CHAPTER 1

Identification of Nonadditive Structural Functions*

Andrew Chesher†

1 INTRODUCTION

When latent variates appear nonadditively in a structural function the effect of a *ceteris paribus* change in an observable argument of the function can vary across people measured as identical. Models that admit nonadditive structural functions permit responses to policy interventions to have probability distributions. Knowledge of the distributions of responses is important for welfare analysis and it is good to know what conditions secure identification of these distributions. This lecture examines some aspects of this problem.

Early studies of identification in econometrics dealt almost exclusively with linear additive "error" models. The subsequent study of identification in nonlinear models was heavily focused on additive error models until quite recently[1] and only within the last ten years has there has been extensive study of identification in nonadditive error models. This lecture examines some of these recent results, concentrating on models which admit no more sources of stochastic variation than there are observable stochastic outcomes. Models with this property are interesting because they are direct generalizations of additive error models and of the classical linear simultaneous equation models associated with the work of the Cowles Commission, and because the addition of relatively weak nonparametric restrictions results in models which identify complete structural functions or specific local features of them.

Nonparametric restrictions are interesting because they are consonant with the information content of economic theory. Even if parametric or semiparametric restrictions are imposed when estimation and inference are done, it is good to know nonparametric identification conditions because they tell us what core elements of the model are essential for identification and which are in

* I thank Lars Nesheim and Whitney Newey for helpful comments. I am grateful for generous support from the Leverhulme Trust through their funding of the Centre for Microdata Methods and Practice and the research project "Evidence, Inference and Enquiry."
† Centre for Microdata Methods and Practice, IFS and UCL.
[1] See, for example, the review in Blundell and Powell (2003).

principle falsifiable. If just-identifying conditions can be found then we know what must be believed if the result of econometric analysis is to be given economic meaning.

Nonadditive error models permit covariates to influence many aspects of the distributions of outcomes. Koenker (2005), motivating the study of estimation of quantile regression functions, argues persuasively for consideration of models more general than the classical location shift, additive error model, that took center stage in the early history of econometrics. Models for discrete outcomes have a natural expression in terms of nonadditive structural functions as do many microeconometric models for continuous outcomes, for example, the following model for durations with heterogeneity.

Example 1 *In the analysis of a continuous duration, Y_1, with distribution function $F_{Y_1|X}$ conditional on X, it is common to use a proportionate hazard model in which the hazard function:*

$$h(y, x) \equiv -\nabla_y \log(1 - F_{Y_1|X}(y|x))$$

is restricted to have the multiplicatively separable form $\lambda(y)g(x)$. With $U_1|X \sim Unif(0, 1)$ values of Y_1 with conditional distribution function $F_{Y_1|X}$ are generated by:

$$Y_1 = \Lambda^{-1}(-\log(1 - U_1)/g(X))$$

in which the structural function is generally nonadditive. Here Λ^{-1} is the inverse of the integrated hazard function: $\Lambda(y) \equiv \int_0^y \lambda(s)ds$. Classical censoring nests the function Λ^{-1} inside a step function. In the mixed, or heterogeneous, proportionate hazard model the hazard function conditional on X and unobserved U_2 is specified as multiplicatively separable: $\lambda(y)g(x)U_2$ and there is the structural equation:

$$Y_1 = \Lambda^{-1}(-U_2^{-1} \log(1 - U_1)/g(X))$$

in which, note, there is effectively just one stochastic term: $V \equiv U_2^{-1} \log(1 - U_1)$. Endogeneity can be introduced by allowing U_2 and elements of X to be jointly dependent.

Imbens and Newey (2003) provide examples of economic contexts in which nonadditive structural functions arise naturally and are an essential feature of an economic problem if endogeneity is to be present. Card (2001) provides a parametric example.

As set out below, identification conditions in nonadditive models are rather natural extensions of those employed in additive models involving, on the one hand, conditions on the sensitivity of structural functions to variation in certain variables and, on the other, restrictions on the dependence of those variables and the latent variables. Local quantile independence conditions are helpful in identifying certain local features of structural functions. Full independence is commonly imposed in order to identify complete structural functions.

