The argument employed in the case in which Y_2 is continuously distributed cannot be used when Y_2 is discrete. When Y_2 varies discretely equation (2) no longer defines a one-to-one relationship between (X, ρ_2) and (X, Y_2) and the crucial step from (3) to (4) cannot be taken.

Let y_2^m denote the m th of M points of support of the distribution of discrete Y_2 . Fixing $Y_2 = y_2^m$ and $X = x$ limits ρ_2 to the following interval.⁹

$$\rho_2 \in (F_{Y_2|X}(y_2^{m-1}|x), F_{Y_2|X}(y_2^m|x)]$$

The quantile on the left hand side of equation (4) is constant for variations in τ_2 in this interval, but $r_1^* \equiv Q_{\rho_1|\rho_2,X}(\tau_1|\tau_2, x)$ which appears as an argument of the right hand side of (4) varies as τ_2 is altered or Y_2 is locally exogenous. So, when Y_2 is discrete and at least locally endogenous equation (4) cannot hold.

Under weak nonparametric restrictions the value of a structural function at specific values of its arguments cannot be point identified when one of the arguments is a discrete endogenous variable. Without point identification of the value of a structural function there is no hope of point identification of the effect on the structural function of varying an endogenous, or indeed an exogenous, argument.

1.6. Set identification. The fact that fixing Y_2 and X restricts ρ_2 to an interval hints at the possibility of *set* identification. Define

$$r_1^* \equiv Q_{\rho_1|\rho_2,X}(\tau_1|\tau_2, x)$$

and

$$p^m(x) \equiv F_{Y_2|X}(y_2^m|x).$$

Theorem 1, given in Section 2, states that if, for some point of support y_2^m and probability τ_2 , there exist values of X , x^{m-1} and x^m such that

$$p^m(x^m) \leq \tau_2 \leq p^{m-1}(x^{m-1}) \tag{5}$$

⁸Note that on the left hand side of (4) changing x by Δx affects $Q_{Y_2|X}(\tau_2|x)$, and thus the Y_2 argument of $Q_{Y_1|Y_2,X}$ and also the X argument of $Q_{Y_1|Y_2,X}$.

⁹To cover the case $m = 1$, $F_{Y_2|X}(y_2^0|x)$ is defined to be zero for all x , and note that $F_{Y_2|X}(y_2^M|x) = 1$.

then under certain conditions and for $x \in \{x^{m-1}, x^m\}$, there is the following interval identifying inequality.

$$\begin{aligned} & \min \left(Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^{m-1}), Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^m) \right) \\ & \leq h_1(y_2^m, x, r_1^*) \leq \\ & \max \left(Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^{m-1}), Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^m) \right) \end{aligned} \quad (6)$$

The conditions under which (6) holds include the following.

1. For r_1 in a sufficiently wide interval around r_1^* , $F_{\rho_1|\rho_2X}(r_1|r_2, x)$ is monotonic in r_2 for r_2 satisfying the following inequality.

$$p^{m-1}(x^m) \leq r_2 \leq p^m(x^{m-1})$$

2. The value of the structural function $h_1(y_2^m, x, r_1^*)$ is insensitive to variation in $x \in \{x^{m-1}, x^m\}$.
3. The value r_1^* is insensitive to variation in $x \in \{x^{m-1}, x^m\}$.

The inequality (6) does not apply when y_2^m is an extreme point of support, so the binary endogenous variable case is excluded from consideration. When y_2^m is the lowest (respectively highest) point of support of Y_2 the probability on the right (respectively left) of (5) is zero (respectively one). Then (5) could only be satisfied if τ_2 were zero (respectively one) but then the probability mass at y_2^m would be zero and the object whose identification is sought vanishes.

The argument leading to the interval identification result is now sketched. The seminal work of Hurwicz (1950) makes clear that in any study of identification one must consider how the information contained in the conditional distribution of outcomes (Y_1 and Y_2) given covariates (X) bears on the structural features whose identification is sought. Thus the route to the interval identification result (6) starts with consideration of the conditional distribution of Y_1 given $Y_2 = y_2^m$ and $X = x$ and the way in which this is related to the structural function evaluated at y_2^m and x .

Lemma 2 in the Appendix to the paper provides bounds on this conditional distribution function and its associated quantile function. The bounds on the quantile function are as follows.

$$\begin{aligned} h^L(\tau_1, y_2^m, x) & \equiv \min \left\{ \begin{array}{l} h_1(y_2^m, x, Q_{\rho_1|\rho_2X}(\tau_1|p^{m-1}(x), x)) \\ h_1(y_2^m, x, Q_{\rho_1|\rho_2X}(\tau_1|p^m(x), x)) \end{array} \right\} \\ & \leq Q_{Y_1|Y_2X}(\tau_1|y_2^m, x) \leq \\ \max \left\{ \begin{array}{l} h_1(y_2^m, x, Q_{\rho_1|\rho_2X}(\tau_1|p^{m-1}(x), x)) \\ h_1(y_2^m, x, Q_{\rho_1|\rho_2X}(\tau_1|p^m(x), x)) \end{array} \right\} & \equiv h^U(\tau_1, y_2^m, x) \end{aligned}$$

Suppose $Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^m) \geq Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^{m-1})$. Under the conditions stated above the lower bound on $Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^m)$, is no smaller than the upper bound

on $Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^{m-1})$ and $h_1(y_2^m, x, r_1^*)$ lies between these bounds, that is, for $x \in \{x^{m-1}, x^m\}$:

$$h^L(\tau_1, y_2^m, x^m) \geq h_1(y_2^m, x, r_1^*) \geq h^U(\tau_1, y_2^m, x^{m-1}).$$

The two quantile functions therefore bound $h_1(y_2^m, x, r_1^*)$. When $Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^m) \leq Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^{m-1})$ the inequalities are reversed and combining the results leads to the interval identifying inequality (6).

1.7. Main results and the organisation of the paper. The precise results concerning identification of the value of a structural function are given in Section 2. In many cases identification of the sensitivity of a structural function to changes in an endogenous variable is of primary interest. Identification of structural values is treated first because this is an essential prerequisite for identification of structural differences.

In Section 2 Theorem 1 defines a model and gives a result which is shown to imply that the model interval identifies the value of a structural function of a discrete endogenous variable at specific values of its arguments.

Theorem 2 shows that the interval identifying bounds are sharp in the sense that there can exist admissible structures such that the bounds are attained. For example the bounds are attained when ρ_1 and ρ_2 are distributed independently given X , in which case Y_2 is exogenous.

Theorem 3 implies that if ρ_1 and ρ_2 co-vary more than minimally then point identification of the value of a structural function is not feasible if Y_1 is continuously distributed and Y_2 is discrete. In that situation there are no values, x^{m-1} and x^m , such that

$$Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^{m-1}) - Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^m) = 0 \quad (7)$$

so the interval (6) always has non-zero length.

Analogue estimation of identifying intervals and the issue of determining minimum length intervals are discussed in Section 2.7.

If y_2^m is very close to y_2^{m-1} , which happens when Y_2 is close to being continuously distributed, then it may be possible to find x^{m-1} and x^m such that (7) is “nearly” satisfied and the identifying interval is short. Considering the limiting case, if Y_2 is continuously distributed and there exists x^* such that $\tau_2 = F_{Y_2|X}(y_2^*|x^*)$, and therefore $y_2^* \equiv Q_{Y_2|X}(\tau_2|x^*)$, then there is point identification of $h_1(y_2^*, x, r_1)$, as described in Section 1.4. This is the subject of Section 2.8.

Section 3 gives results on the identification of structural *partial differences*. Identification of a partial difference can be achieved if each component of the difference is interval identified and if the structural function and the value of its ρ_1 argument are insensitive to the variations in X required to identify the components. Section 3.1 and 3.2 deal with interval identification of a structural partial difference with respect to respectively an endogenous variate and a covariate.

Section 4 examines the implications of these results for the well known study of Angrist and Krueger (1991) (AK) which uses wage information and discrete data on

years of schooling to estimate returns to schooling. Analysis of the AK data suggests that the weak quarter of birth instruments have insufficient power to identify the returns to schooling in the absence of parametric restrictions of the sort imposed by AK.¹⁰ Results are presented suggesting that, even if schooling were continuously recorded, the weakness and discreteness of the quarter of birth instruments severely limits the structural information that can be extracted from this data without imposing the sort of parametric restrictions used by AK.

Section 5 concludes with a discussion of the merits of local identification analysis, the extension of the analysis to cases in which there are many endogenous variables, the issue of identification when an endogenous variate is binary, and the relationship of this analysis to the classical Cowles Commission analysis of identification.

2. IDENTIFICATION OF A VALUE DELIVERED BY A STRUCTURAL FUNCTION

2.1. Introduction and definitions. This Section considers nonparametric identification of the value of a discrete or continuous random variable Y_1 delivered by a structural function when one of the arguments of the structural function, Y_2 , is a discrete endogenous variable. Identification of structural partial differences is addressed in Section 3.

First there are the following definitions.

Definition 1. $X \equiv \{X_k\}_{k=1}^K$ is a vector of covariates. The conditional distribution and quantile functions of ρ_1 given ρ_2 and X are denoted by respectively $F_{\rho_1|\rho_2 X}$ and $Q_{\rho_1|\rho_2 X}$. ρ_1 and ρ_2 are jointly continuously distributed latent variates and ρ_2 is normalised uniformly distributed on $(0, 1)$ independent of X .

Definition 2. Y_1 is a discrete, continuous or mixed discrete-continuous random variable. The conditional distribution of Y_2 given $X = x$ is discrete with countable points of support $\{y_2^m\}_{m=1}^M$, invariant with respect to x , and with positive probability masses $\{p_m(x)\}_{m=1}^M$. Cumulative probabilities $\{p^m(x)\}_{m=1}^M$ are defined as:

$$p^m(x) \equiv \sum_{l=0}^m p_l(x), \quad m \in \{1, \dots, M\}$$

with $p^0(x) \equiv 0$ and $p^M(x) = 1$.

Restriction A1 imposes conditions on the structural functions determining Y_1 and Y_2 and will be maintained throughout.

(A1). **Structural functions.** At any value of X , ρ_1 and ρ_2 , the value of the outcomes Y_1 and Y_2 is uniquely determined by the following equations.

$$Y_1 = h_1(Y_2, X, \rho_1) \tag{8}$$

$$Y_2 = h_2(X, \rho_2) \tag{9}$$

¹⁰This is additional to the problem weak instruments pose for *inference* discussed in Bound, Jaeger and Baker (1995).

The function h_1 is weakly monotonic with respect to variation in ρ_1 , normalised caglad and non-decreasing. The function h_2 evaluated at $X = x$, $h_2(x, \rho_2)$ is $Q_{Y_2|X}(\rho_2|x)$, the conditional quantile function of Y_2 given $X = x$, as follows.

$$h_2(x, \rho_2) = \begin{cases} y_2^1 & p^0(x) = 0 < \rho_2 \leq p^1(x) \\ y_2^m & p^{m-1}(x) < \rho_2 \leq p^m(x), \\ y_2^M & p^{M-1}(x) < \rho_2 \leq p^M(x) = 1 \end{cases} \quad m \in \{2, \dots, M-1\}$$

The structural equations (8) and (9) are in triangular form. This is essential in what follows but is less restrictive than it seems at first sight. The second equation can be regarded as a “reduced form” equation in a fully simultaneous structural system, so the analysis which follows can be regarded as an analysis of the identification of features of a *single* structural equation embedded in a *fully simultaneous system*.

2.2. Identifying restrictions. This Section provides nonparametric restrictions under which there is identification of the value, $h_1(y_2^m, x, r_1^*)$, delivered by the structural function h_1 . The value y_2^m , $m \in \{2, \dots, M-1\}$, is an interior point of support of the distribution of Y_2 . The term r_1^* is a value of ρ_1 defined as

$$r_1^* \equiv Q_{\rho_1|\rho_2 X}(\tau_1|\tau_2, x) \quad (10)$$

where $\tau \equiv \{\tau_1, \tau_2\} \in (0, 1)^2$ and x lies in a set of *instrumental values* of X to be specified shortly.¹¹

Four restrictions additional to A1 are now introduced. These define a set, \tilde{x}_m , of instrumental values of X and ensure that r_1^* and $h_1(y_2^m, x, r_1^*)$ are invariant with respect to choice of x within the set \tilde{x}_m .

