Conditional Moment Restrictions and Triangular Simultaneous Equations*

Jinyong Hahn[†] and Geert Ridder[‡]

October, 2009

Abstract

It is shown that in a nonparametric nonseparable triangular system the conditional moment restriction (CMR) does not identify the average structural function (ASF). The CMR identifies the ASF only if the model is structurally separable in observable covariates and unobservable random errors. This excludes for instance random coefficient models in which the CMR in general does not identify the average response. An implication of our results is that empirical researchers should use other methods than CMR if they want to estimate the average response in models that are not additively separable.

EconLit code: C140, C200

1 Introduction

Often we are interested in the relation between a (vector of) dependent variable(s) Y and a (vector of) independent variable(s) X. Because the list of independent variables is incomplete, the relation involves one or more unobservable "errors " that are treated as random variables. In the absence of prior knowledge of the nature of the relation, it is reasonable to adopt a nonparametric framework, and try to identify and estimate such a relationship. The literature has used two approaches within the nonparametric framework. In the first approach, the nonparametric relationship has an additively separable error that satisfies some conditional

*We are grateful for comments by the editor and two anonymous referees. Financial support for this research was generously provided through NSF SES 0819612 and 0819638.

[†]Department of Economics, UCLA

[‡]Department of Economics, University of Southern California

moment restriction (see, e.g. Newey and Powell (2003), Hall and Horowitz (2005), and Carrasco, Florens, and Renault (2006)). In the second approach, it has an error that is not separable, but satisfies some independence restriction (see, e.g. Imbens and Newey (2009)). It is of interest to clarify the relationship between the two nonparametric approaches.

Common sense suggests that there could be a relationship between the two approaches. In order to understand why, consider the simple case where X is exogenous. If we adopt the nonseparable error approach, we would write

$$Y = f(X, \varepsilon), \tag{1}$$

and assume that X and ε are independent of each other. As argued by Blundell and Powell (2004) and others, a main function/parameter of interest is the average structural function (ASF) $\phi(x)$ defined by

$$\phi(x) = \mathbb{E}[f(x, \varepsilon)], \tag{2}$$

where the expectation is over the marginal distribution of ε. The ASF is the average outcome if the value $X = x$ is assigned independently of the unobservable ε. Therefore we can think of the ASF as the average causal relation between Y and X. In model (1) the ASF can be obtained using the following conditional moment restriction (CMR)

$$\mathbb{E}[Y - \phi(x)|X = x] = 0. \tag{3}$$

This follows because by independence of X and ε we have[1]

$$\mathbb{E}[Y|X = x] = \mathbb{E}[f(X, \varepsilon)|X = x] = \mathbb{E}[f(x, \varepsilon)] = \phi(x). \tag{4}$$

The main result of this paper is that such a relationship does *not* exist when X is endogenous. If we adopt the nonseparable error approach, the relation in which X is endogenous can be represented by a triangular system

$$Y = f\left(X, \varepsilon\right) \tag{5}$$

$$X = g\left(Z, V\right) \tag{6}$$

with ε and V correlated and Z an instrumental variable. The first equation is the structural equation and the second the first stage equation that relates the endogenous variable to the instrument(s) Z. Although the first stage equation can be a structural equation, in most applications it is a reduced form equation, i.e. it can be considered as the solution of a possibly large system that involves other endogenous variables.[2]

Newey and Powell (2003) show (see also Darolles, Florens, and Renault (2003)) that under some regularity conditions (that are satisfied in the examples we consider) there exists a (unique) function ψ that satisfies the CMR

$$\mathbb{E}[Y - \psi(X) | Z = z] = 0,$$

and by analogy with the case with exogenous X we might expect that ψ and the ASF are related. However we show that the function ψ identified by the CMR is different from the ASF.[3]

We also show that the CMR approach is sensitive to the first stage, because if the correlation between the first stage error V and the structural error changes, then the function ψ identified by the CMR changes as well. Sensitivity to the first stage implies that the identified relation is not *structural* if we adopt the definition by Haavelmo (1944) (see chapter II, section 8).[4][5] Equation (5) is structural if it is invariant to changes in (6) that can be considered as a reduced-form equation that relates the endogenous variable X to the exogenous variable Z.[6] The ASF is clearly a structural object in this sense. This definition of a structural relation is also consistent with the notion that a structural relation is determined by economic principles as optimization and (the tendency to) equilibrium. However, we should emphasize that there are parameters that are of interest and that are sensitive to the first stage. In the sequel we consider an average derivative in Section 5 and a local causal effect in the spirit of LATE in Section 4. Such parameters have drawn a lot of attention in the profession. We show that the CMR cannot identify even these nonstructural parameters without further assumptions.