The precise form of the independence condition employed has significant implications for the way in which identification is *achieved*, by which I mean the way in which information about a structural function is carried in the distribution of outcomes about which data is informative. That in turn has implications for the challenges to be faced when developing estimators and conducting inference.

In the context of identification of a single structural function involving two jointly dependent outcomes the two contrasting cases considered here involve an independence restriction involving (i) a single latent variable and (ii) two latent variables. The latter conveys more information, is more tolerant of limitations in the support of instruments, facilitates identification of local structural features when global features may not be identifiable and motivates more benign estimation and inference. However, these benefits come at the cost of a more restrictive model whose conditions are difficult to justify in some economic problems.

Sections 2 and 3 consider identification of nonadditive structural functions under the two types of condition. Section 4 concludes.

2 MARGINAL INDEPENDENCE

In the first class of models considered the structural function delivering the value of an outcome Y_1 involves a single latent random variable.

$$Y_1 = h(Y_2, X, U_1)$$

The function h is restricted to be strictly monotonic in U_1 and is normalized increasing. There are covariates, X, whose appearance in h will be subject to exclusion-type restrictions that limit the sensitivity of h to variations in X. The variable Y_2 is an observed outcome which may be jointly dependent with U_1 and thus "endogenous."

The latent variable, U_1, and X are restricted to be independently distributed. The independence condition can be weakened to τ-quantile independence, that is, that $q_\tau \equiv Q_{U_1|X}(\tau|x)$ be independent of x for some value of τ. Then there can be identification of $h(Y_2, X, q_\tau)$. In all cases of econometric interest U_1 will be continuously distributed and the strict monotonicity condition on h then implies that Y_1 given X is continuously distributed.

This is in the family of models studied in Chernozhukov and Hansen (2005), and is essentially the model studied in Chernozhukov, Imbens, and Newey (2007). Matzkin (2003) studies the special case in which Y_2 is "exogenous," independent of U_1.

There is, using the independence condition at the first step, and the monotonicity restriction at the second, for all x:

$$\begin{aligned}\tau &= P[U_1 \leq q_\tau | X = x] \\ &= P[h(Y_2, X, U_1) \leq h(Y_2, X, q_\tau) | X = x] \\ &= P[Y_1 \leq h(Y_2, X, q_\tau) | X = x].\end{aligned}$$