(B1) *There exist instrumental values of X , $\tilde{x}_m \equiv \{x^{m-1}, x^m\}$, such that*

$$p^m(x^m) \equiv F_{Y_2|X}(y_2^m|x^m) \leq \tau_2 \leq F_{Y_2|X}(y_2^{m-1}|x^{m-1}) \equiv p^{m-1}(x^{m-1}). \quad (11)$$

(B2) *For $x \in \tilde{x}_m$ and all r_1 , and $r_2 \in R_2 \equiv (p^{m-1}(x^m), p^m(x^{m-1})]$, $F_{\rho_1\rho_2 X}(\tau_1|r_2, x)$ is a monotonic function of r_2 .*

(B3) *The value of $r_1^* \equiv Q_{\rho_1|\rho_2 X}(\tau_1|\tau_2, x)$ is invariant with respect to choice of $x \in \tilde{x}_m$.*

(B4) *The value of $h_1(y_2^m, x, r_1^*)$ is invariant with respect to choice of $x \in \tilde{x}_m$, that is:*

$$h_1(y_2^m, x^m, r_1^*) = h_1(y_2^m, x^{m-1}, r_1^*).$$

¹¹In a classical analysis of identification, one identifies a subset of the covariates as *instrumental variables*. In this analysis of local identification it is allowed that variables can play the classical role of instrumental variables at some values but not at others. Hence the term *instrumental values* is used.

2.3. Remarks on the restrictions. Restrictions B1 and B4 taken together are analogous to the *rank condition* in the classical analysis of identification¹² of the values of parameters of a linear structural equation, Restriction B1 requiring the distribution of Y_2 given X to have a degree of sensitivity to variation in X , Restriction B4 limiting the sensitivity of the structural function to variation in X

It follows directly from (11) that

$$Q_{Y_2|X}(\tau_2|x^{m-1}) \leq y_2^{m-1} < y_2^m \leq Q_{Y_2|X}(\tau_2|x^m) \quad (12)$$

the strong inequality, $y_2^{m-1} < y_2^m$, holding because Y_2 has a discrete distribution with positive probability masses on each point of support. So Restriction B1 requires X to have impact on the conditional distribution of Y_2 given X sufficient that both y_2^m and y_2^{m-1} are bounded by τ_2 -quantiles of the distribution of Y_2 given X and that there exist values of X such that this bound can be achieved.

If $\{x^{m-1}, x^m\}$ are such that equalities hold throughout in (11) then y_2^m and y_2^{m-1} are τ_2 -quantiles of Y_2 given X equal to respectively x^m and x^{m-1} . Such pairs of values of X may exist if X has rich support, for example if X is continuously distributed, and the impact of x on $Q_{Y_2|X}(\tau_2|x)$ is sufficiently strong.

Restriction B4 is a weak “local exclusion” restriction requiring insensitivity of the structural function h_1 to variation in X within \tilde{x}_m at the point of interest. It would hold if X were entirely excluded from h_1 . It would hold if certain elements of X were excluded from h_1 , in which case the non-excluded elements would be required to be equal in x^{m-1} and x^m . Restriction B4 is rather weak, requiring only local insensitivity of h_1 to variation in X .

The essential requirement covered by Restriction B2 is that $F_{\rho_1\rho_2X}$, and therefore $Q_{\rho_1\rho_2X}$, have a degree of bounded variation with respect to ρ_2 sufficient to allow the impact of ρ_2 on h_1 (via the dependence of ρ_1 on ρ_2) to be limited when $Y_2 = y_2^m$. The monotonicity restriction of B2 is sufficient for this purpose and can be slightly weakened.

Restriction B3 limits the dependence of ρ_1 on X given ρ_2 . It would hold if ρ_1 were distributed independently of X given ρ_2 but is much weaker being local to τ and to the values of x in \tilde{x}_m .

Identification of $h_1(y_2^m, x, r_1^*)$ when y_2^m is the lowest or highest point of support of Y_2 is *excluded* from consideration. Thus the identification results developed here have no force when Y_2 is binary.

Restriction B1, which plays a key role in this nonparametric analysis, cannot be satisfied in either of these cases. Consider the case $m = M$, so that y_2^m is the maximum value of Y_2 . The cumulative probability $p^M(x^M)$ must be *equal* to one for *all* x^M and there can be no value of $\tau_2 \in (0, 1)$ such that $\tau_2 \geq p^M(x^M) = 1$. If $\tau_2 = 1$ were admitted then Restriction B1 could only be satisfied for x^{M-1} such that $p^{M-1}(x^{M-1}) = 1$ but then there is no probability mass on y_2^M which violates the requirement that all points of support of Y_2 carry positive probability.

¹²Koopmans, Rubin and Leipnik (1950).

Consider the case $m = 1$, so that y_2^m is the minimum value of Y_2 . The cumulative probability $p^0(x) = 0$ for all x so Restriction B1 requires $\tau_2 \leq 0$ which cannot be satisfied if $\tau_2 \in (0, 1)$. If $\tau_2 = 0$ were admitted then Restriction B1 requires x^1 to be chosen such that $p^1(x^1) = 0$ but then there is no probability mass on y_2^1 which violates the requirement that all points of support of Y_2 carry positive probability.

The latent variate ρ_2 is not an argument of the structural function h_1 . Chesher (2002, 2003) and Imbens and Newey (2003), both effectively considering the case in which Y_2 is continuously distributed, demonstrate identification of the value of a structural function when ρ_1 and ρ_2 are arguments of h_1 . This case can be addressed using the methods of this paper if Restriction B2 is modified. The case in which the structural function (8) has the form

$$Y_1 = h_1(Y_2, X, \rho_1, \rho_2)$$

can be addressed if Restriction B2 is replaced by the following restriction.

(B2[§]) For $x \in \tilde{x}_m$ and all t_1 , and $r_2 \in R_2 \equiv (p^{m-1}(x^m), p^m(x^{m-1})]$,

$$h_1(y_2^m, x, Q_{\rho_1 \rho_2 X}(t_1 | r_2, x), r_2)$$

is a monotonic function of r_2 .

The results which follow apply with this modification. Like Restriction B2, B2[§] can be weakened to a local (to τ_1) bounded variation restriction.

2.4. Set identification. Hurwicz (1950) defines the concept of point identification of a feature of a structure. The definition is extended to cover set identification as follows.

Definition 3. A model identifies the value of a feature of a structure S_0 within a set A_0 if in all structures admitted by the model and observationally equivalent to S_0 the value of the feature lies in A_0 .

The model defined by the restrictions given in Section 2.2 set identifies $h_1(y_2^m, x, r_1^*)$. This is the subject of Theorem 1. Define

$$q^L(\tau, y_2^m, \tilde{x}_m) \equiv \min(Q_{Y_1|Y_2X}(\tau_1 | y_2^m, x^{m-1}), Q_{Y_1|Y_2X}(\tau_1 | y_2^m, x^m)) \quad (13)$$

$$q^U(\tau, y_2^m, \tilde{x}_m) \equiv \max(Q_{Y_1|Y_2X}(\tau_1 | y_2^m, x^{m-1}), Q_{Y_1|Y_2X}(\tau_1 | y_2^m, x^m)) \quad (14)$$

and recall the definition (10) of r_1^* .

Theorem 1. Under Restrictions A1 and B1 - B4 there is the inequality:

$$q^L(\tau, y_2^m, \tilde{x}_m) \leq h_1(y_2^m, x, r_1^*) \leq q^U(\tau, y_2^m, \tilde{x}_m) \quad (15)$$

where $x \in \tilde{x}_m \equiv \{x^{m-1}, x^m\}$.

The proof is in the Appendix. The following argument shows that Theorem 1 implies set identification of the value of $h_1(y_2^m, x, r_1^*)$.

The objects $q^L(\tau, y_2^m, \tilde{x}_m)$ and $q^U(\tau, y_2^m, \tilde{x}_m)$ are well-defined, single valued functionals of the distribution of Y_1 and Y_2 given X . Consider any *admissible* structure, S , that is a structure satisfying Restrictions A1 and B1 - B4, in which $h_1(y_2^m, x, r_1^*) = h^S$. Theorem 1 states that

$$q_S^L \leq h_1^S \leq q_S^U$$

where q_S^L and q_S^U are the values obtained when the functionals (13) and (14) are applied to the distribution of Y_1 and Y_2 given X generated by the structure S . Consider any admissible structure S' observationally equivalent to S with $h_1(y_2^m, x, r_1^*) = h^{S'}$. Since S' is observationally equivalent to S it generates the *same* distribution for Y_1 and Y_2 given X as does the structure S , and so it yields the same values of q^L and q^U , that is $q_S^L = q_{S'}^L$ and $q_S^U = q_{S'}^U$ and Theorem 1 tells us that

$$q_S^L \leq h_1^{S'} \leq q_S^U$$

whatever the value of $h_1^{S'}$. Theorem 1 therefore implies that the model defined by the restrictions of Theorem 1 set identifies the value of $h_1(y_2^m, x, r_1^*)$ in the sense of Definition 3.

2.5. Sharpness of the bounds and locally exogenous Y_2 . The bounds of Theorem 1 are sharp in the sense that there can exist admissible structures in which equalities hold throughout in (15). One such case is the subject of Theorem 2. Consider the following restriction in place of Restriction B2.

(B2*) For each $x \in \tilde{x}_m$ and all r_1 , $F_{\rho_1 \rho_2 X}(r_1 | r_2, x)$ is constant for variation in r_2 in $R_2 \equiv (p^{m-1}(x^m), p^m(x^{m-1}))$.

Theorem 2. Under Restrictions A1, B1, B2*, B3 and B4 there is the equality

$$h_1(y_2^m, x, r_1^*) = Q_{Y_1|Y_2X}(\tau_1 | y_2^m, x^m) = Q_{Y_1|Y_2X}(\tau_1 | y_2^m, x^{m-1})$$

where $x \in \tilde{x}_m$.

Theorem 2 not only demonstrates the sharpness of the bounds of Theorem 1, but also shows that when Y_2 is “locally exogenous” point identification of $h_1(y_2^m, x, r_1^*)$ is possible if suitable instrumental values can be found.¹³

¹³This result was given in Chesher (2002).

2.6. Point and interval identification. Theorem 1 implies that the value of $h_1(y_2^m, x, r_1^*)$ is point identified if:

$$Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^{m-1}) = Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^m). \quad (16)$$

When Y_1 is a discrete random variable this equality can hold because the discretely varying quantiles of Y_1 given Y_2 and X can be insensitive to variation in $x \in \tilde{x}_m$.

When Y_1 is continuously distributed the equality (16) cannot hold unless the conditions of Theorem 2 prevail. This is the subject of Theorem 3. Consider the following restriction in place of Restriction B2.

(B2**) For each $x \in \tilde{x}_m$ and all r_1 , $F_{\rho_1\rho_2|X}(r_1|r_2, x)$ is monotonic and non-constant for variation in r_2 in $R_2 \equiv (p^{m-1}(x^m), p^m(x^{m-1}))$. h_1 is strictly monotonic in ρ_1 .

Theorem 3. Under Restrictions A1, B1, B2**, B3 and B4

$$Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^{m-1}) \neq Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^m).$$

and $h_1(y_2^m, x, r_1^*)$ is not point identified.

An implication of Theorem 3 is that when Y_1 is continuous and Y_2 is discrete, the model defined by the restrictions of the Theorem 1 only *interval* identifies the value of $h_1(y_2^m, x, r_1^*)$.

There may be many values $\{x^{m-1}, x^m\}$ for which the conditions of Theorem 1 are satisfied and so many identifying intervals. Clearly the interval with minimum length is obtained using $\{x_{opt}^{m-1}, x_{opt}^m\}$ defined as:

$$\{x_{opt}^{m-1}, x_{opt}^m\} = \arg \min_{\{x^{m-1}, x^m\} \in B} |Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^{m-1}) - Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^m)|$$

where B is the set of values of X for which Restrictions B1 - B4 hold.

There is interval *overidentification* of $h_1(y_2^m, x, r_1^*)$ if there are many pairs \tilde{x}_m satisfying the conditions of Theorem 1 leading to the same interval (15). Such a situation can arise, for example, when X contains many elements with more than one of them “excluded” from the structural function h_1 .