This paper is about the *interpretation* of the identified function. As such, we simply assume that the sufficient conditions for identification are satisfied for both approaches. The focus on interpretation means that we do not provide an in depth discussion of identification or estimation, since these topics are discussed in many existing papers. However our results have implications for empirical research. Although the CMR does not recover the ASF, using the first stage error as a control variate in the structural relation does identify the ASF if the instrument induces sufficient variation in X or at least the ASF for a subpopulation that depends on the first stage if the instrument does not induce sufficient variation in X. Although we mainly consider nonparametric relations, our conclusions also hold in a linear regression model if we allow for population heterogeneity in responses, i.e. if we consider a random coefficient model. In a random coefficient model the CMR does not identify the average response, but the first stage error used as a control variate does.

2 2SLS in a Random Coefficient Model

Although we are mainly concerned with the nonparametric system (5) and (6), we first consider the estimation of a linear model with random coefficients by 2SLS. The fact that 2SLS does not typically recover the average treatment response is now well-established in the literature, see Imbens and Angrist (1994) or Angrist, Graddy, and Imbens (2000). Our contribution is to go beyond the interpretation of 2SLS as a weighted average treatment response, as was done in the previous literature, and recognize that the discrepancy between 2SLS and the average treatment response can be attributed to the use of conditional moment restrictions in general. This is done by developing a simple mathematical example, which will be used throughout the paper.

In order to appreciate the importance of average treatment response in empirical practice, consider identification of an Engel curve, where Y is the budget share of a certain good[7] and X is the (log) of total expenditure. In applications total expenditure is possibly endogenous. Estimation proceeds by some instrumental variable method, either parametric, semi-parametric, or even fully non-parametric.[8] In most cases, models for Engel curves assume fixed parameters, but because the Engel curves refer to heterogeneous households, it is likely that these households have different responses to a change in total expenditure. This can be captured by a random coefficient model. The treatment effect literature has made researchers aware of the importance of allowing for variation in individual responses. As in that literature the econometrician is interested in the average response to some exogenous change in total expenditure. If the independent variables are exogenous least squares will recover this average response.

Our main results are best understood in a simple example. Consider a triangular system of equations where

$$\begin{aligned} Y &= X^2 \varepsilon \\ X &= Z + V \qquad Z \perp V \end{aligned} \tag{7}$$

We assume that $\varepsilon \sim N(0,1)$, and $V \sim N(0,1)$. We also assume that the first stage error V is correlated with ε through

$$\varepsilon = \frac{1}{2}V + \frac{\sqrt{3}}{2}V^* \qquad V^* \perp V, Z$$

and $V^* \sim N(0,1)$. We will also impose a technical assumption that the support of Z is an open set containing a non-empty interval, which will be satisfied if Z has a normal distribution.

The function $f(X,\varepsilon) = X^2\varepsilon$ is obviously not additively separable in X and ε. *If* X were randomly assigned independent of ε, the level of Y given $X = x$ is equal to $\phi(x) = \mathbb{E}[x^2\varepsilon] =$

$x^2\mathbb{E}[\varepsilon] = 0$, where the expectation is taken with respect to the marginal distribution of ε treating x as nonstochastic. In other words, the average treatment response is summarized by the function $\phi(x) = 0$. On the other hand, it is shown in Section 3 that there is a unique function ψ that satisfies $\mathbb{E}[Y - \psi(X)|Z = z] = 0$, and this function is $\psi(x) = x$.

What is the implication? To adapt our example to the Engel curve case, we add an intercept and write the structural equation as $Y = s^* + X^2\varepsilon$, maintaining the first stage $X = Z + V$. The ε is now the random coefficient that captures the heterogeneous response. It has mean 0. The ASF $\phi(x)$ is now $\phi(x) = s^*$. The joint distribution of (Y, X, Z) is such that $\mathbb{E}[Y - s^* - X|Z = z] = 0$. Therefore, if an econometrician estimates a linear model

$$Y = \alpha + \beta X + \gamma X^2 + u \tag{8}$$

using e.g. 2SLS, as is often done in empirical practice, the econometrician identifies the linear regression parameters $\beta = 1, \gamma = 0$. Obviously the causal parameters should be $\beta = 0, \gamma = 0$, because the ASF $\phi(x) = s^*$ is constant. In the context of Engel curve estimation, even though the good in question has a constant budget share, the econometrician would conclude that the good is a luxury good because the budget share increases with total expenditure.