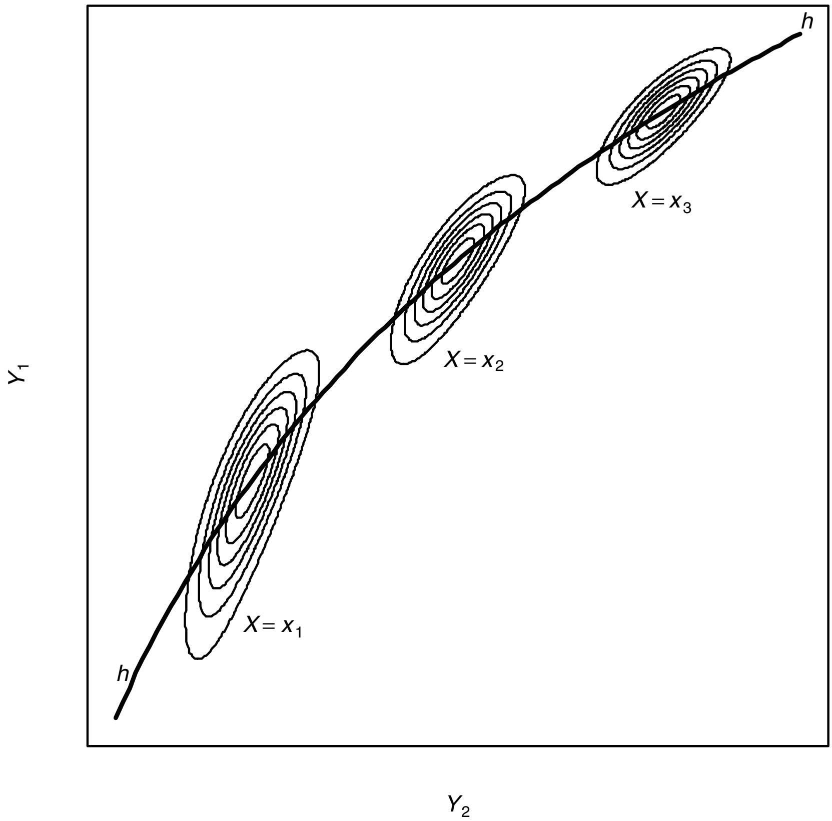

Figure 1.1. Contours of a joint density function of Y_1 and Y_2 conditional on X at three values of X. The line marked hh is the structural function $h(Y_2, X, q_\tau)$, not varying across $X \in \{x_1, x_2, x_3\}$, drawn with $q_\tau = 0.5$.

So the function $h(Y_2, X, q_\tau)$ is such that the moment condition:

$$E_{Y_1 Y_2|X}[\mathbf{1}[Y_1 \leq h(Y_2, x, q_\tau)] - \tau|x] = 0 \tag{1.1}$$

is satisfied for all x.

The structural function $h(\cdot, \cdot, q_\tau)$ satisfies (1.1) for all x in the support of X. It is identified if there are restrictions on the structural function, h, the joint distribution of U_1 and Y_2 given X and the support of X sufficient to ensure that no function distinct from h satisfies this condition for all x. Chernozhukov, Imbens, and Newey (2007) show that with a joint normality restriction there is an identification condition like the classical rank condition. In the absence of some parametric restrictions it is difficult to find primitive identification conditions which can be entirely motivated by economic considerations.

Figure 1.1 illustrates the situation. It shows contours of a joint distribution of Y_1 and Y_2, conditional on X, at three values of X: $\{x_1, x_2, x_3\}$. The joint distribution does vary with X and there is an exclusion restriction so that the structural function $h(\cdot, x, q_\tau)$, marked hh, is unaffected by variation in X across $x \in \{x_1, x_2, x_3\}$. The value q_τ is the τ-quantile of U_1 given X which is restricted to be invariant with respect to $x \in \{x_1, x_2, x_3\}$ and the result is that at each value of x there is the same probability mass (τ) falling below the structural function. In the case illustrated here $\tau = 0.5$. If the structural function h is the only function which has that property then h is identified.

Clearly, there could be many functions with this property, and so no identification of h, if there were, as illustrated, only three values of X at which the various conditions hold. It is evident that identification of h (in the absence of parametric restrictions) is critically dependent on the nature of the support of X which must be at least as rich as the support of the endogenous Y_2. If Y_2 is binary then binary X may suffice for identification and if X has richer support there can be nonparametric overidentification. If Y_2 is continuous then X must be continuous if h is to be identified.

The role in identification of the support of instruments is easily appreciated in the case in which there is discrete $Y_2 \in \{y_2^1, \ldots, y_2^M\}$ and discrete $X \in \{x_1, \ldots, x_J]$ and X does not feature in h. The values of h at the M points of support of Y_2 and at $U_1 = q_\tau$ are denoted by $\theta_\tau^m \equiv h(y_2^m, q_\tau)$, $m \in \{1, \ldots, M\}$.

The J values of X yield equations which are *nonlinear* in the terms θ_τ^m

$$\sum_{m=1}^{M} G(\theta_\tau^m, y_2^m | x_j) = \tau \qquad j \in \{1, \ldots, J\} \tag{1.2}$$

where G is a conditional probability distribution – probability mass function defined as follows.

$$G(a, b|x) \equiv P[Y_1 \leq a \cap Y_2 = b | X = x]$$

The M values of the θ_τ^m's do satisfy (1.2) because those values are instrumental in determining the function G. However, without further restriction there can be other solutions to these nonlinear simultaneous equations and the θ_τ^m's will not be identified. Without further restriction there must be as many equations as unknowns and so the requirement that the support of X is at least as rich as the support of discrete Y_2. The J equations must be distinct which requires that X does have influence on the distribution of Y_2 given X. Conditions like the classical instrumental variables inclusion and exclusion restrictions are necessary. Chernozhukov and Hansen (2005) give a rank condition sufficient for local identification.