The support of X may limit the values of Y_2 and τ_1 at which interval identification can be obtained. Even if there is rich support, when instruments (X) are weak, having only a small influence on the conditional distribution of Y_2 given X , it may not be possible to find values of X satisfying (11) for any value of τ_2 . The combination of weak instruments and sparse support can combine to produce underidentification at *all* values of Y_2 , τ_1 and τ_2 , as is illustrated in Section 4 using data employed in the study of the returns to schooling reported in Angrist and Krueger (1991).

2.7. Estimation. Identifying intervals can be estimated by applying the analog principle, employing non-, semi- or fully- parametric estimates of the conditional distribution or quantile functions of Y_2 given X and of Y_1 given Y_2 and X .¹⁴

Given a value, y_2^m of interest, a value τ_2 and an estimate $\hat{F}_{Y_2|X}$ of the conditional distribution function of Y_2 given X one seeks values \hat{x}^{m-1} and \hat{x}^m such that

$$\hat{F}_{Y_2|X}(y_2^m|\hat{x}^m) \leq \tau_2 \leq \hat{F}_{Y_2|X}(y_2^{m-1}|\hat{x}^{m-1})$$

and calculates the values of an estimate of the conditional τ_1 -quantile function of Y_1 given $Y_2 = y_2^m$ and $X = x$ with $x \in \{\hat{x}^{m-1}, \hat{x}^m\}$. This procedure delivers an estimated interval

$$\left[\hat{q}^L(\tau, y_2^m, \hat{x}_m), \hat{q}^U(\tau, y_2^m, \hat{x}_m) \right].$$

A minimum length interval can be calculated using $\{\hat{x}_{opt}^{m-1}, \hat{x}_{opt}^m\}$ obtained as

$$\{\hat{x}_{opt}^{m-1}, \hat{x}_{opt}^m\} = \arg \min_{\{x^{m-1}, x^m\} \in B \cap \hat{C}} \left| \hat{Q}_{Y_1|Y_2X}(\tau_1|y_2^m, x^{m-1}) - \hat{Q}_{Y_1|Y_2X}(\tau_1|y_2^m, x^m) \right|$$

where, as above B defines the set of values of X satisfying Restrictions B1 - B4 and

$$\hat{C} \equiv \{(x^{m-1}, x^m) : \hat{F}_{Y_2|X}(y_2^m|x^m) \leq \tau_2 \leq \hat{F}_{Y_2|X}(y_2^{m-1}|x^{m-1})\}.$$

This throws up challenging inferential issues, not addressed here.¹⁵

2.8. Continuous endogenous variables. If Y_2 is continuously distributed then in (11) we can set $y_2^m = y_2^{m-1} = y_2^*$ say. The monotonicity restriction B2 is then irrelevant. Suppose there exist $\{x', x''\}$ such that

$$F_{Y_2|X}(y_2^*|x') \leq \tau_2 \leq F_{Y_2|X}(y_2^*|x'') \quad (17)$$

and that $r_1^* \equiv Q_{\rho_1|\rho_2X}(\tau_1|\tau_2, x)$ and $h_1(y_2^*, x, r_1^*)$ are invariant with respect to choice of $x \in \{x', x''\}$. Then the argument employed in the proof of Theorem 1 leads to the identifying inequality

$$\begin{aligned} \min (Q_{Y_1|Y_2X}(\tau_1|y_2^*, x'), Q_{Y_1|Y_2X}(\tau_1|y_2^*, x'')) \\ \leq h_1(y_2^*, x, r_1^*) \leq \end{aligned} \quad (18)$$

$$\max (Q_{Y_1|Y_2X}(\tau_1|y_2^*, x'), Q_{Y_1|Y_2X}(\tau_1|y_2^*, x''))$$

¹⁴For the case in which Y_1 is continuous see Koenker and Bassett (1978), Koenker and d'Orey (1987) and Ma and Koenker (2003) for parametric estimation, see Chaudhuri, Doksum and Samarov (1997), Kahn (2001) and Lee (2003a, 2003b) for semi-parametric estimation and see Chaudhuri (1991) for nonparametric estimation. There are fewer options for the case in which Y_1 is discrete. Machado and Santos Silva (2002) propose a procedure for the parametric case. One could proceed by estimating the conditional distribution function of Y_1 given Y_2 and X , and obtain an estimated quantile function by inversion.

¹⁵See Chernozhukov, Hong and Tamer (2003) and Imbens and Manski (2003) for results on inference when there is interval identification of structural features.

where $x \in \{x', x''\}$.

When X has rich support there may exist x^* such that $F_{Y_2|X}(y_2^*|x^*) = \tau_2$ and setting $x' = x'' = x^*$ in (18) leads to the point identifying equality

$$h_1(y_2^*, x^*, r_1^*) = Q_{Y_1|Y_2X}(\tau_1|y_2^*, x^*)$$

where $y_2^* \equiv Q_{Y_2X}(\tau_2|x^*)$.¹⁶

3. IDENTIFICATION OF STRUCTURAL PARTIAL DIFFERENCES

3.1. Identification of a partial difference with respect to an endogenous variable. This Section provides conditions under which there is identification of the value of the structural partial difference $\Delta_{ij}h_1^*$ defined as follows.

$$\Delta_{ij}h_1^* \equiv h_1(y_2^i, x, r_1^*) - h_1(y_2^j, x, r_1^*)$$

Here y_2^i and y_2^j are two interior points of support of the distribution of Y_2 with, r_1^* is a value of ρ_1 defined as

$$r_1^* \equiv Q_{\rho_1|\rho_2X}(\tau_1|\tau_2, x)$$

where $\tau \equiv \{\tau_1, \tau_2\} \in (0, 1)^2$ and r_1^* is invariant with respect to choice of $x \in \tilde{x}_{ij}$, a set of *instrumental values* of X .

The identification strategy involves putting in place conditions sufficient to identify the values of the two structural functions that appear in the partial difference. Then additional restrictions (C1 and C2 below) are introduced which make these values invariant with respect to the variations in x required to identify the two values of the structural functions that appear in $\Delta_{ij}h_1^*$.

For $k \in \{i, j\}$, let \tilde{x}_k be a set of instrumental values for which conditions B1 - B4 hold, conditions sufficient to interval identify $h_1(y_2^k, x, r_1^*)$ with $x \in \tilde{x}_k$. Define $\tilde{x}_{ij} \equiv \tilde{x}_i \cup \tilde{x}_j$ and consider conditions C1 - C2.

(C1) *The value of r_1^* is invariant with respect to choice of $x \in \tilde{x}_{ij}$.*

(C2) *For $k \in \{i, j\}$ and $\{x', x''\} \in \tilde{x}_{ij}$:*

$$h_1(y_2^k, x', r_1^*) = h_1(y_2^k, x'', r_1^*).$$

Under these conditions there are the interval identifying inequalities:

$$\begin{aligned} q^L(\tau, y_2^i, \tilde{x}_i) &\leq h_1(y_2^i, x, r_1^*) \leq q^U(\tau, y_2^i, \tilde{x}_i) \\ -q^U(\tau, y_2^j, \tilde{x}_j) &\leq -h_1(y_2^j, x, r_1^*) \leq -q^L(\tau, y_2^j, \tilde{x}_j) \end{aligned}$$

where $x \in \tilde{x}_{ij}$. Addition yields an inequality which interval identifies the structural partial difference $\Delta_{ij}h_1^*$, a result formally expressed in Theorem 4.

¹⁶This local point identification result is given in Chesher (2003).

Theorem 4. *Suppose that (i) Restriction A1 holds, (ii) with $\{i, j\} \subseteq \{2, \dots, M-1\}$, for $k \in \{i, j\}$, $\tilde{x}_k \equiv \{x^{k-1}, x^k\}$ is such that the conditions of Theorem 1 hold when identification of the value of $h_1(y_2^k, x, r_1^*)$ is considered and $x \in \tilde{x}_k$, (iii) restrictions C1 and C2 hold. Then :*

$$q^L(\tau, y_2^i, \tilde{x}_i) - q^U(\tau, y_2^j, \tilde{x}_j) \leq \Delta_{ij} h_1^* \leq q^U(\tau, y_2^i, \tilde{x}_i) - q^L(\tau, y_2^j, \tilde{x}_j)$$

where

$$\Delta_{ij} h_1^* \equiv h_1(y_2^i, x, r_1^*) - h_1(y_2^j, x, r_1^*)$$

and $x \in \tilde{x}_{ij} \equiv \tilde{x}_i \cup \tilde{x}_j$.

The proof is omitted, being straightforward following the argument outlined above.

3.2. Identification of a partial difference with respect to covariates. Partition x thus $x = [\hat{x}, \check{x}]$ and write the structural function thus: $h_1(y_2, \hat{x}, \check{x}, \rho_1)$ making explicit the dependence on the components of x . This Section provides conditions under which there is identification of the value of the structural partial difference $\Delta_{ab} h_1^*$ defined as follows.

$$\Delta_{ab} h_1^* \equiv h_1(y_2^i, \hat{x}_a, \check{x}, r_1^*) - h_1(y_2^i, \hat{x}_b, \check{x}, r_1^*)$$

Here y_2^i is a point of support of the distribution of Y_2 with $i \in \{2, \dots, M-1\}$, r_1^* is a value of ρ_1 defined as

$$r_1^* \equiv Q_{\rho_1|\rho_2 X}(\tau_1|\tau_2, x)$$

where $\tau \equiv \{\tau_1, \tau_2\} \in (0, 1)^2$ and r_1^* is invariant with respect to choice of $x \in \tilde{x}_{ab}$, a set of *instrumental values* of X .

For $c \in \{a, b\}$ let \tilde{x}_c be a set of instrumental values for which conditions B1 - B4 hold, conditions sufficient to interval identify $h_1(y_2^i, \hat{x}_c, \check{x}, r_1^*)$ with $x \in \tilde{x}_c \equiv \{x_c^{i-1}, x_c^i\}$ where

$$\begin{aligned} x_c^{i-1} &\equiv [\hat{x}_c, \check{x}_c^{i-1}] \\ x_c^i &\equiv [\hat{x}_c, \check{x}_c^i] \end{aligned}$$

so that the component \hat{x} takes a *common* value, \hat{x}_c , in x_c^{i-1} and x_c^i . Define $\tilde{x}_{ab} \equiv \tilde{x}_a \cup \tilde{x}_b$ and consider conditions D1 and D2.

(D1) *The value of r_1^* is invariant with respect to choice of $x \in \tilde{x}_{ab}$.*

(D2) *For $c \in \{a, b\}$ and $\{\check{x}', \check{x}''\} \in \{\check{x}_a^{i-1}, \check{x}_a^i, \check{x}_b^{i-1}, \check{x}_b^i\}$:*

$$h_1(y_2^i, \hat{x}_c, \check{x}', r_1^*) = h_1(y_2^i, \hat{x}_c, \check{x}'', r_1^*).$$

Condition D2 is clearly satisfied if \tilde{x} is excluded from h_1 . Under these conditions there are the interval identifying inequalities:

$$\begin{aligned} q^L(\tau, y_2^i, \tilde{x}_a) &\leq h_1(y_2^i, \hat{x}_a, \tilde{x}, r_1^*) \leq q^U(\tau, y_2^i, \tilde{x}_a) \\ -q^U(\tau, y_2^i, \tilde{x}_b) &\leq -h_1(y_2^i, \hat{x}_b, \tilde{x}, r_1^*) \leq -q^L(\tau, y_2^i, \tilde{x}_b) \end{aligned}$$

where $x \in \tilde{x}_{ab}$. Addition yields an inequality which interval identifies the structural partial difference $\Delta_{ab}h_1^*$, a result formally expressed in the following Theorem.