We note that the problem here is not due to an incorrect statistical specification. The pseudo-parameters $\beta = 1, \gamma = 0$ are such that the linear specification is 'correct' in the statistical sense that, with $\beta = 1, \gamma = 0$, the u is mean independent of Z. The problem lies in the statistical model with the conditional moment restriction

$$\mathbb{E}[Y - \alpha - \beta X - \gamma X^2|Z] = 0,$$

In other words, instrumental variable methods as 2SLS that are based on the conditional moment restriction, identify a pseudo-parameter that does not correspond to the average response in general.

Does the preceding phenomenon disappear if we avoid parametric models? Next we show that a similar observation can be made in the nonparametric context.

3 Does The Conditional Moment Restriction Identify a Structural Parameter?

Our main results are illustrated through the example (7) in the previous section. For technical reasons, we will now explicitly assume that Z has a normal distribution. This ensures that the support condition in Imbens and Newey (2009) is satisfied, so the identification of the ASF does not present any conceptual challenge. It also ensures that the Newey and Powell (2003) completeness assumption is satisfied.

In example (7), $f(X,\varepsilon) = X^2\varepsilon$ is not additively separable in X and ε. Because the ASF is defined with respect to the marginal distribution of ε, we can see that the ASF $\phi(x)$ is

$$\phi(x) \equiv 0,$$

where the expectation is taken with respect to the marginal distribution of ε treating x as nonstochastic.

We now consider the conditional moment restriction

$$\mathbb{E}[Y - \psi\left(X\right)|Z = z] \tag{9}$$

Is there such a $\psi\left(X\right)$? If so, is it unique? The answer to both of these questions is affirmative because, under normality of Z, the Newey and Powell (2003) completeness assumption is satisfied. Because

$$\begin{aligned}\mathbb{E}[Y|Z] &= \mathbb{E}\left[(Z+V)^2\left(\frac{1}{2}V + \frac{\sqrt{3}}{2}V^*\right)\middle| Z\right] \\ &= Z^2\mathbb{E}\left[\frac{1}{2}V + \frac{\sqrt{3}}{2}V^*\middle| Z\right] + Z\mathbb{E}\left[V^2 + \sqrt{3}VV^*\middle| Z\right] \\ &+ \mathbb{E}\left[\frac{1}{2}V^3 + \frac{\sqrt{3}}{2}V^2V^*\middle| Z\right] = Z\end{aligned}$$

and

$$\mathbb{E}[X|Z] = \mathbb{E}\left[Z+V\middle| Z\right] = Z$$

we have

$$\mathbb{E}[Y - X|Z] = 0$$

from which we find that $\psi(x) = x$. Therefore, we conclude that (9) does not identify the ASF $\phi(x) \equiv 0$.

The function ψ identified by the CMR is not invariant to the 'first stage' and therefore not structural in the definition of Haavelmo (1944). Consider a slightly different relationship between the first stage error V and the structural error ε given by[9]

$$\varepsilon = \frac{1}{4}V + \frac{\sqrt{15}}{4}V^* \qquad V^* \perp V, Z$$

We can see that

$$\begin{aligned}\mathbb{E}[Y|Z] &= \mathbb{E}\left[(Z+V)^2\left(\frac{1}{4}V + \frac{\sqrt{15}}{4}V^*\right)\middle| Z\right] \\ &= Z^2\mathbb{E}\left[\frac{1}{4}V + \frac{\sqrt{15}}{4}V^*\middle| Z\right] + Z\mathbb{E}\left[\frac{1}{2}V^2 + \frac{\sqrt{15}}{2}VV^*\middle| Z\right] \\ &+ \mathbb{E}\left[\frac{1}{4}V^3 + \frac{\sqrt{15}}{4}V^2V^*\middle| Z\right] = \frac{Z}{2}\end{aligned}$$

but we continue to have $\mathbb{E}[X|Z] = Z$. Therefore, we have

$$\mathbb{E}\left[Y - \frac{1}{2}X\middle| Z\right] = 0,$$

and the CMR (9) now identifies the function $\psi(x) = \frac{1}{2}x$. The CMR continues to identify a function different from the ASF. Moreover, the function changes as a result of the change in the correlation between the first stage and structural errors. It should be stressed that our discussion does not rely on a failure of the completeness assumption. The result suggests problems with the interpretation of the CMR, and it has nothing to do with the identifiability of the function ψ using the CMR.

