

# Conditional Moment Restrictions and Triangular Simultaneous Equations\*

Jinyong Hahn<sup>†</sup> and Geert Ridder<sup>‡</sup>

December 1, 2008

## Abstract

We examine whether a causal interpretation can be given to the function identified by the conditional moment restriction (CMR). It is shown that in a general nonseparable triangular system the CMR does not identify the average structural function (ASF) nor any other structural object. This implies that the CMR identifies a causal relation 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. Because we discuss identification by CMR in a general nonseparable triangular system, we provide a condition under which this system is nonparametrically just identified from the population distribution of the observables, so that under this condition there is a one-to-one correspondence between the population distribution of the observables and the triangular nonseparable system. An implication of our results is that empirical researchers should use other methods than CMR if they want to estimate the average response in a random coefficient model.

## 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

---

\*Financial support for this research was generously provided through NSF grant SES 0452590.

<sup>†</sup>Department of Economics, UCLA

<sup>‡</sup>Department of Economics, University of Southern California

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

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  $\varepsilon$  are independent of each other. As argued by Blundell and Powell (2004) and others, the 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  $\varepsilon$ . The ASF is the average outcome if the value  $X = x$  is assigned independently of the unobservable  $\varepsilon$ . This indicates that 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<sup>1</sup> because by independence of  $X$  and  $\varepsilon$

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

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

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

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

with  $\varepsilon$  and  $V$  correlated and  $Z$  an instrumental variable. The CMR in this model is

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

---

<sup>1</sup>See ? for similar results in conditional quantile restrictions.

and by analogy with the case with exogenous  $X$  we would expect that  $\psi$  and the ASF are related. However we show that the function  $\psi(\cdot)$  identified by the CMR is different from the ASF<sup>2</sup>. Even stronger we argue that the CMR does *not* identify anything structural. To be specific, we show that the CMR approach is sensitive to the first stage in the sense that, if the correlation between the first stage error  $V$  and the structural error changes, then the function  $\psi(\cdot)$  identified by the CMR changes as well. This sensitivity to the first stage implies that the CMR does not identify a structural parameter.

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 in general identify the ASF. 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 identify the average response.

## 2 Does The Conditional Moment Restriction Identify Anything Structural?

In this section, we argue that the conditional moment restriction does not identify the average structural function. Even stronger, we argue that the function that is identified by the conditional moment restriction is not structural.

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

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

---

<sup>2</sup>? 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.

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  $\varepsilon$  through

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

and  $V^* \sim N(0, 1)$ . Finally, we will assume that  $Z$  has a normal distribution<sup>3</sup>.

The relationship between  $Y$  and  $X$  is obviously nonseparable. Because the ASF is defined with respect to the marginal distribution of  $\varepsilon$ , we can see that the ASF  $\phi(x)$  is

$$\phi(x) \equiv \mathbb{E}[x^2\varepsilon] = x^2\mathbb{E}[\varepsilon] = 0,$$

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

We now consider the conditional moment restriction

$$\mathbb{E}[Y - \psi(X) | Z = z] \tag{7}$$

Is there such a  $\psi(X)$ ? 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

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

and

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

we have

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

from which we find that  $\psi(x) = x$ . Therefore, we conclude that (7) does not identify the ASF  $\phi(x) \equiv 0$ .

The preceding analysis may be subject to the criticism that, even though the CMR may not identify the ASF, it may still identify some ‘structural’ object. In order to address such

---

<sup>3</sup>This ensures that the support condition in Imbens and Newey (2003) 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.

a concern, it is first necessary to define what the term ‘structural’ object means. We loosely define it as something solely based on the ‘second stage’. In other words, we argue that a structural object should not be affected by a change of the ‘first stage’ or its correlation with the ‘first stage’.

Now consider a slightly different relationship<sup>4</sup> between the first stage error  $V$  and the second stage error  $\varepsilon$

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

We can see that

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

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 (7) 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 second stage errors. We may therefore conclude that the CMR does not identify a ‘structural’ function. This should not come as a surprise because in a sense the ASF is the only meaningful CMR-based structural object. After all, if we consider the degenerate first stage where  $X = Z$ , so that there is no endogeneity problem, then the CMR does identify the ASF.

It should be stressed that our discussion here does not rely on a failure of the completeness assumption. The result above suggests interpretational problems with the CMR, and it has nothing to do with the identifiability of the function  $\psi(\cdot)$  using the CMR.