Theorem 5. *Suppose that (i) Restriction A1 holds, (ii) with $i \in \{2, \dots, M-1\}$, for $c \in \{a, b\}$, $\tilde{x}_c \equiv \{x_c^{i-1}, x_c^i\}$ with*

$$\begin{aligned} x_c^{i-1} &\equiv [\hat{x}_c, \tilde{x}_c^{i-1}] \\ x_c^i &\equiv [\hat{x}_c, \tilde{x}_c^i] \end{aligned}$$

is such that the conditions of Theorem 1 hold when identification of the value of $h_1(y_2^i, \hat{x}_c, \tilde{x}_c, r_1^)$ is considered, (iii) restrictions D1 and D2 hold. Then :*

$$q^L(\tau, y_2^i, \tilde{x}_a) - q^U(\tau, y_2^i, \tilde{x}_b) \leq \Delta_{ab}h_1^* \leq q^U(\tau, y_2^i, \tilde{x}_a) - q^L(\tau, y_2^i, \tilde{x}_b)$$

where

$$\Delta_{ab}h_1^* \equiv h_1(y_2^i, \hat{x}_a, \tilde{x}, r_1^*) - h_1(y_2^i, \hat{x}_b, \tilde{x}, r_1^*)$$

and $x \in \tilde{x}_{ab} \equiv \tilde{x}_a \cup \tilde{x}_b$ and $\tilde{x} \in \{\tilde{x}_a^{i-1}, \tilde{x}_a^i, \tilde{x}_b^{i-1}, \tilde{x}_b^i\}$.

The proof is omitted, being straightforward following the argument outlined above. If an element of X exhibits continuous variation and h_1 is differentiable with respect to that element then a simple extension of Theorem 5 leads to a result on interval identification of a partial derivative of h_1 with respect to a covariate.

3.3. Minimum length identifying intervals. Consider the length of an interval that identifies a partial difference, for example with respect to a discrete endogenous variable.

$$L_{ij} \equiv (q^U(\tau, y_2^i, \tilde{x}_i) - q^L(\tau, y_2^j, \tilde{x}_j)) - (q^L(\tau, y_2^i, \tilde{x}_i) - q^U(\tau, y_2^j, \tilde{x}_j))$$

This is necessarily nonnegative and can be written as the sum of the lengths of the identifying intervals for the two components of the difference, also nonnegative, as follows.

$$L_{ij} = (q^U(\tau, y_2^i, \tilde{x}_i) - q^L(\tau, y_2^i, \tilde{x}_i)) + (q^U(\tau, y_2^j, \tilde{x}_j) - q^L(\tau, y_2^j, \tilde{x}_j))$$

Therefore choices of \tilde{x}_i and \tilde{x}_j which minimise the lengths of the identifying intervals for respectively $h_1(y_2^i, x, r_1^*)$ and $h_1(y_2^j, x, r_1^*)$ yield minimum length identifying intervals for the partial difference $\Delta_{ij}h_1^*$. So an estimate of a minimum length identifying interval for a structural partial difference can be obtained by combining estimates of minimum length intervals for each component of the difference.

Point identification of a structural partial difference cannot be achieved when Y_1 is continuous and Y_2 is discrete under the weak nonparametric restrictions considered here because, under those conditions, point identification of values delivered by structural functions is infeasible.

4. WEAK INSTRUMENTS

Crucial to the identification of the value of a structural function at a point of support, y_2^m , of a discrete endogenous variable is the ability to find instrumental values, x^{m-1} and x^m such that for a chosen value of $\tau_2 \in (0, 1)$ the following inequality of Restriction B1 of Theorem 1 is satisfied.

$$P[Y_2 \leq y_2^m | X = x^m] \equiv p^m(x^m) \leq \tau_2 \leq p^{m-1}(x^{m-1}) \equiv P[Y_2 \leq y_2^{m-1} | X = x^{m-1}]$$

When X has only a weak effect on the conditional distribution of Y_2 given X or when the support of X is limited this may not be possible. The model and data studied by Angrist and Krueger (1991) (AK) illustrate the difficulty.

AK consider a parametric model in which the log wage, Y_1 , is a linear function of endogenous years of schooling, Y_2 . The returns to schooling is the focus of interest, that is the coefficient on Y_2 in the linear structural log wage equation.

AK propose binary indicators of quarter of birth (QOB) as instrumental variables. They argue that QOB is an influence on years of schooling because people born in different quarters of the year enter school at different ages, yet can leave school at any point in the year. The influence of QOB on attained years of schooling is evident in the data, but the effect is rather weak. AK argue that QOB does not influence the wage given years of schooling and assume that the latent variable in the structural wage equation is mean independent of QOB. Their parametric model, which incorporates these restrictions, identifies the returns to schooling.

In this Section the identifying power of the QOB instruments in the absence of AK's parametric restrictions is considered.

4.1. Discrete years of schooling. The AK data record years of schooling as integers, so Y_2 is discrete. The instrumental variables employed by AK are binary variables, $X \equiv [X_1, \dots, X_4]$ indicating quarter of birth with $X_i = 1$ for a person born in quarter i and equal to zero otherwise. These are weak instruments so far as Restriction B1 is concerned because variation in their values can only exert a small influence on the distribution of years of schooling given X . For someone who leaves school at the earliest legal opportunity, say at age 15, the difference in schooling comparing people born in quarter 1 and quarter 4 is only a few months.

Table 1 shows nonparametric estimates of the distribution function of years of schooling for each quarter of birth.¹⁷ The data come from 329,509 members of the 1930-39 cohort observed in 1980, around 80,000 born in each of the four quarters.¹⁸ The QOB 1 and QOB 4 estimated distribution functions are plotted in Figure 1.

¹⁷These are the empirical distribution functions. The estimate of $P[Y_2 \leq y_2^m | X_i = 1]$ is simply the proportion of those with $X_i = 1$ having recorded years of schooling no greater than y_2^m .

¹⁸Sample sizes used in producing each entry in a row of Table 1 exceed 1,000 for $Y_2 > 4$ and exceed 7,000 for $Y_2 > 7$ after which the estimates are very accurate. For example the estimated standard errors at $Y_2 = 8$ are all close to 0.0033.

Table 1: Estimated distribution function of years of schooling for each quarter of birth

Years of schooling	QOB 1	QOB 2	QOB 3	QOB 4
1	.002	.003	.002	.002
2	.005	.005	.004	.004
3	.008	.009	.007	.007
4	.013	.013	.011	.010
5	.020	.019	.016	.016
6	.032	.032	.028	.027
7	.052	.050	.046	.045
8	.104	.101	.094	.090
9	.146	.142	.133	.128
10	.197	.194	.182	.177
11	.238	.236	.223	.218
12	.612	.604	.598	.595
13	.666	.658	.653	.652
14	.742	.734	.734	.732
15	.771	.763	.764	.763
16	.877	.873	.874	.872
17	.913	.909	.911	.909
18	.947	.945	.947	.946
19	.966	.964	.966	.966
20	1.00	1.00	1.00	1.00

Inspecting Table 1 it can be seen that for every value of m

$$\max_i(\hat{P}[Y_2 \leq y_2^{m-1} | X_i = 1]) < \min_i(\hat{P}[Y_2 \leq y_2^{m-1} | X_i = 1])$$

and it follows that there is *no* value of τ_2 and m for which there are configurations of the four binary indicators, x^m and x^{m-1} such that:

$$\hat{p}^m(x^m) \leq \tau_2 \leq \hat{p}^{m-1}(x^{m-1}).$$

This strongly suggests that the AK quarter of birth instruments are too weak to nonparametrically identify the returns to schooling.

4.2. “Continuous” years of schooling. It is interesting to consider what the identifying power of the quarter of birth instruments would have been had years of schooling been recorded *continuously*. This also provides the opportunity to illustrate some of the results set out earlier in the paper.

For this exercise the estimated distribution functions reported in Table 1 are treated as if they were the *actual* distribution functions. The integer years of schooling recorded in the AK data are treated as if they were the result of rounding exact figures to the nearest integer in a situation in which a value recorded as y_2 is equally likely to be any value in the interval $y_2 \pm 0.5$. Figure 2 shows the resulting hypothetical distribution functions of continuous years of schooling for QOB 1 (solid) and QOB 4 (dashed).¹⁹

Write the structural log wage function as

$$Y_1 = h_1(Y_2, X, \rho_1)$$

where Y_1 denotes the log wage, $X \equiv \{X_1, X_2, X_3, X_4\}$ lists the QOB indicators and Y_2 denotes schooling. Let $x_{[i]}$ denote the configuration of QOB indicators in which $X_i = 1$.

Section 2.8 considered nonparametric identification in the presence of a continuous endogenous variable. Suppose that, for some τ_2 and y_2^* , there are QOB configurations $x_{[i]}$ and $x_{[j]}$, $\{i, j\} \subset \{1, 2, 3, 4\}$, not necessarily distinct, such that the inequality

$$F_{Y_2|X}(y_2^*|x_{[i]}) \leq \tau_2 \leq F_{Y_2|X}(y_2^*|x_{[j]}) \quad (19)$$

is satisfied. *The discussion of the remainder of this Section is predicated on the validity of the other required identifying restrictions.*²⁰

If (19) holds there is interval identification of the value of the log wage structural function evaluated at $Y_2 = y_2^*$, $X = x \in \{x_{[i]}, x_{[j]}\}$ and

$$\rho_1 = r_1^*(\tau_2) \equiv Q_{\rho_1|\rho_2 X}(\tau_1|\tau_2, x).$$

It is convenient to make the dependence of r_1^* on τ_2 explicit here. The additional identifying restrictions ensure that h_1 and $r_1^*(\tau_2)$ are invariant to choice of $x \in \{x_{[i]}, x_{[j]}\}$.

Figure 3 zooms in on Figure 2, focussing on years of schooling between 5 and 12 years. The vertical line in Figure 3 marks 8 years of schooling. It intersects the distribution functions at $a = .109$ and at $b = .125$. Considering the inequality (19) there is therefore interval identification of $h_1(8, x, r_1^*(\tau_2))$ for $x \in \{x_{[1]}, x_{[4]}\}$ at all values $\tau_2 \in [.109, .125]$, and for τ_2 in this interval:

$$\begin{aligned} & \min(Q_{Y_1|Y_2 X}(\tau_1|8, x_{[1]}), Q_{Y_1|Y_2 X}(\tau_1|8, x_{[4]})) \\ & \leq h_1(8, x, r_1^*(\tau_2)) \leq \\ & \max(Q_{Y_1|Y_2 X}(\tau_1|8, x_{[1]}), Q_{Y_1|Y_2 X}(\tau_1|8, x_{[4]})) \end{aligned}$$

¹⁹These conjectured distribution functions are obtained by joining midpoints of each successive step of each of the estimated distribution functions graphed in Figure 1. The midpoints are joined with straight lines because the result is intended to reflect the distribution of “continuous” years of schooling in the situation in which schooling is uniformly distributed within each one year interval. Alternative “assumptions” about the distribution of schooling within one year intervals are captured using alternative increasing functions to join the midpoints.

²⁰Thus it is supposed that h_1 and $Q_{\rho_1|\rho_2 X}(\tau_1|\tau_2, x)$ are insensitive to variations in $x \in \{x_{[i]}, x_{[j]}\}$ and the required monotonicity conditions hold.

where $x \in \{x_{[1]}, x_{[4]}\}$.

At $\tau_2 = .109$ and $\tau_2 = .125$ the structural log wage function $h_1(8, x, r_1^*(\tau_2))$ is *point* identified because in (19)

$$\begin{aligned} F_{Y_2|X}(8|x_{[1]}) &\leq .109 \leq F_{Y_2|X}(8|x_{[1]}) \\ F_{Y_2|X}(8|x_{[4]}) &\leq .125 \leq F_{Y_2|X}(8|x_{[4]}) \end{aligned}$$

which are of course equalities, and so:

$$h_1(8, x, r_1^*(.109)) = Q_{Y_1|Y_2X}(\tau_1|8, x_{[1]}) \quad (20)$$

$$h_1(8, x, r_1^*(.125)) = Q_{Y_1|Y_2X}(\tau_1|8, x_{[4]}) \quad (21)$$

where $x \in \{x_{[1]}, x_{[4]}\}$.