Our explanation of the result is very simple. We note that the CMR is inherently defined by the joint density of (ε, V) whereas the ASF is defined by the marginal density of ε. This fundamental tension in their definitions naturally leads to the discrepancy. More specifically, we have

$$\mathbb{E}[\psi(X)|Z=z] = \mathbb{E}[Y|Z=z] = \mathbb{E}[f(g(z,V),\varepsilon)] = \tag{10}$$
$$\int\int f(g(z,v),\varepsilon)p(\varepsilon,v)\mathrm{d}v\mathrm{d}\varepsilon = \int\int \frac{p(\varepsilon,v)}{p_V(v)p_\varepsilon(\varepsilon)} f(g(z,v),\varepsilon)p_V(v)p_\varepsilon(\varepsilon)\mathrm{d}v\mathrm{d}\varepsilon$$

The conditional expectation of the ASF in (2) given $Z = z$ is

$$\mathbb{E}[\phi(X)|Z=z] = \mathbb{E}\left[\mathbb{E}[f(g(Z,V),\varepsilon)]\right] = \int\int f(g(z,v),\varepsilon)p_V(v)p_\varepsilon(\varepsilon)\mathrm{d}v\mathrm{d}\varepsilon \tag{11}$$

We conclude that $\mathbb{E}[\psi(X)|Z=z] \neq \mathbb{E}[\phi(X)|Z=z]$, because the former expectation is over the joint distribution of (the dependent) ε, V while the latter is over the marginal distributions of ε and V. In both conditional expectations setting $Z = z$ does not fix X so that we at most identify some average causal object. The problem with the CMR (9) is that if Z is fixed, there is still variation in X due to variation in V that is correlated with variation in ε. The ASF as a proper causal response function is for the case that X varies independently of ε as in (11). Note that in (10) X and ε can be made to vary independently, but that is for a function that is not the structural equation.

4 Structural Separability

We argue that 'additive separability' can be obtained by mathematical manipulation with vacuous economic content, and some care is needed to interpret an additively separable model. Consider the example discussed in Section 3. If we "define" our error to be $e \equiv X^2\varepsilon - X$, we can easily obtain an additive separable model

$$Y = X + e,$$

where e satisfies the conditional moment restriction $\mathbb{E}[e|Z = z] = 0$. Despite the superficial familiarity, the function $\psi(x) = x$ thus identified is not the ASF.

If the relation $Y = f(X, \varepsilon)$ satisfies the restriction

$$Y = k(X, \varepsilon_1) + \varepsilon_2 \qquad X, Z \perp \varepsilon_1 \quad Z \perp \varepsilon_2$$

with ASF $\mathbb{E}_{\varepsilon_1}[k(x, \varepsilon_1)]$, then the CMR identifies the ASF, because $\varepsilon_1 \perp X|Z$ so that

$$\mathbb{E}[Y|Z = z] = \mathbb{E}[k(X, \varepsilon_1)|Z = z] + \mathbb{E}[\varepsilon_2|Z = z] = \mathbb{E}[\mathbb{E}_{\varepsilon_1}[k(X, \varepsilon_1)]|Z = z].$$

We call such a model for Y *structurally separable*, because the error that makes X endogenous is separable. The structurally separable triangular system can be expressed as

$$Y = m(X) + \eta \tag{12}$$

$$X = g(Z, V) \tag{13}$$

with $m(x) = \mathbb{E}_{\varepsilon_1}[k(x, \varepsilon_1)]$ and $\eta = k(X, \varepsilon_1) - \mathbb{E}_{\varepsilon_1}[k(X, \varepsilon_1)] + \varepsilon_2$ where $\mathbb{E}[\eta|Z] = 0$ because $V \perp \varepsilon_1$.

Imbens and Newey (2009) propose an approach to the identification and estimation of the nonseparable triangular simultaneous equation model in (5) and (6) that also applies if (5) is not structurally separable. Their method recovers the function f in the case that the random error ε is a scalar random variable. The key insight is that because Z and V are independent, the distribution of X given $V = v$ is the same as that of $g(Z, v)$. Because ε, V are independent of Z, we have that

$$X \perp \varepsilon | V = v \tag{14}$$

The variable V is called the control variate. The control variate approach seems to use the representation of the first-stage relation in (6). However, this equation does not impose any restriction on the joint distribution of X, Z.