When X has sparse support relative to Y_2 and $P[Y_2 = y_2^m | x] \in (0, 1)$ for all x then no value of h is point identified. Parametric restrictions on h can lead to point identification in this case. When Y_2 is continuous there is generally point identification of the structural function nowhere unless instruments have continuous variation.

In the additive error model $Y_1 = h(Y_2, X) + U_1$ with the conditional expectation restriction: $E[U_1 | X = x] = 0$ the moment condition corresponding to (1.1) is

$$E_{Y_1 Y_2 | X}[Y_1 - h(Y_2, x, q_\tau) | x] = 0$$

which reduces to

$$E_{Y_1|X}[Y_1|x] - E_{Y_2|X}[h(Y_2, x)]|x] = 0$$

and to *linear* equations in the values of h when Y_2 is discrete, a case studied in Das (2005) and Florens and Malavolti (2003). When Y_2 is continuous the structural function h is the solution to an integral equation which constitutes an ill-posed *linear* inverse problem and leads to challenging problems in estimation and inference studied in, for example, Darolles, Florens, and Renault (2003), Florens (2003), Blundell and Powell (2003), Hall and Horowitz (2005).

In the nonadditive case with continuous Y_2 the structural function h is a solution to the integral equation

$$\int_{-\infty}^{\infty} G(h(y_2, x, q_\tau), y_2|x) dy_2 = \tau \tag{1.3}$$

which constitutes an ill-posed *nonlinear* inverse problem with significant challenges for estimation and inference studied in Chernozhukov, Imbens, and Newey (2007), Chernozhukov and Hong (2003), and Chernozhukov and Hansen (2005).

The inverse problem (1.3) is nonlinear because of the use of quantile independence conditions and arises in the additive error case as well if a quantile independence rather than a mean independence condition is employed. In the additive error model there is, with quantile independence:

$$\int_{-\infty}^{\infty} G(h(y_2, x) + q_\tau, y_2|x) dy_2 = \tau$$

and there can be overidentification because $h(y_2, x)$ may be identified up to an additive constant at any value of τ.

3 JOINT INDEPENDENCE

In the second class of models considered here the structural function delivering the value of an outcome Y_1 involves one or two latent random variables.

$$Y_1 = h(Y_2, X, U_1, U_2)$$

The function h is restricted to be monotonic in U_1 and is normalized increasing. There are covariates, X, whose appearance in h will be subject to exclusion-type restrictions and the variable Y_2 is an observed outcome which may be jointly dependent with U_1 and thus "endogenous."

In this model the latent variable U_2 is the *sole* source of stochastic variation in Y_2 given X and there is the equation

$$Y_2 = g(X, U_2)$$

with g strictly monotonic in U_2, normalized increasing. If U_2 is specified as uniformly distributed on [0, 1] independent of X then $g(X, U_2) = Q_{Y_2|X}(U_2|X)$. In all cases of econometric interest the latent variables are continuously distributed and the strict monotonicity restriction on g means that Y_2 is continuously distributed given X which is not required by the model of Section 2. However, in this model Y_1 can have a discrete distribution which is not permitted in the model of Section 2.

For global identification of h the latent variables (U_1, U_2) and X are restricted to be independently distributed. Note this is more general than the pair of marginal independence conditions: $U_1 \perp X$ and $U_2 \perp X$.[2] This can be substantially weakened to a local quantile independence condition with $Q_{U_1|U_2X}(\tau_1|\tau_2, x) = a(\tau_1, \tau_2)$ a function constant with respect to variation in x for some τ_1 and τ_2.[3] Then there can be identification of h when $U_1 = a(\tau_1, \tau_2)$ and $U_2 = \tau_2$. This is similar to the model studied by Imbens and Newey (2003) and Chesher (2003).