Now consider the horizontal line drawn in Figure 3 at $\tau_2 = 0.2$. This intersects the distribution functions at $c = 9.58$ and $d = 10.06$. Considering the inequality (19) there is

$$\begin{aligned} F_{Y_2|X}(9.58|x_{[1]}) &\leq 0.2 \leq F_{Y_2|X}(9.58|x_{[1]}) \\ F_{Y_2|X}(10.06|x_{[4]}) &\leq 0.2 \leq F_{Y_2|X}(10.06|x_{[4]}) \end{aligned}$$

and therefore at $y_2 = 9.58$ and $y_2 = 10.06$ the value of the structural function $h_1(y_2, x, r_1^*(0.2))$ is point identified, thus:

$$h_1(9.58, x, r_1^*(0.2)) = Q_{Y_1|Y_2X}(\tau_1|9.58, x_{[1]}) \quad (22)$$

$$h_1(10.06, x, r_1^*(0.2)) = Q_{Y_1|Y_2X}(\tau_1|10.06, x_{[4]}) \quad (23)$$

where $x \in \{x_{[1]}, x_{[4]}\}$.

For any choice of τ_1 the structural functions in (22) and (23) are evaluated at the *same* value of ρ_1 , that is

$$r_1^*(0.2) \equiv Q_{\rho_1|\rho_2X}(\tau_1|0.2, x).$$

Therefore the difference between the quantile functions in (22) and (23) point identifies a *partial difference* of the structural function with respect to years of schooling. The returns to schooling to increasing schooling by 0.48 years, from 9.58 years to 10.06 years, for a person at the 0.2-quantile of the distribution of ρ_2 and the τ_1 -quantile of the distribution of ρ_1 conditional on $\tau_2 = 0.2$ and $x \in \{x_{[1]}, x_{[4]}\}$ is identified by

$$\frac{1}{0.48} (Q_{Y_1|Y_2X}(\tau_1|10.06, x_{[4]}) - Q_{Y_1|Y_2X}(\tau_1|9.58, x_{[1]})).$$

Returning to Figure 3, moving the horizontal line away from $\tau_2 = 0.2$ produces a sequence of pairs of values of years of schooling generated by the intersection of the horizontal line with the two estimated distribution functions. These are the pairs of values at which the returns to schooling can be identified.

The binary and weak nature of the QOB instruments have significant implications in the absence of parametric restrictions. Because the two QOB instruments are binary each pair of values is associated with a different value of τ_2 . One could never learn about the returns to schooling at low levels of schooling for a person at a high quantile of the distribution of ρ_2 . Because the instruments are weak the pairs of values exhibit very little separation. One could not learn about the returns to a large change in schooling.

5. CONCLUDING REMARKS

5.1. Local identification. The objects whose identification is studied in this paper are *local* features of structures: values and partial differences of structural functions at specific values of their arguments. The conditions under which they are identified are specific to the objects considered and *local* in character: for example they require structural functions to be locally insensitive to particular variations in covariates; they require insensitivity of a specific quantile of a latent variate to particular variations in covariates.

Local conditions of this sort allow one to understand when a structural interpretation can be placed on particular features of estimated conditional quantile and distribution functions. Nonparametric identification of entire nonseparable structural functions seems to require the unpalatable restriction that latent variates and covariates are *independently* distributed as in the analysis of Roehrig (1988), Imbens and Newey (2003) and Matzkin (2003). One might be prepared to believe in a degree of local independence even in situations in which global independence seems untenable.

5.2. Many endogenous variables. The results given here are for the case in which there is a single endogenous variable. Extension to the case in which one endogenous variable (Y_2) is discrete and there is another *continuous* endogenous variable (Y_3) is straightforward. Among the conditions under which there is interval identification of the value of a structural function at $Y_2 = y_2^m$ and $Y_3 = y_3$ is an extension of Restriction B1 in which in the inequality (11) probabilities are conditional on X and on $Y_3 = y_3$. Further, the instrumental values must be chosen so that y_3 is a quantile of the distribution of Y_3 given X at all of the instrumental values. This may be difficult to achieve if X does not have rich support but arguments like those employed here still yield interval identification when y_3 cannot be set to a chosen quantile of Y_3 by choice of X . When Y_3 is discrete the extension of the results given here requires extension of Lemma 2 to provide bounds on $F_{Y_1|Y_2Y_3X}(y_1|y_2^m, y_3^n, x)$ and thus bounds on conditional quantiles of Y_1 .

5.3. Binary endogenous variables. The conditions employed in this paper exploit the ordering of discrete outcomes but when an endogenous variable is binary there is no ordering of outcomes. Nonparametric identification of features of nonseparable structural functions that depend on binary endogenous variables requires stronger conditions than those developed here.

5.4. Classical identification analysis. The classical analysis of the identification problem encapsulated in Koopmans, Rubin and Leipnik (1950) was set in the context of a linear simultaneous equations system. That analysis produced the well known *rank condition* which, under the restriction that latent variates are *mean* independent of covariates (instruments), is necessary and sufficient for the identification of the values of the coefficients of a linear structural function of the system.

This rank condition requires that there are covariates, variation in which has effect on the conditional means of the endogenous variables that appear in the structural function and has no *direct* effect on the structural functions.

The identification conditions deployed in this paper are extensions of these classical conditions, constructed within the framework for the study of parametric and nonparametric identification proposed by Hurwicz (1950).²¹ The mean independence covariation restriction of Koopmans, Rubin and Leipnik (1950) must be modified when a structural function is nonseparable in its latent variate because, when a structural function is nonseparable in its latent variate, conditional mean independence is an uninformative restriction unless there are highly restrictive conditions on the functional form of the structural function. If the structural function is restricted to be monotonic in its latent variate, a restriction of course satisfied in a separable model, a conditional *quantile* restriction is informative when a structural function is nonseparable.

The rank condition of the classical analysis appears here in the requirement that (a) there exist variation in covariates with an effect on the conditional distribution function of an endogenous variate given covariates of sufficient magnitude, as set out in Restriction B1, and (b) that the structural function is insensitive to this variation.

²¹This early development by Leonid Hurwicz is remarkable given that it was done at a time when linear separable parametric models preoccupied most people working in econometrics. The contribution is clearly acknowledged in Koopmans and Reiersøl (1950): “We shall therefore use the terms and concepts introduced by Hurwicz ... which cover both parametric and non-parametric specifications”. Hurwicz (1950) is one of five papers which Koopmans “draws freely on” in his expository paper Koopmans (1949), the others authored by Koopmans and Rubin, Wald, Koopmans and Reiersøl, and Leonid Hurwicz himself.

REFERENCES

- ANGRIST, JOSHUA, D., AND ALAN B. KRUEGER (1991): "Does compulsory schooling attendance affect schooling and earnings?" *Quarterly Journal of Economics*, 106, 979 - 1014.
- BECKER, GARY S., AND BARRY R. CHISWICK (1966): "Education and the distribution of earnings," *American Economic Review*, 56, 358-369.
- BOUND, JOHN, DAVID A. JAEGER AND REGINA M. BAKER (1995): "Problems with Instrumental Variables Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, 90, 443 - 450.
- CARD, DAVID (2001): "Estimating the returns to schooling: Progress on some persistent econometric problems," *Econometrica*, 69, 1127-1160.
- CHAUDHURI, PROBAL (1991): "Nonparametric estimation of regression quantiles and their local Bahadur representation," *Annals of Statistics*, 19, 760-777.
- CHAUDHURI, PROBAL, KJELL DOKSUM AND ALEXANDER SAMAROV (1997): "On average derivative quantile regression," *Annals of Statistics*, 25, 715-744.
- CHERNOZHUKOV, VICTOR, HAN HONG AND ELIE TAMER (2003): "Parameter Set Inference in a class of Econometric Models," manuscript.
- CHESHER, ANDREW D. (2002): "Instrumental values," Centre for Microdata Methods and Practice Working Paper 17/02.
- CHESHER, ANDREW D. (2003): "Identification in nonseparable models," *Econometrica*, 71, 1404 - 1441.
- CHISWICK, BARRY R. (1974): *Income inequality: regional analyses within a human capital framework*. New York: Columbia University Press.
- CHISWICK, BARRY R., AND JACOB MINCER (1972): "Time series changes in personal income inequality," *Journal of Political Economy*, 80, S34-S66.
- COCKBURN, IAIN, AND ZVI GRILICHES (1988): "The Estimation and Measurement of Spillover Effects of R&D Investment," *American Economic Review*, 78, No. 2, *Papers and Proceedings of the One-Hundredth Annual Meeting of the American Economic Association*, 419 - 423.
- DAROLLES, SERGE, JEAN-PIERRE FLORENS AND ERIC RENAULT, (2000): "Nonparametric instrumental regression," CREST Documents de Travail 2000-17.
- DAS, MITALI (2003): "Instrumental Variables Estimation of Nonparametric Models with Discrete Endogenous Variables," forthcoming, *Journal of Econometrics*.
- DUBIN, JEFFREY, A., DANIEL. L. MCFADDEN (1984): "An Econometric Analysis of Residential Electrical Appliance Holdings and Consumption," *Econometrica*, 52, 345 - 362.
- FLORENS, JEAN-PIERRE AND LAETITIA MALAVOLTI (2003): "Instrumental Regression with Discrete Variables," presented at the 2003 European Meeting of the Econometric Society, Stockholm.
- GRILICHES, ZVI (1990): "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature*, 28, 1661 - 1707.

- HURWICZ, LEONID (1950): "Generalization of the concept of identification," in *Statistical inference in dynamic economic models*. Cowles Commission Monograph 10, New York, John Wiley.
- IMBENS, GUIDO W., AND WHITNEY K. NEWEY (2003): "Identification and estimation of triangular simultaneous equations models without additivity," Manuscript, paper presented at the 14th EC² Meeting, cemmap, London, December 12th - 13th 2003.
- IMBENS, GUIDO W., AND CHARLES F. MANSKI (2003): "Confidence Intervals for Partially Identified Parameters," Centre for Microdata Methods and Practice Working Paper 09/03.
- KAHN, SHAKEEB (2001): "Two-stage rank estimation of quantile index models," *Journal of Econometrics*, 100, 319-355.
- KOENKER, ROGER W., AND GILBERT W. BASSETT JR. (1978): "Regression quantiles," *Econometrica*, 46, 33-50.
- KOENKER, ROGER W. AND VASCO D'OREY (1987): "Computing regression quantiles," *Journal of the Royal Statistical Society, Series C*, 36, 383-393.
- KOOPMANS, TJALLING C., (1949): "Identification Problems in Economic Model Construction," *Econometrica*, 17, 125 - 144.
- KOOPMANS, TJALLING C., AND OLAV REIERSØL (1950): "The identification of structural characteristics," *Annals of Mathematical Statistics*, 21, 165-181.
- KOOPMANS, TJALLING C., HERMAN RUBIN AND ROY B. LEIPNIK (1950): "Measuring the Equation Systems of Dynamic Economics," in *Statistical inference in dynamic economic models*. Cowles Commission Monograph 10, New York, John Wiley.
- LEE, SOKBAE (2003a): "Efficient Semiparametric Estimation of a Partially Linear Quantile Regression Model," *Econometric Theory*, 19, 1-31.
- LEE, SOKBAE (2003b): "Endogeneity in Quantile Regression Models: a Control Function Approach," Manuscript.
- MA, LINGJIE AND ROGER W. KOENKER (2003): "Quantile Regression Methods for Recursive Structural Equation Models," Manuscript.
- MACHADO, JOSÉ A.F., AND JOAO M.C. SANTOS SILVA (2002): "Quantiles for Counts," Centre for Microdata Methods and Practice Working Paper 22/02.
- MATZKIN, ROSA L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339 - 1376.
- MINCER, JACOB (1974): *Schooling experience and earnings*. New York: Columbia University Press.
- NEWEY, WHITNEY K., AND JAMES L. POWELL (2003): "Instrumental Variables Estimation for Nonparametric Models," *Econometrica*, 71, 1565 - 1578.
- NEWEY, WHITNEY K., JAMES L. POWELL, AND FRANK VELLA (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica* 67, 565-603.
- PINKSE, JORIS (2000): "Nonparametric two-step regression functions when regressors and errors are dependent," *Canadian Journal of Statistics*, 28, 289-300.

ROEHRIG, CHARLES S. (1988): “Conditions for identification in nonparametric and parametric models”, *Econometrica* 56, 433-447.

SEVERINI, THOMAS, AND GAUTAM TRIPATHI (2003): “Some Identification Issues in Nonparametric Linear Regression with Endogenous Regressors,” Manuscript.