We now provide an explanation of the example by providing an interpretation of the function  $\psi(X)$  identified by the CMR. We do so under the assumption that the joint population distribution of  $Y, X, Z$  is such that there exists a triangular simultaneous equations model  $Y = f(X, \varepsilon)$ ,  $X = g(Z, V)$  with the ‘instrument’  $Z$  jointly independent of  $(\varepsilon, V)$ . We provide a justification of this assumption in Section 5.

---

<sup>4</sup>Note that the marginal distributions of  $\varepsilon$  and  $V$  are not changed.

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

$$\mathbb{E}[\psi(X)|Z = z] = \mathbb{E}[Y|Z = z] = \mathbb{E}[f(g(z, V), \varepsilon)] = \int \int f(g(z, v), \varepsilon)p(\varepsilon, v)dv d\varepsilon \quad (8)$$

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

$$\mathbb{E}[\phi(X)|Z = z] = \mathbb{E}[\mathbb{E}[f(g(Z, V), \varepsilon)]] = \int \int f(g(z, v), \varepsilon)p_V(v)p_\varepsilon(\varepsilon)dv d\varepsilon \quad (9)$$

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)  $\varepsilon, V$  while the latter is over the marginal distributions of  $\varepsilon$  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 (7) is that if  $Z$  is fixed, there is still variation in  $X$  due to variation in  $V$  that is correlated with variation in  $\varepsilon$ . The ASF as a proper causal response function is for the case that  $X$  varies independently of  $V$  as in (9).

### 3 Does the Conditional Moment Restriction At Least Identify the Average Derivative?

Having concluded that the CMR does not identify anything structural, we may want to ask if the CMR at least identifies an interesting reduced form parameter. This is too general a question, so we restrict our attention to a reduced form parameter that was considered by, among others Imbens and Newey (2003). Given a triangular system  $Y = f(X, \varepsilon)$ ,  $X = g(Z, V)$ , the average derivative is defined to be

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

which corresponds to the effect of increasing  $X$  by a unit, holding fixed the relationship between  $X$  and  $\varepsilon$ . Obviously it is defined with respect to the joint distribution of  $(X, \varepsilon)$  that is determined by the first stage, and it is not ‘structural’ according to our loose definition in the previous section. The question is whether the  $\psi(\cdot)$  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. 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  $z^5$ . 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  $X^2\varepsilon = (ZV)^2(V + V^*) = V^3Z^2 + V^2Z^2V^*$ , and therefore,  $\mathbb{E}[Y|Z] = 8Z^2$ . We also note that  $\mathbb{E}[X^2|Z] = 2Z^2$ , and hence, we conclude that

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

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

Now, we calculate the average derivative by noting that

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

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

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

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

---

<sup>5</sup>By Theorem 2.2 in Newey and Powell (2003), the completeness condition is satisfied if  $Z$  has a support that contains an open interval.

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

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

or

$$0 = \int \int (f(g(z) + v, \varepsilon) - \psi(g(z) + v)) p(v, \varepsilon) dv d\varepsilon$$

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

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

Dividing by  $g'(z)$  we obtain

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

Now, using the law of iterated expectations, we obtain

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

or

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

Note that if the first stage relation between  $X$  and  $Z$  is expressed as a projection of  $X$  on  $Z$  we apparently obtain a relation that is separable. However, only if the first stage is structurally separable, in the sense defined below, can we identify the average derivative by the CMR.

## 4 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<sup>6</sup> of the form

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

---

<sup>6</sup>See ?, e.g., for related discussion on local IV.

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. 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 3, 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.

**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)  $\phi$  is such that  $\Pr[f(x, \varepsilon) \leq \phi(x)] = \tau$ . Note that the latter is with respect to the marginal distribution of  $\varepsilon$ . ? showed that, if the  $\varepsilon$  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  $\varepsilon$  has two or more components<sup>7</sup>. Our preliminary analysis shows that it does not.

For this purpose, we now consider a model

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

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

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

and  $V \sim N(0, 1)$ ,  $V^* \sim N(0, 1)$  for some  $t$  such that  $|t| \leq 1$ . We assume that  $\varepsilon_1, V$  and  $V^*$  are independent of each other. The  $\psi(\cdot)$  identified by the conditional median restriction solves  $\Pr[Y - \psi(X) \leq 0|Z = z] = \frac{1}{2}$ . It can be shown that this can be rewritten as

$$\frac{1}{2} = \mathbb{E} \left[ \Phi \left( \frac{-tV + \psi(z + V)}{\sqrt{(z + V)^2 + 1 - t^2}} \right) \right] \quad \forall z \quad (10)$$

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

---

<sup>7</sup>See ? for related discussion when  $X$  is exogenous.