The results in Blundell and Powell (2004) and Imbens and Newey (2009) imply that the control variate V allows us to obtain the ASF by a conditional moment restriction, even if the model is not structurally separable. To see this we observe that

$$\mathbb{E}[Y|X=x,V=v]=\mathbb{E}[f(X,\varepsilon)|X=x,V=v]=\mathbb{E}[f(x,\varepsilon)|V=v]$$

To obtain the ASF ϕ we integrate this expression with respect to the marginal distribution of V that is the uniform distribution on $[0,1]$. Hence ϕ satisfies the average conditional moment restriction (ACMR)

$$\int_0^1 \mathbb{E}[Y-\phi(X)|X=x,V=v]\mathrm{d}v \tag{15}$$

It is important to note that this works even if ε is a vector.

The ACMR requires that for $X=x$ the control variate V has a distribution with support $[0,1]$. Because $V=F(X|Z)$ with F the conditional cdf of X given Z, this implies that for all x in the support of X the random variable $V(x)=F(x|Z)$ where Z has the conditional distribution of Z given $X=x$ has support $[0,1]$. This holds under the assumptions made above. In general, the full support will only hold on a subset of the support of X. In that case the ASF is can be recovered for a set of values of x that depends on the first stage (6) so that the ASF is not structural in the definition of Haavelmo (1944). It is still an average causal effect. In the case that the support of $V(x)$ is $[v_l(x),v_h(x)]$ we can define an average causal effect of a change from x_1 to x_2 as longs as $[v_l(x_1),v_h(x_1)]$ and $[v_l(x_2),v_h(x_2)]$ overlap. If we denote the intersection of these two sets by $[v_l(x_1,x_2),v_u(x_1,x_2)]$ then the local average causal effect of a change from x_1 to x_2 is

$$\int_{v_l(x_1,x_2)}^{v_u(x_1,x_2)} \left(\mathbb{E}[Y|X=x_2,V=v]-\mathbb{E}[Y|X=x_1,V=v]\right)\mathrm{d}v$$

which is a causal, but not a structural parameter.

5 Does the Conditional Moment Restriction Identify the Average Derivative?

We now ask if the CMR at least identifies the average derivative that is considered by Imbens and Newey (2009), among others. Given a triangular system $Y=f\left(X,\varepsilon\right)$, $X=g\left(Z,V\right)$, the average derivative is defined to be

$$\mathbb{E}\left[\frac{\partial f\left(X,\varepsilon\right)}{\partial x}\right]=\int\int\frac{\partial f(x,\varepsilon)}{\partial x}p(x,\varepsilon)\mathrm{d}x\mathrm{d}\varepsilon,$$

which corresponds to the effect of increasing X by a unit, holding fixed the relationship between X and ε, i.e. it is an average partial effect where the average is over joint distribution of X, ε in which X and ε are correlated. This joint distribution depends on the first stage (6). Therefore the average derivative is not a structural parameter in the definition of Haavelmo (1944) and in general is not equal to the average of the derivative of the ASF. The question is whether the ψ identified by the CMR $\mathbb{E}[Y - \psi(X) | Z = z]$ is such that $\mathbb{E}[\psi'(X)] = \mathbb{E}[\partial f(X, \varepsilon) / \partial x]$, in which case we can argue that the CMR does identify something interesting, i.e. the average derivative. We show that this is not the case by a counter-example.

Suppose that the first stage is such that the conditional density of X given $Z = z$ is exponential with mean equal to z.[10] We can then write the first stage as

$$X = ZV$$

where V is independent of Z, and has a unit exponential distribution. Now, we let

$$Y = X^2 \varepsilon$$

where

$$\varepsilon = V + V^*$$

and V^* is unit exponential independent of V.

We first show that the ASF is different from the function identified by the CMR. Because V is unit exponential, we have $\mathbb{E}[V] = 1$, $\mathbb{E}[V^2] = 2$, and $\mathbb{E}[V^3] = 6$. Likewise, we have the same result for V^*. It follows that the ASF is equal to $2X^2$. Now, note that $X^2 \varepsilon = (ZV)^2 (V + V^*) = V^3 Z^2 + V^2 Z^2 V^*$, and therefore, $\mathbb{E}[Y | Z] = 8Z^2$. We also note that $\mathbb{E}[X^2 | Z] = 2Z^2$, and hence, we conclude that

$$\mathbb{E}\left[Y - 4X^2 \middle| Z \right] = 0,$$

which identifies $\psi(X) = 4X^2$, which is different from the ASF.