APPENDIX: PROOFS OF THEOREMS

First, two lemmata employed in the proofs of the theorems are stated and proved. The inverse function h_1^{-1} will be used, defined as follows.

$$h_1^{-1}(y_2^m, x, y_1) \equiv \sup_p \{p : h_1(y_2^m, x, p) \leq y_1\}$$

When $h_1(y_2, x, \rho_1)$ varies *strictly* monotonically with ρ_1 this is the conventional inverse function satisfying

$$y_1 = h_1(y_2^m, x, h_1^{-1}(y_2^m, x, y_1))$$

for all y_2^m, x and y_1 .

Lemma 1

Under Restriction (I) the conditional distribution and τ_1 -quantile functions of Y_1 given $Y_2 = y_2^m$ and $X = x$ are as follows.

$$F_{Y_1|Y_2X}(y_1|y_2^m, x) = \frac{1}{p_m(x)} \int_{p^{m-1}(x)}^{p^m(x)} F_{\rho_1|\rho_2X}(h_1^{-1}(y_2^m, x, y_1)|r_2, x) dr_2$$

$$Q_{Y_1|Y_2X}(\tau_1|y_2^m, x) = \inf_q \left\{ q : \frac{1}{p_m(x)} \int_{p^{m-1}(x)}^{p^m(x)} F_{\rho_1|\rho_2X}(h_1^{-1}(y_2^m, x, q)|r_2, x) dr_2 \geq \tau_1 \right\}$$

Proof of Lemma 1

The conditional distribution function of Y_1 given $Y_2 = y_2^m$ and $X = x$ is defined as follows.

$$F_{Y_1|Y_2X}(y_1|y_2^m, x) \equiv \frac{P[Y_1 \leq y_1 \cap Y_2 = y_2^m | X = x]}{P[Y_2 = y_2^m | X = x]}$$

Given $X = x$, in terms of events, since h_1 is a non-decreasing function of ρ_1 ,

$$\{Y_1 \leq y_1 \cap Y_2 = y_2^m\} = \{\rho_1 \leq h_1^{-1}(y_2^m, x, y_1)\} \cap \{p^{m-1}(x) < \rho_2 \leq p^m(x)\}$$

and so

$$P[Y_1 \leq y_1 \cap Y_2 = y_2^m | X = x] = \int_{p^{m-1}(x)}^{p^m(x)} F_{\rho_1|\rho_2X}(h_1^{-1}(y_2^m, x, y_1)|r_2, x) dr_2 \quad (\text{L1.1})$$

and the expression for $F_{Y_1|Y_2X}(y_1|y_2^m, x)$ is delivered on dividing by $p_m(x) \equiv P[Y_2 = y_2^m | X = x]$. The expression for the conditional quantile function follows directly. \square

Lemma 2 places bounds on the conditional quantile function of Y_1 conditional on $Y_2 = y_2^m$ and $X = x$.

Lemma 2

Assume Restriction A1 holds. If for all r_1 , $F_{\rho_1|\rho_2 X}(r_1|r_2, x)$ is monotonic in r_2 for $r_2 \in (p^{m-1}(x), p^m(x))$ then

$$\begin{aligned} \min \left\{ \begin{array}{l} h_1(y_2^m, x, Q_{\rho_1|\rho_2 X}(\tau_1|p^{m-1}(x), x)) \\ h_1(y_2^m, x, Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x), x)) \end{array} \right\} \\ \leq Q_{Y_1|Y_2 X}(\tau_1|y_2^m, x) \leq \\ \max \left\{ \begin{array}{l} h_1(y_2^m, x, Q_{\rho_1|\rho_2 X}(\tau_1|p^{m-1}(x), x)) \\ h_1(y_2^m, x, Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x), x)) \end{array} \right\}. \end{aligned}$$

If ρ_1 and ρ_2 are independently distributed given $X = x$ over the interval $[p^{m-1}(x), p^m(x)]$ in the sense that for r_2 in this interval $Q_{\rho_1|\rho_2 X}(\tau_1|r_2, x) = r_1^*(x)$, a constant, then:

$$Q_{Y_1|Y_2 X}(\tau_1|y_2^m, x) = h_1(y_2^m, x, r_1^*(x)).$$

Remark A1. The weak monotonicity condition on $F_{\rho_1|\rho_2 X}(r_1|r_2, x)$ is sufficient but not necessary, a degree of bounded variation being all that is required. Specifically, the result of Lemma 2 also applies if, for all r_1 in a wide enough interval around $Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x), x)$, there exists $\theta \in [0, 1]$ such that for all $r_2 \in (p^{m-1}(x), p^m(x))$

$$F_{\rho_1|\rho_2 X}(r_1|r_2, x) = \theta F_{\rho_1|\rho_2 X}(r_1|p^{m-1}(x), x) + (1 - \theta)F_{\rho_1|\rho_2 X}(r_1|p^m(x), x).$$

Remark A2. If the structural function depends on both latent variates, thus: $h_1(Y_2, X, \rho_1, \rho_2)$, then the result of Lemma 2 applies if for all t_1 in a wide enough interval around τ_1 , there exists $\theta \in [0, 1]$ such that for all $r_2 \in (p^{m-1}(x), p^m(x))$

$$\begin{aligned} h_1(y_2^m, x, Q_{\rho_1|\rho_2 X}(\tau_1|r_2, x), r_2) &= \theta h_1(y_2^m, x, Q_{\rho_1|\rho_2 X}(\tau_1|p^{m-1}(x), x), p^{m-1}(x)) \\ &\quad + (1 - \theta)h_1(y_2^m, x, Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x), x), p^m(x)) \end{aligned}$$

for which the monotonicity of $h_1(y_2^m, x, Q_{\rho_1|\rho_2 X}(\tau_1|r_2, x), r_2)$ with respect to r_2 is sufficient.

Remark A3. The bounds of Lemma 2 apply when $m = M$ but note that in that case $p^M(x) = 1$ for all x . They also apply when $m = 1$ with $p^0(x) \equiv 0$.

Proof of Lemma 2

First consider the case in which $F_{\rho_1|\rho_2 X}$ is non-decreasing in ρ_2 for $\rho_2 \in (p^{m-1}(x), p^m(x))$. Recalling the result of Lemma 1, replacing ρ_2 in the integrand first by $p^{m-1}(x)$ and then by $p^m(x)$ yields the following inequality.

$$\frac{1}{p_m(x)} \int_{p^{m-1}(x)}^{p^m(x)} F_{\rho_1|\rho_2 X}(h_1^{-1}(y_2^m, x, y_1)|p^{m-1}(x), x) d\rho_2$$

$$\begin{aligned} &\leq F_{Y_1|Y_2X}(y_1|y_2^m, x) \leq \\ &\frac{1}{p_m(x)} \int_{p^{m-1}(x)}^{p^m(x)} F_{\rho_1|\rho_2X}(h_1^{-1}(y_2^m, x, y_1)|p^m(x), x) d\rho_2 \end{aligned}$$

Since, for example,

$$\begin{aligned} &\frac{1}{p_m(x)} \int_{p^{m-1}(x)}^{p^m(x)} F_{\rho_1|\rho_2X}(h_1^{-1}(y_2^m, x, y_1)|p^m(x), x) d\rho_2 \\ &= F_{\rho_1|\rho_2X}(h_1^{-1}(y_2^m, x, y_1)|p^m(x), x) \end{aligned}$$

there is, for all y_1 , the following inequality.

$$\begin{aligned} &F_{\rho_1|\rho_2X}(h_1^{-1}(y_2^m, x, y_1)|p^{m-1}(x), x) \\ &\leq F_{Y_1|Y_2X}(y_1|y_2^m, x) \leq \\ &F_{\rho_1|\rho_2X}(h_1^{-1}(y_2^m, x, y_1)|p^m(x), x) \end{aligned} \tag{L2.1}$$

When $F_{\rho_1|\rho_2X}$ is non-increasing in ρ_2 for $\rho_2 \in (p^{m-1}(x), p^m(x)]$ the inequality (L2.1) is reversed.

Since the functions on the left and the right of the inequality (L2.1) bound $F_{Y_1|Y_2X}(y_1|y_2^m, x)$ respectively *below* and *above*, the values, q_{m-1} and q_m defined by:

$$\begin{aligned} q_{m-1} &\equiv \inf_q \{q : F_{\rho_1|\rho_2X}(h_1^{-1}(y_2^m, x, q)|p^{m-1}(x), x) \geq \tau_1\} \\ q_m &\equiv \inf_q \{q : F_{\rho_1|\rho_2X}(h_1^{-1}(y_2^m, x, q)|p^m(x), x) \geq \tau_1\} \end{aligned}$$

bound

$$q \equiv Q_{Y_1|Y_2X}(\tau_1|y_2^m, x) \equiv \inf_q \{q : F_{Y_1|Y_2X}(q|y_2^m, x) \geq \tau_1\}$$

respectively *above* and *below*, that is:

$$q_m \leq q \leq q_{m-1}.$$

The final step is to obtain the relationships between the bounding functions, q_{m-1} and q_m , and the structural function h_1 .

First consider q_m - a similar argument will apply for q_{m-1} . For any choice of q^*

$$F_{\rho_1|\rho_2X}(h_1^{-1}(y_2^m, x, q^*)|p^m(x), x) \geq \tau_1$$

if and only if

$$Q_{\rho_1|\rho_2X}(\tau_1|p^m(x), x) \leq h_1^{-1}(y_2^m, x, q^*).$$

Therefore, recalling the definition of h_1^{-1} given at the start of the Appendix,

$$q_m = \inf_q \{q : Q_{\rho_1|\rho_2X}(\tau_1|p^m(x), x) \leq \sup_p \{p : h_1(y_2^m, x, p) \leq q\}\}.$$

When h_1 is a *strictly* increasing function of ρ_1

$$q_m = \inf_q \{q : Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x), x) \leq h_1^{-1}(y_2^m, x, q)\}$$

and therefore

$$q_m = h_1(y_2^m, x, Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x), x)).$$

It is now shown that the same expression obtains when h_1 is a *weakly* increasing function of ρ_1 .

Consider

$$q^+ \equiv h_1(y_2^m, x, Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x), x))$$

which is a candidate value for q_m in the weakly increasing case because,

$$\sup_p \{p : h_1(y_2^m, x, p) \leq q^+\} \geq Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x), x), \quad (\text{L2.2})$$

as Figure 4 makes clear.

In Figure 4 a non-decreasing caglad step function of ρ_1 , h_1 , is drawn with ends of steps at which h_1 is continuous from the left drawn as filled circles. The graph shows a value of $Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x), x)$ and the candidate value q^+ . The value of

$$p^+ \equiv \sup_p \{p : h_1(y_2^m, x, p) \leq q^+\}$$

is indicated in Figure 4 and clearly $p^+ \geq Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x), x)$, demonstrating that the inequality (L2.2) does hold.

It is evident from Figure 4 that for $q < q^+$ there is no $p \geq Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x), x)$ such that

$$\sup_p \{p : h_1(y_2^m, x, p) \leq q\} \geq Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x), x)$$

and so $q_m = q^+$.

A similar argument produces the following expression for q_{m-1} .

$$q_{m-1} = h_1(y_2^m, x, Q_{\rho_1|\rho_2 X}(\tau_1|p^{m-1}(x), x))$$

Therefore, in the case in which $F_{\rho_1|\rho_2 X}$ is non-decreasing in ρ_2 for $\rho_2 \in (p^{m-1}(x), p^m(x)]$, there is the following inequality.

$$\begin{aligned} & h_1(y_2^m, x, Q_{\rho_1|\rho_2 X}(\tau_1|p^m(x), x)) \\ & \leq Q_{Y_1|Y_2 X}(\tau_1|y_2^m, x) \leq \\ & h_1(y_2^m, x, Q_{\rho_1|\rho_2 X}(\tau_1|p^{m-1}(x), x)) \end{aligned} \quad (\text{L2.3})$$

A similar argument for the case in which $F_{\rho_1|\rho_2 X}$ is non-increasing in ρ_2 for $\rho_2 \in (p^{m-1}(x), p^m(x)]$ produces the reverse of this inequality

$$h_1(y_2^m, x, Q_{\rho_1|\rho_2 X}(\tau_1|p^{m-1}(x), x))$$

$$\begin{aligned} &\leq Q_{Y_1|Y_2X}(\tau_1|y_2^m, x) \leq \\ &h_1(y_2^m, x, Q_{\rho_1|\rho_2X}(\tau_1|p^m(x), x)) \end{aligned} \quad (\text{L2.4})$$

and combining (L2.3) and (L2.4) gives the result of the first part of the Lemma.