3.1 Continuous Y_2

An argument leading to identification in the continuous Y_2 case is as follows. Holding $U_2 = \tau_2$ and setting U_1 equal to its τ_1-quantile given $U_2 = \tau_2$, there is, in view of the monotonicity restriction on h and the equivariance property of quantiles, the following conditional quantile of Y_1 given U_2 and X.

$$Q_{Y_1|U_2X}(\tau_1|\tau_2, x) = h(g(x, \tau_2), x, a(\tau_1, \tau_2), \tau_2)$$

Because of the strict monotonicity restriction on g, which ensures a one-to-one correspondence between U_2 and Y_2 given X, there is the conditional quantile of Y_1 given Y_2 and X:

$$Q_{Y_1|Y_2X}(\tau_1|Q_{Y_2|X}(\tau_2|x), x) = h(y_2, x, a(\tau_1, \tau_2), \tau_2) \tag{1.4}$$

where on the right-hand side: $y_2 \equiv Q_{Y_2|X}(\tau_2|x)$. This identifies the *value* of the structural function at the indicated arguments of h. Note that this step involving a change in conditioning (from U_2 to Y_2) could not be taken if Y_2 were discrete, a case returned to shortly.

Now suppose that h is insensitive through its X-argument to a movement in X from x' to x'', which is in the nature of an exclusion restriction. Define $y_2' \equiv Q_{Y_2|X}(\tau_2|x')$ and $y_2'' \equiv Q_{Y_2|X}(\tau_2|x'')$. Then for $x^* \in \{x', x''\}$

$$Q_{Y_1|Y_2X}(\tau_1|Q_{Y_2|X}(\tau_2|x'), x') = h(y_2', x^*, a(\tau_1, \tau_2), \tau_2)$$
$$Q_{Y_1|Y_2X}(\tau_1|Q_{Y_2|X}(\tau_2|x''), x'') = h(y_2'', x^*, a(\tau_1, \tau_2), \tau_2)$$

and the difference in the iterated conditional quantiles

$$Q_{Y_1|Y_2X}(\tau_1|Q_{Y_2|X}(\tau_2|x'), x') - Q_{Y_1|Y_2X}(\tau_1|Q_{Y_2|X}(\tau_2|x''), x'') \tag{1.5}$$

[2] The notation $A \perp B$ indicates that the random variables A and B are statistically independent.
[3] There is $Q_{U_1|X}(\tau_2|x) = \tau_2$ by virtue of the definition of U_2.

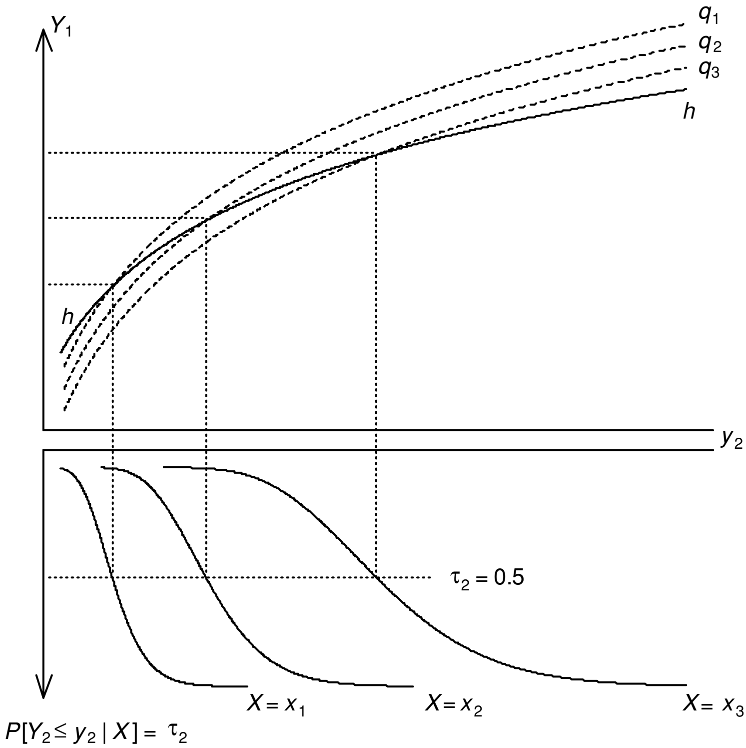

Figure 1.2. Identification of partial differences. Upper graph – q_1, q_2, q_3 are conditional τ_1-quantiles of Y_1 given Y_2 and X at $x \in \{x_1, x_2, x_3\}$ and hh is the structural function $h(y_2, x, a(\tau_1, \tau_2), \tau_2)$. Lower graph – conditional quantiles of Y_2 given $X = x$ for $x \in \{x_1, x_2, x_3\}$ with conditional medians displayed.

identifies the partial difference of the structural function

$$h(y_2', x^*, a(\tau_1, \tau_2), \tau_2) - h(y_2'', x^*, a(\tau_1, \tau_2), \tau_2) \tag{1.6}$$

and so the *ceteris paribus* effect on h of a change in Y_2 from y_2' to y_2''.