Now, we calculate the average derivative by noting that

$$\begin{aligned}
\mathbb{E}\left[\frac{\partial f(X, \varepsilon)}{\partial x}\right] &= \mathbb{E}[2X\varepsilon] = 2\mathbb{E}[(ZV)(V + V^*)] \\
&= 2\mathbb{E}\left[ZV^2 + ZVV^*\right] = 2\left(\mathbb{E}[Z]\,\mathbb{E}\left[V^2\right] + \mathbb{E}[Z]\,\mathbb{E}[V]\,\mathbb{E}[V^*]\right) \\
&= 2\mathbb{E}[Z](2 + 1 \cdot 1) = 6\mathbb{E}[Z].
\end{aligned}$$

On the other hand, we have

$$\mathbb{E}[\psi'(X)] = \mathbb{E}[8X] = 8\mathbb{E}[ZV] = 8\mathbb{E}[Z]\,\mathbb{E}[V] = 8\mathbb{E}[Z],$$

from which we obtain that

$$\mathbb{E}\left[\frac{\partial f\left(X,\varepsilon\right)}{\partial x}\right]\neq\mathbb{E}\left[\psi'\left(X\right)\right].$$

In other words, the CMR does not identify the average derivative in general.

We do note that there is an interesting exception, in which the CMR does identify the average derivative. Suppose that the first stage is separable in Z and V and can be written $X = g(Z) + V$ with g differentiable and $V \perp Z$. The CMR identifies

$$\begin{aligned}0 &= \mathbb{E}\left[Y-\psi\left(X\right)\right|Z=z] = \mathbb{E}\left[f\left(g(Z)+V,\varepsilon\right)-\psi\left(g(Z)+V\right)\right|Z=z]\\ &= \mathbb{E}\left[f\left(g(z)+V,\varepsilon\right)-\psi\left(g(z)+V\right)\right],\end{aligned}$$

or

$$0=\int\int\left(f\left(g(z)+v,\varepsilon\right)-\psi\left(g(z)+v\right)\right)p\left(v,\varepsilon\right)\mathrm{d}v\mathrm{d}\varepsilon.$$

Now differentiate both sides with respect to z, and we obtain

$$0=\int\int\left(\frac{\partial f\left(g(z)+v,\varepsilon\right)}{\partial x}-\psi'\left(g(z)+v\right)\right)g'(z)p\left(v,\varepsilon\right)\mathrm{d}v\mathrm{d}\varepsilon.$$

Dividing by $g'(z)$ we obtain

$$0=\mathbb{E}\left[\left(\frac{\partial f\left(X,\varepsilon\right)}{\partial x}-\psi'\left(X\right)\right)\right|Z=z\right].$$

Now, using the law of iterated expectations, we obtain

$$0=\mathbb{E}\left[\frac{\partial f\left(X,\varepsilon\right)}{\partial x}-\psi'\left(X\right)\right],$$

or

$$\mathbb{E}\left[\frac{\partial f\left(X,\varepsilon\right)}{\partial x}\right]=\mathbb{E}\left[\psi'\left(X\right)\right].$$

We conclude that the CMR identifies the ASF if the structural equation is structurally separable, and the average derivative if the first stage is structurally separable.

6 Related Discussion

In this section, we will make a small digression, and examine some other objects considered in the literature.

Local IV In an exactly identified linear simultaneous equations model, the IV estimator can be written as a ratio of two least squares estimators. Perhaps motivated by such an indirect least squares interpretation, the local IV (see, e.g. Florens, Heckman, Meghir, and Vytlacil (2008) and Schennach, White, and Chalak (2007)) of the form

$$\lambda(z) = \frac{\partial \mathbb{E}[Y|Z=z]/\partial z}{\partial \mathbb{E}[X|Z=z]/\partial z}$$

has been suggested as a way of summarizing some causal relation between Y and X. Our analysis suggests that $\lambda(z)$ is not a 'structural' object either.[11] In the first example above, we have $\mathbb{E}[Y|Z] = Z$ and $\mathbb{E}[X|Z] = Z$ so that $\lambda(z) = 1$. On the other hand, in the second example with slightly different correlation between the first stage and second stage errors, we have $\mathbb{E}[Y|Z] = \frac{1}{2}Z$ and $\mathbb{E}[X|Z] = Z$ so that $\lambda(z) = \frac{1}{2}$. It follows that the local IV is affected by the first stage, and therefore cannot be a structural object.