Suppose $Q_{\rho_1|\rho_2X}(\tau_1|r_2, x)$ is constant and equal to $r_1^*(x)$ for variation in r_2 over the interval $[p^{m-1}(x), p^m(x)]$. Then in each of (L2.3) and (L2.4) the lower and upper bounds are equal and so:

$$\begin{aligned} Q_{Y_1|Y_2X}(\tau_1|y_2^m, x) &= h_1(y_2^m, x, Q_{\rho_1|\rho_2X}(\tau_1|p^{m-1}(x), x)) \\ &= h_1(y_2^m, x, Q_{\rho_1|\rho_2X}(\tau_1|p^m(x), x)) \\ &= h_1(y_2^m, x, r_1^*(x)) \end{aligned}$$

which gives the result of the final part of the Lemma. \square

Proof of Theorem 1.

Consider the case in which for all r_1 , $F_{\rho_1\rho_2X}(r_1|r_2, x)$ is a weakly *increasing* function of r_2 for $r_2 \in R_2$. In this case, for all τ , $Q_{\rho_1\rho_2X}(\tau|r_2, x)$ is a weakly decreasing function of r_2 for $r_2 \in R_2$.

Lemma 2 implies the following inequalities.

$$Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^{m-1}) \leq h_1(y_2^m, x^{m-1}, Q_{\rho_1|\rho_2X}(\tau_1|p^{m-1}(x^{m-1}), x^{m-1})) \quad (\text{T1.1})$$

$$h_1(y_2^m, x^m, Q_{\rho_1|\rho_2X}(\tau_1|p^m(x^m), x^m)) \leq Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^m) \quad (\text{T1.2})$$

Because $p^m(x^m) \leq \tau_2 \leq p^{m-1}(x^{m-1})$ and in the case under consideration $Q_{\rho_1\rho_2X}(\tau_1|r_2, x)$ is a weakly decreasing function of r_2 for $r_2 \in R_2$ there are the inequalities

$$Q_{\rho_1\rho_2X}(\tau_1|p^{m-1}(x^{m-1}), x^{m-1}) \leq Q_{\rho_1\rho_2X}(\tau_1|\tau_2, x^{m-1})$$

$$Q_{\rho_1\rho_2X}(\tau_1|\tau_2, x^m) \leq Q_{\rho_1\rho_2X}(\tau_1|p^m(x^m), x^m)$$

and because h_1 is monotonic in ρ_1 and normalised non-decreasing there are the following inequalities.

$$h_1(y_2^m, x^{m-1}, Q_{\rho_1|\rho_2X}(\tau_1|p^{m-1}(x^{m-1}), x^{m-1})) \leq h_1(y_2^m, x^{m-1}, Q_{\rho_1|\rho_2X}(\tau_1|\tau_2, x^{m-1})) \quad (\text{T1.3})$$

$$h_1(y_2^m, x^m, Q_{\rho_1|\rho_2X}(\tau_1|\tau_2, x^m)) \leq h_1(y_2^m, x^m, Q_{\rho_1|\rho_2X}(\tau_1|p^m(x^m), x^m)) \quad (\text{T1.4})$$

Combining (T1.1) with (T1.3) and (T1.2) with (T1.4), produces the following inequalities.

$$Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^{m-1}) \leq h_1(y_2^m, x^{m-1}, Q_{\rho_1|\rho_2X}(\tau_1|\tau_2, x^{m-1})) \quad (\text{T1.5})$$

$$h_1(y_2^m, x^m, Q_{\rho_1|\rho_2X}(\tau_1|\tau_2, x^m)) \leq Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^m)$$

Finally, under the covariation and exclusion restriction, B3 and B4 there is the following inequality for $x \in \{x^{m-1}, x^m\}$.

$$Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^{m-1}) \leq h_1(y_2^m, x, r_1^*) \leq Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^m)$$

Arguing as above for the case in which for all r_1 , $F_{\rho_1\rho_2X}(r_1|r_2, x)$ is a weakly decreasing function of r_2 for $r_2 \in R_2$ there is

$$Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^m) \leq h_1(y_2^m, x, r_1^*) \leq Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^{m-1})$$

and combining the two cases gives the inequality of Theorem 1. \square

Proof of Theorem 2

Since τ_2 lies in R_2 and

$$\begin{aligned} (p^{m-1}(x^{m-1}), p^m(x^{m-1})) &\subseteq R_2 \\ (p^{m-1}(x^m), p^m(x^m)) &\subseteq R_2 \end{aligned}$$

the constant value of $F_{\rho_1\rho_2X}(\tau_1|r_2, x)$ in Restriction B2* is $r_1^* \equiv F_{\rho_1\rho_2X}(\tau_1|\tau_2, x)$ which, by Restriction B3 is invariant with respect to choice of $x \in \tilde{x}_m$. The final result of Lemma 2 implies that

$$\begin{aligned} Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^m) &= h_1(y_2^m, x, r_1^*) \\ Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^{m-1}) &= h_1(y_2^m, x, r_1^*) \end{aligned}$$

the terms on the right hand sides of these equations being invariant with respect to choice of $x \in \tilde{x}_m$ by virtue of Restriction B4. \square

Lemma 3 is used in the proof of Theorem 3.

Lemma 3

Assume h_1 is strictly increasing in ρ_1 . If, for all r_1 , $F_{\rho_1|\rho_2X}(r_1|r_2, x)$ is weakly monotonic and non-constant in r_2 for $r_2 \in (p^{m-1}(x), p^m(x))$ over a set of values of nonzero measure then the inequalities of Lemma 2 are strong.

Proof of Lemma 3

Consider the case in which $F_{\rho_1|\rho_2X}(r_1|r_2, x)$ is non-decreasing for $r_2 \in (p^{m-1}(x), p^m(x))$. Then with $r_1 = h_1^{-1}(y_2^m, x, y_1)$:

$$\begin{aligned} F_{\rho_1|\rho_2X}(r_1|r_2, x) &= F_{\rho_1|\rho_2X}(r_1|p^{m-1}(x), x) + A^{m-1}(r_1, r_2, x) \\ F_{\rho_1|\rho_2X}(r_1|r_2, x) &= F_{\rho_1|\rho_2X}(r_1|p^m(x), x) + A^m(r_1, r_2, x) \end{aligned}$$

with $A^{m-1} > 0$ and $A^m < 0$ for some $r_2 \in (p^{m-1}(x), p^m(x))$ in a set of values of nonzero measure. The following strong inequalities result.

$$F_{\rho_1|\rho_2X}(h_1^{-1}(y_2^m, x, y_1)|p^{m-1}(x), x)$$

$$\begin{aligned} &< F_{Y_1|Y_2X}(y_1|y_2^m, x) < \\ &F_{\rho_1|\rho_2X}(h_1^{-1}(y_2^m, x, y_1)|p^m(x), x) \end{aligned}$$

The inequalities are reversed if $F_{\rho_1|\rho_2X}(r_1|r_2, x)$ is non-increasing for $r_2 \in (p^{m-1}(x), p^m(x)]$.

Considering the solutions obtained on setting each term in the preceding inequalities equal to τ_1 there is, arguing as in Lemma 2,

$$\begin{aligned} &h_1(y_2^m, x, Q_{\rho_1|\rho_2X}(\tau_1|p^m(x), x)) \\ &< Q_{Y_1|Y_2X}(\tau_1|y_2^m, x) < \\ &h_1(y_2^m, x, Q_{\rho_1|\rho_2X}(\tau_1|p^{m-1}(x), x)) \end{aligned}$$

with the inequalities reversed if $F_{\rho_1|\rho_2X}(r_1|r_2, x)$ is non-increasing for $r_2 \in (p^{m-1}(x), p^m(x)]$.

Thus the inequalities of Lemma 2 are strong under the stated conditions. \square

Remark A4. The conditions of Lemma 3 are somewhat stronger than required since the non-constancy of $F_{\rho_1|\rho_2X}(r_1|r_2, x)$ need only hold for r_1 in a sufficiently large region containing $h_1^{-1}(y_2^m, x, Q_{Y_1|Y_2X}(\tau_1|y_2^m, x))$.

Proof of Theorem 3

Consider the case in which $F_{\rho_1|\rho_2X}(r_1|r_2, x)$ is a weakly *increasing* function of r_2 for $r_2 \in R_2$.

Under Restriction B2** the conditions of Lemma 3 are satisfied. Therefore in the proof of Theorem 1, the weak inequalities (T1.1) and (T1.2) are strong. Therefore the inequalities (T1.3) and (T1.4) are strong.

The same argument applies, with inequalities reversed, in the case in which $F_{\rho_1|\rho_2X}(r_1|r_2, x)$ is a weakly *decreasing* function of r_2 for $r_2 \in R_2$. It follows that the inequalities of Theorem 1 are strong and therefore that

$$Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^{m-1}) \neq Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^m).$$

\square

Figure 1: Estimated distribution functions of discrete years of schooling for quarter of birth 1 and 4

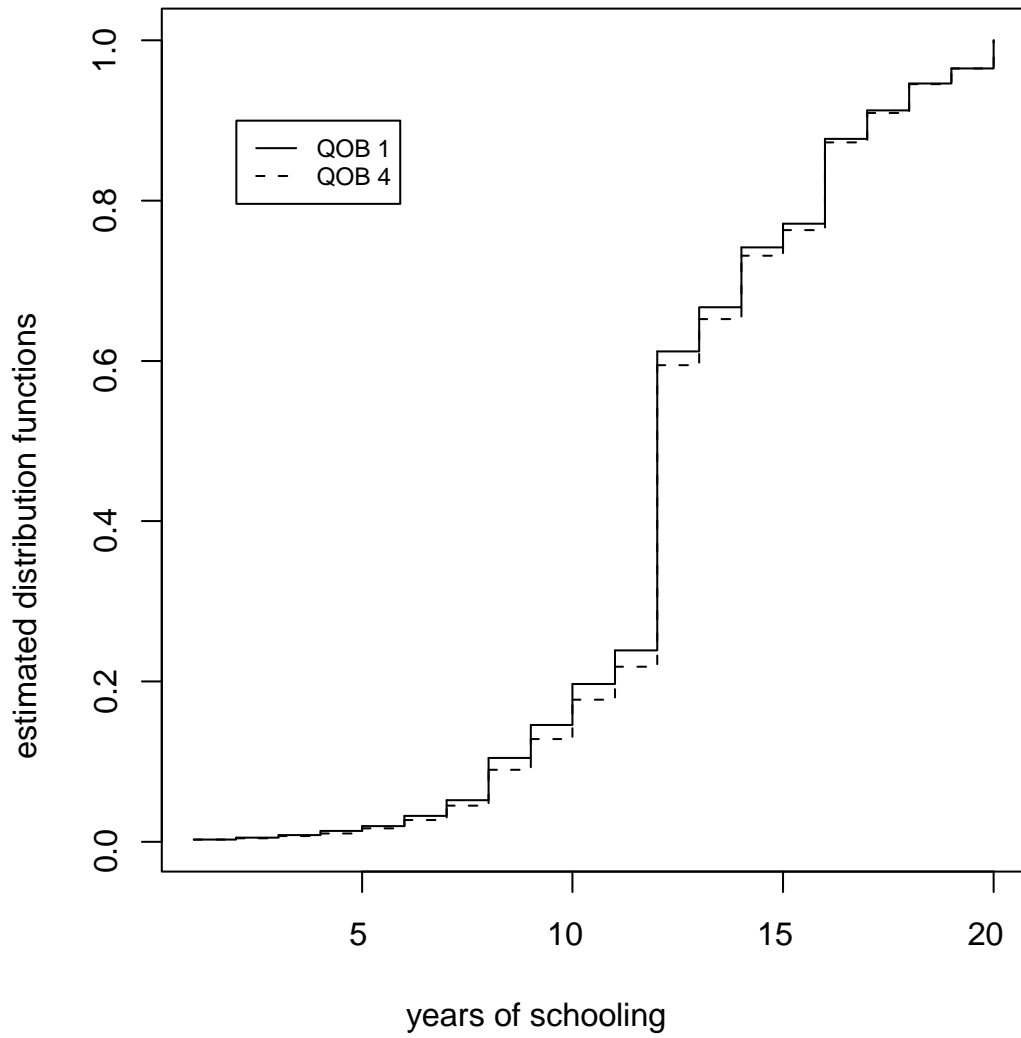


Figure 2: Conjectured distribution functions of continuous years of schooling for quarter of birth 1 and 4