Under the iterated covariation condition considered here there is identification of values delivered by the structural function however limited the support of X. However, the locations at which identification can be achieved is critically dependent on the support of X and on the dependence of the distribution of Y_2 on X.[4] When this dependence is weak it may only be possible to identify the structural function over a very limited range of values of Y_2. In this case parametric restrictions have substantial identifying power allowing extrapolation away from the locations at which nonparametric identification can be achieved.

[4] Identification of partial differences using discrete instruments is discussed in Chesher (2007).

Figure 1.2 illustrates the situation. In the upper and lower parts of Figure 1.2 values of Y_2 are measured along the horizontal axes. The vertical axis in the upper part of Figure 1.2 measures values of Y_1. In the lower part of Figure 1.2 the vertical axis measures values of the distribution function of Y_2 given X.

In the upper part of Figure 1.2 the three dashed curves marked q_1, q_2, q_3 are conditional τ_1-quantiles of Y_1 given $Y_2 = y_2$ and $X = x$ drawn for some value of τ_1 as functions of y_2 for $x \in \{x_1, x_2, x_3\}$. In the upper part of the picture the solid curve (marked hh) is the structural function $h(y_2, x, U_1, U_2)$ drawn as a function of y_2 with $U_1 = a(\tau_1, \tau_2)$ and $U_2 = \tau_2$ for the same value τ_1 that determines the dashed conditional quantile functions and for $\tau_2 = 0.5$. There is an exclusion-type restriction so that the structural function remains fixed as x varies across $\{x_1, x_2, x_3\}$.

The three curves in the lower part of Figure 1.2 are conditional distribution (viewed another way, quantile) functions of Y_2 given X for $x \in \{x_1, x_2, x_3\}$ – the *same* values that generate the conditional quantile functions of Y_1 shown in the upper part of the picture. The median ($\tau_2 = 0.5$) values of Y_2 given $X \in \{x_1, x_2, x_3\}$ are marked off on the horizontal axes and, in consequence of equation (1.4), it is at these values that the iterated conditional quantile functions of Y_1 given Y_2 and X intersect with the structural function in the upper part of Figure 1.2 which as noted has been drawn for the case $\tau_2 = 0.5$. Differences of values of Y_1 at these intersections, as given in equation (1.5), identify partial differences of the structural function as given in equation (1.6). The number of feasible comparisons is clearly constrained by the support of X.

When X varies continuously there is the possibility of identification of derivatives of the structural function. In (1.5) and (1.6) divide by $\Delta x \equiv x' - x''$ (X is now assumed scalar), consider $\Delta x \to 0$ and suppose the required derivatives exist. From (1.5) there is, noting that x affects $Q_{Y_1|Y_2X}$ through its X argument and via $Q_{Y_2|X}$ though its Y_2 argument:

$$\begin{aligned}&\nabla_{y_2} Q_{Y_1|Y_2X}(\tau_1|y_2, x)|_{y_2=Q_{Y_2|X}(\tau_2|x)} \nabla_x Q_{Y_2|X}(\tau_2|x)\\ &\quad + \nabla_x Q_{Y_1|Y_2X}(\tau_1|y_2, x)|_{y_2=Q_{Y_2|X}(\tau_2|x)}\end{aligned}$$

and from (1.6) there is, on using $y_2 \equiv Q_{Y_2|X}(\tau_2|x)$ and assuming that h does not vary with x:

$$\nabla_{y_2} h(y_2, x^*, a(\tau_1, \tau_2), \tau_2)|_{y_2=Q_{Y_2|X}(\tau_2|x)} \nabla_x Q_{Y_2|X}(\tau_2|x)$$

and so the following correspondence identifying the y_2-derivative of h.