With the example considered in Section 5, we can examine whether the local IV identifies the average derivative. Because $\mathbb{E}[Y|Z] = 8Z^2$ and $\mathbb{E}[X|Z] = Z$, we can see that $\lambda(z) = 16z$. It follows that $\mathbb{E}[\lambda(Z)] = 16\mathbb{E}[Z] \neq 6\mathbb{E}[Z] = \mathbb{E}[\partial f(X,\varepsilon)/\partial x]$, so the local IV does not identify the average derivative either in general.

On the other hand, the local IV can be used to identify the average derivative if the first stage is structurally separable in Z and V and can be written $X = g(Z) + V$ with g differentiable and $V \perp Z$. Because

$$\frac{\partial \mathbb{E}[Y|Z=z]}{\partial z} = \int\int \left(\frac{\partial f(g(z)+v,\varepsilon)}{\partial x}\right) g'(z) p(v,\varepsilon)\, \mathrm{d}v\mathrm{d}\varepsilon$$

and

$$\frac{\partial \mathbb{E}[X|Z=z]}{\partial z} = g'(z)$$

we obtain

$$\lambda(z) = \int\int \left(\frac{\partial f(g(z)+v,\varepsilon)}{\partial x}\right) p(v,\varepsilon)\, \mathrm{d}v\mathrm{d}\varepsilon = \mathbb{E}\left[\left.\frac{\partial f(X,\varepsilon)}{\partial x}\right| Z=z\right]$$

It follows that the average derivative can be identified by $\mathbb{E}[\lambda(Z)]$ if the first stage is structurally separable in Z and V.

Conditional Quantile Restriction The conditional quantile restriction (CQR) solves $\Pr[Y \leq \psi(X)|Z=z] = \tau$ for some $\tau \in (0,1)$, whereas the quantile structural function (QSF) ϕ is such that $\Pr[f(x,\varepsilon) \leq \phi(x)] = \tau$. Note that the latter is with respect to the marginal distribution of ε. Chernozhukov, Imbens, and Newey (2007) showed that, if the ε is a *scalar* random variable, then the CQR identifies the QSF, i.e., $\psi = \phi$. It is difficult to give a plausible interpretation to the QSF with two or more dimensional errors, which may be a justification to focus on cases with scalar errors, but the question remains whether the CQR identifies anything structural in the case that ε has two or more components.[12] Our preliminary analysis shows that it does not.

For this purpose, we now consider a model

$$Y = X\varepsilon_1 + \varepsilon_2 \quad \varepsilon_1, \varepsilon_2 \overset{iid}{\sim} N(0,1)$$

so that the median structural function $\phi(x) \equiv 0$. We consider the first stage

$$X = Z + V \qquad Z \perp V, \qquad \varepsilon_2 = tV + \sqrt{1-t^2}V^* \qquad V \perp V^*$$

and $V \sim N(0,1)$, $V^* \sim N(0,1)$ for some t such that $|t| \leq 1$. We assume that ε_1, V and V^* are independent of each other. The $\psi(\cdot)$ identified by the conditional median restriction solves $\Pr\left[Y - \psi(X) \leq 0 \middle| Z = z\right] = \frac{1}{2}$. It can be shown that this can be rewritten as

$$\frac{1}{2} = \mathbb{E}\left[\Phi\left(\frac{-tV + \psi(z+V)}{\sqrt{(z+V)^2 + 1 - t^2}}\right)\right] \quad \forall z \tag{16}$$

where $\Phi(\cdot)$ denotes the CDF of $N(0,1)$.

Now, we consider $t = 0$. This is the case with X independent of $\varepsilon_1, \varepsilon_2$, and we can easily see that $\psi(\cdot)$ should be identically equal to zero if it has to satisfy (16). We now ask whether (16) is satisfied with $\psi(\cdot) = 0$ for other values of t. With $t = \frac{1}{2}$ and $z = 1$, we found (by Monte Carlo with 100,000 runs) that the expectation on the right is equal to 0.5381.

We note that the discrepancy between the CQR and QSF can be potentially used as a basis of a specification test, because there should not be any discrepancy if the second stage error is indeed a scalar.

7 Conclusion

Our main conclusion is that in nonseparable models, the usual CMR based on mean independence of the error from the instrument identifies a structural or causal relation between the dependent variable and an endogenous covariate, only if the model is structurally separable. If the CMR is used in a population where the relation is nonseparable, the average structural function that gives the average response given an exogenously assigned level of $X = x$, is not recovered.