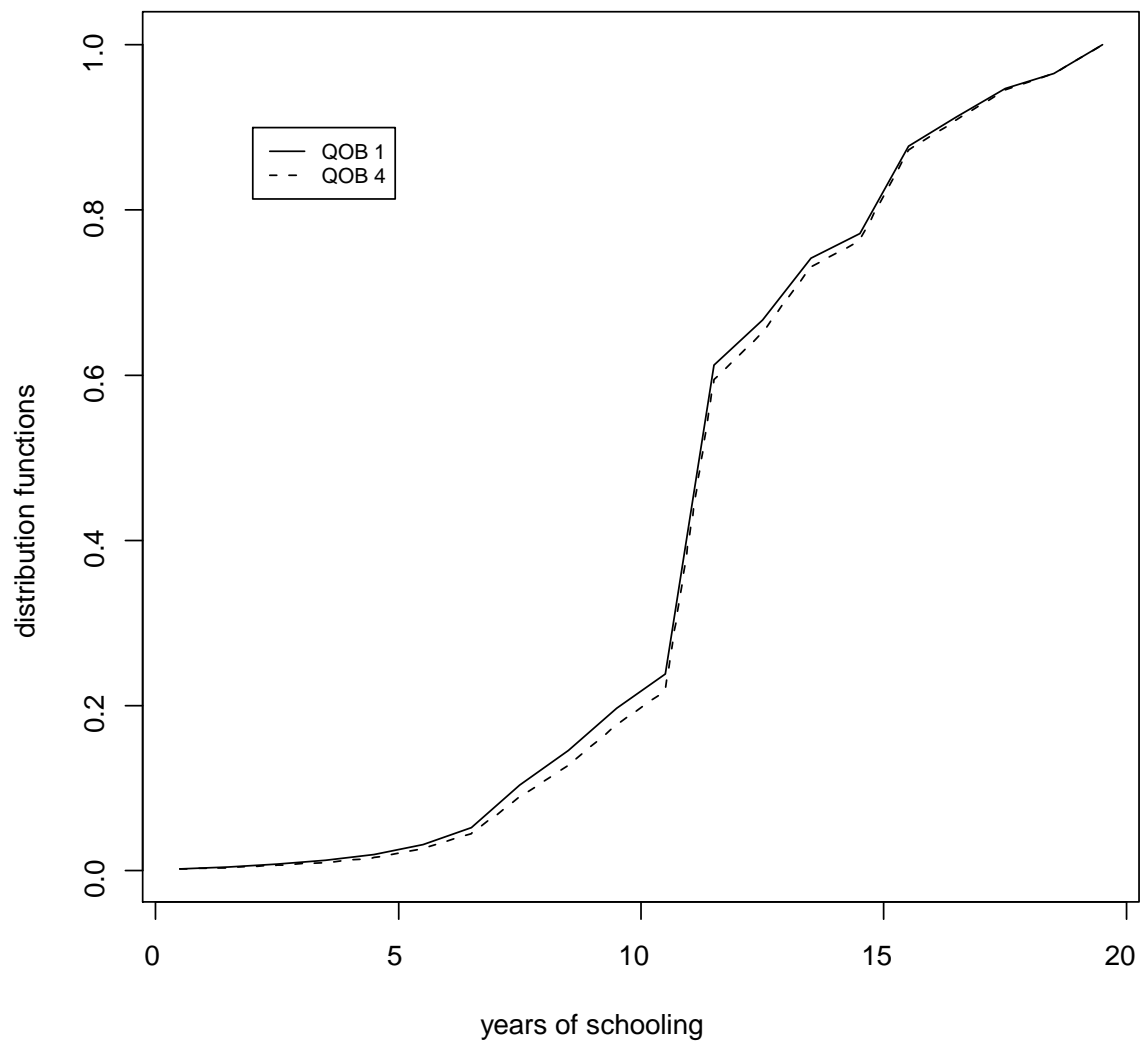
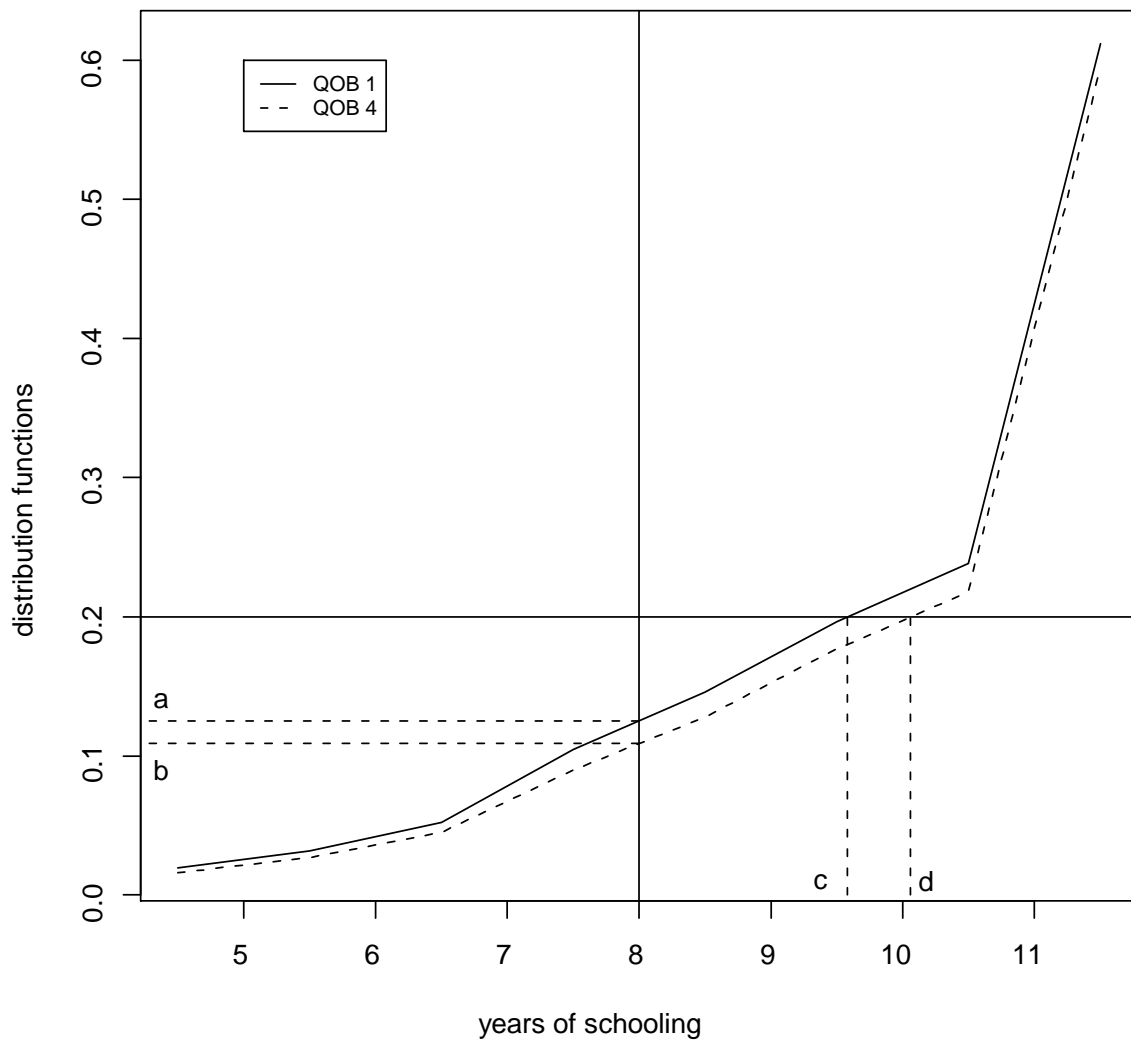


Figure 3: Conjectured distribution functions of continuous years of schooling (5 - 12 years) for quarter of birth 1 and 4



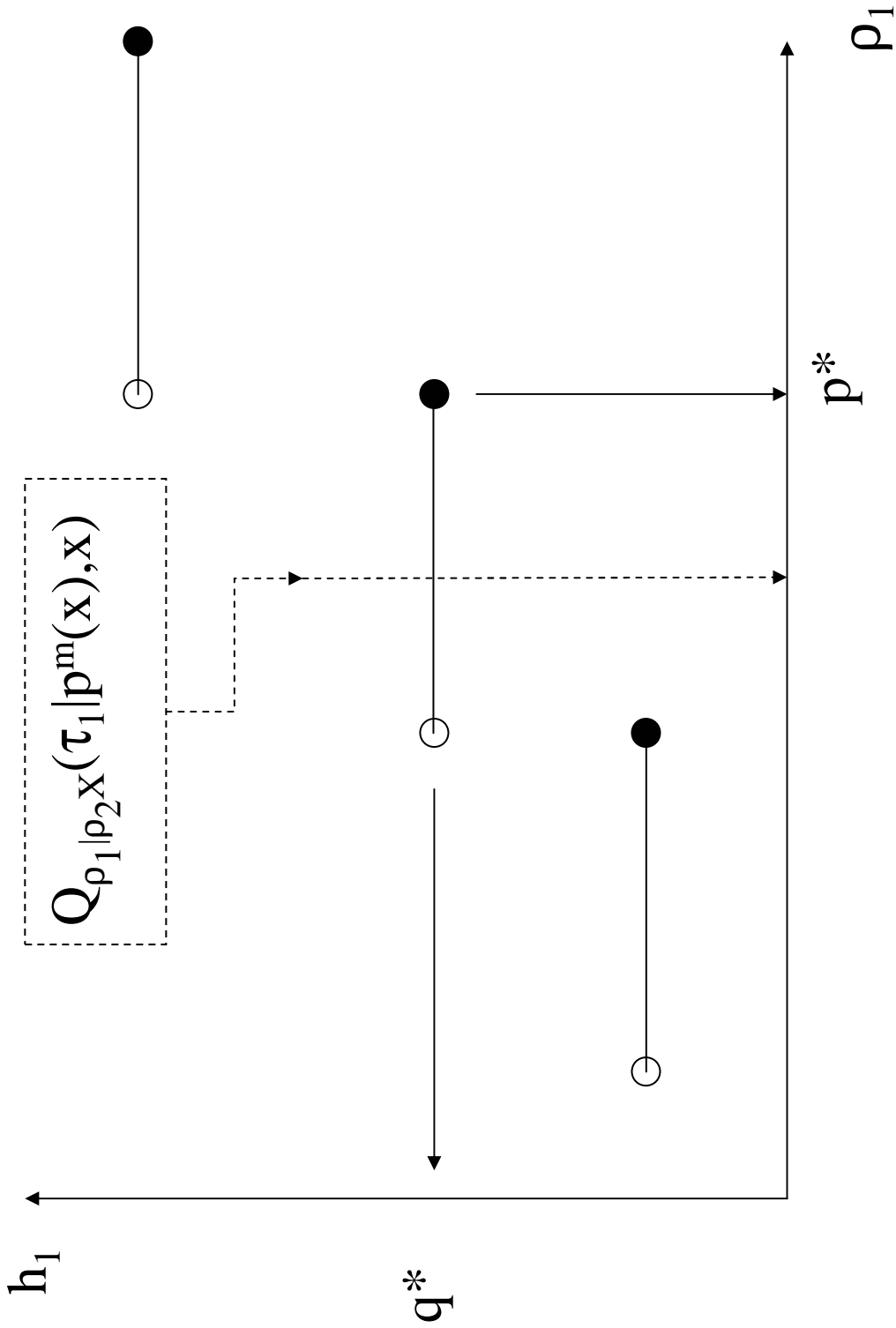


Figure 4: Proof of Lemma 2 – discontinuous h_1