$$\begin{aligned}&\nabla_{y_2} h(y_2, x^*, a(\tau_1, \tau_2), \tau_2)|_{y_2=Q_{Y_2|X}(\tau_2|x)}\\ &\quad = \nabla_{y_2} Q_{Y_1|Y_2X}(\tau_1|y_2, x)|_{y_2=Q_{Y_2|X}(\tau_2|x)}\\ &\qquad + \frac{\nabla_x Q_{Y_1|Y_2X}(\tau_1|y_2, x)|_{y_2=Q_{Y_2|X}(\tau_2|x)}}{\nabla_x Q_{Y_2|X}(\tau_2|x)}\end{aligned}$$

Imbens and Newey (2003) propose a two-step estimation procedure for estimating h by nonparametric quantile regression of Y_1 on X and a first step estimate of the conditional distribution function of Y_2 given X, the latter serving as a control variate. Ma and Koenker (2006) and Lee (2004) consider control variate estimation procedures under parametric and semiparametric restrictions. Chesher (2003) suggests estimating derivatives and differences of h by plugging unrestricted estimates of the quantile regressions of Y_1 on Y_2 and X and of Y_2 on X into expressions like that given above. Ma and Koenker (2006) study a weighted average version of this estimator in a semiparametric model.

3.2 Discrete Y_2

The joint independence model is useful when X is discrete because, unlike the marginal independence model, it can identify the *ceteris paribus* effect of continuous Y_2 on h. This advantage apparently vanishes once Y_2 is discrete.

The difficulty is that when Y_2 is discrete the conditional independence model does not point identify the structural function without further restriction. This is so however rich the support of discrete Y_2. The reason is that holding discrete Y_2 at a particular quantile does *not* hold U_2 at a quantile of its distribution. In Figure 1.2 the quantile functions in the lower graph are step functions when Y_2 is discrete and so there is a *range* of values of U_2 associated with any particular conditional quantile of Y_2. With U_2 uniformly distributed on (0, 1), when $Y_2 = y_2^m$ and $X = x$ the value of U_2 must lie in the interval $(F_{Y_2|X}(y_2^{m-1}|x), F_{Y_2|X}(y_2^m|x)]$ which, note, may be narrow when Y_2 has many points of support.

If the dependence of U_1 on U_2 can be restricted then this interval restriction on U_2 can imply an interval restriction on the quantiles of the distribution of Y_1 given Y_2 and X and interval identification of the structural function.

The effect of restricting the conditional quantile function of U_1 given U_2 to be a monotonic function of U_2 is studied in Chesher (2005a). The case in which U_2 does not appear in h in its own right is considered here. If there exist values of X, $\tilde{x}_m \equiv \{x_{m-1}, x_m\}$ such that

$$F_{Y_2|X}[y_2^m|x_m] \leq \tau_2 \leq F_{Y_2|X}[y_2^{m-1}|x_{m-1}] \tag{1.7}$$

then $Q_{Y_1|Y_2X}(\tau_1|y_2^m, x_m)$ and $Q_{Y_1|Y_2X}(\tau_1|y_2^m, x_{m-1})$ bound the value of $h(y_2^m, x, a(\tau_1, \tau_2))$ where $x \in \tilde{x}_m$. It is necessary that h is insensitive to changes in x within $\tilde{x}_m$. Specifically, with:

$$q^L(y_2^m, \tilde{x}_m) \equiv \min\left(Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^{m-1}), Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^m)\right)$$
$$q^U(y_2^m, \tilde{x}_m) \equiv \max\left(Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^{m-1}), Q_{Y_1|Y_2X}(\tau_1|y_2^m, x^m)\right)$$

there is the bound:

$$q^L(y_2^m, \tilde{x}_m) \leq h_1(y_2^m, x, a(\tau_1, \tau_2)) \leq q^U(y_2^m, \tilde{x}_m).$$

A bound on the *ceteris paribus* effect of *changing* Y_2 from y_2^m to y^n is available if for the *same* value τ_2 there exist $\tilde{x}_n \equiv \{x_{n-1}, x_n\}$ such that

$$F_{Y_2|X}[y_2^n|x_n] \leq \tau_2 \leq F_{Y_2|X}[y_2^{n-1}|x_{n-1}] \tag{1.8}$$