Nonseparable parametric models as random coefficient models have become more popular recently. In these models the CMR identifies the average response only if the covariates and random coefficients are independent. If that is not true, then 2SLS does not recover the average response. The control variate estimator of Blundell and Powell (2004) and Imbens and Newey (2009) does recover the average response in that case.

If the first stage model is structurally separable, then the CMR (and the local IV) identifies the average derivative. The common practice to express the first stage as a separable model has therefore implications for identification.

Notes

[1]See Hoderlein and Mammen (2007) for similar results in conditional quantile restrictions.

[2]We can find this reduced form equation by inverting the identity $V = F(X|Z)$ so that we only need to know the exogenous variables of the system.

[3]Florens, Heckman, Meghir, and Vytlacil (2008) argued that the exclusion restriction along with the conditional moment restriction are not sufficient to identify the average treatment effect. Our result goes further, and establishes that the function identified by the conditional moment restriction is *different* from the average structural function.

[4]Haavelmo (1944) calls this invariance property *autonomy*, and it is clear from his discussion that a structural relation must be autonomous. He also makes the distinction between 'two different kind of variations of economic variables, namely hypothetical variations, and variations which are restricted by a system of simultaneous equations.' This is why in the ASF we fix $X = x$ and average over ε (create hypothetical variation), breaking the relation between X and ε induced by the first stage equation (the restricted variation).

[5]We thank Guido Imbens for guiding us to this original literature on structural relations.

[6]See also Goldberger (1991), p. 344 and p. 346, for a similar definition.

[7]We acknowledge that the analogy is not perfect. A budget share should lie in the $[0, 1]$ interval, but our Y can in principle take any value in the real line. This could be achieved by a monotone transformation that can take any real value.

[8]A recent contribution to this literature is Blundell, Chen, and Kristensen (2007)

[9]Note that the marginal distributions of ε and V are not changed.

[10]By Theorem 2.2 in Newey and Powell (2003), the completeness condition is satisfied if Z has a support that contains an open interval.

[11]Here, a 'structural' object is defined in terms of invariance, as discussed at the end of Section 3.

[12]See Hoderlein and Mammen (2007) for a related discussion when X is exogenous.

References

ANGRIST, J., K. GRADDY, AND G. IMBENS (2000): "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish," *Review of Economic Studies*, 67, 499–528.

BLUNDELL, R., X. CHEN, AND D. KRISTENSEN (2007): "Semi-Nonparametric IV Estimation of Shape Invariant Engel Curves," *Econometrica*, 75, 1613–1670.

BLUNDELL, R., AND J. POWELL (2004): "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, 71, 665–679.

CARRASCO, M., J. P. FLORENS, AND E. RENAULT (2006): "Linear Inverse Problems in Structural Econometrics: Estimation Based on Spectral Decomposition and Regularization," in *Handbook of Econometrics, Vol. 6*, ed. by J. Heckman, and E. Leamer. North Holland.

CHERNOZHUKOV, V., G. W. IMBENS, AND W. K. NEWEY (2007): "Instrumental Variable Estimation of Nonseparable Models," *Journal of Econometrics*, 139, 4–14.

DAROLLES, S., J. FLORENS, AND E. RENAULT (2003): "Non Parametric Instrumental Regression," IDEI Working Paper, n. 228.

FLORENS, J. P., J. J. HECKMAN, C. MEGHIR, AND E. VYTLACIL (2008): "Identification of Treatment Effects Using Control Functions in Models with Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica*, 76, 1191–1206.

GOLDBERGER, A. (1991): *A Course in Econometrics*. Harvard University Press.

HAAVELMO, T. (1944): "The probability approach to econometrics," *Econometrica*, 12(Supplement).

HALL, P., AND J. HOROWITZ (2005): "Nonparametric Methods for Inference in the Presence of Instrumental Variables," *Annals of Statistics*, 33, 2904–2929.

HODERLEIN, S., AND E. MAMMEN (2007): "Identification of Marginal Effects in Nonseparable Models without Monotonicity," *Econometrica*, 75, 1513–1518.

IMBENS, G., AND J. ANGRIST (1994): "Identification and estimation of local average treatment effects," *Econometrica*, 62, 467–475.

Imbens, G., and W. Newey (2009): "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," Working Paper, Department of Economics, Harvard.

Newey, W., and J. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

Schennach, S. M., H. White, and K. Chalak (2007): "Estimating Average Marginal Effects in Nonseparable Structural Systems," Working Paper, Department of Economics, UCSD.