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 (10). We now ask whether (10) 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.

## 5 Creation of Nonparametric Triangular Simultaneous Equations Model

In Section 2, we made an assumption that the joint population distribution of  $Y, X, Z$  is such that there exists a triangular simultaneous equations model  $Y = f(X, \varepsilon)$ ,  $X = g(Z, V)$  with the ‘instrument’  $Z$  independent of both  $(\varepsilon, V)$ . We now provide a justification by showing that, under some reasonable assumption, it is always possible to represent the population distribution of  $Y, X, Z$  by such a triangular system.

We first begin by the construction of the first stage. The most general model for the relation between a dependent variable  $X$  and a vector of independent variables  $Z$  has  $g(Z, V)$  monotone in  $V$  for all values of  $Z$ . If we assume that  $V$  and  $Z$  are independent and we normalize the distribution of  $V$  as uniform on  $[0, 1]$ , then as shown by Matzkin (2003)

$$g(z, V) = G^{-1}(V|z) \tag{11}$$

with  $G(x|z)$  the conditional CDF of  $X$  given  $Z = z$ . The model (11) can be ‘constructed’ from the joint distribution of  $X, Z$  by defining the random variable  $V$  as

$$V = G(X|Z) \tag{12}$$

Because

$$\Pr(V \leq e|Z = z) = \Pr(G(X|Z) \leq e|Z = z) = \Pr(X \leq G^{-1}(e|z)|Z = z) = e \tag{13}$$

the error  $V$  has a uniform distribution that is independent of  $Z$ . Upon inversion we obtain (11). This construction implies that there is a one-one correspondence between the model (11) with  $V$  independent of  $Z$  and the observed joint distribution of  $X, Z$ . In other words, the model with  $V$  independent of  $Z$  is nonparametrically just identified.

Now, let  $H(y|z, v) = \Pr(Y \leq y|Z = z, V = v)$  be the conditional CDF of  $Y$  given  $(Z, V)$ , and let  $U = H(Y|Z, V)$ . Note that  $U$  and  $V$  have a uniform distribution that is independent of  $Z$ . Hence if we define  $h(z, v, u) = H^{-1}(u|z, v)$ , we have

$$Y = h(Z, V, U) \quad U \perp Z, V \quad (14)$$

In the same way

$$X = g(Z, V) \quad Z \perp V \quad (15)$$

Note that by construction  $g, h$  are increasing in their last argument and that  $Z, U, V$  are mutually independent.

To construct a triangular simultaneous equations model we would like to invert (15) with respect to  $Z$  and express  $Z$  as a function of  $X, V$ . The simplest case is that  $g(z, v)$  is strictly monotonic, without loss of generality strictly increasing, in  $z$  for (almost all)  $v$ . This is equivalent to assuming that the joint distribution of  $X, Z$  is such that if  $z > z'$  then  $G(x|z) < G(x|z')$  for all  $x$  on the union of the supports of these distributions. If  $Z = z$  is the assigned level of  $Z$ , then  $X(z) = g(z, V)$  is the resulting level of  $X$  and the assumption is equivalent to the assumption that this level is strictly increasing in  $z$  for all members of the population. If  $g(z, v)$  is strictly increasing in  $z$  for (almost all)  $v$ , then

$$Z = g^{-1}(X, V) \quad (16)$$

where  $g^{-1}$  is the inverse with respect to the first argument. Note that in (16)  $X$  and  $V$  are not independent. Substitution in equation (14) gives

$$Y = h(g^{-1}(X, V), V, U) = f(X, \varepsilon) \quad (17)$$

with  $\varepsilon$  the vector  $(U, V)$ . Hence we have constructed a triangular system (15) and (17) with errors  $\varepsilon, V$  that are independent of  $Z$ . Because monotonicity is equivalent to  $G(x|z)$  being decreasing in  $z$  for all  $x$  we can check whether this assumption holds.

## 6 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 2. 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 does not have any structural interpretation.

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

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

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

$$Y = m(X) + \eta \tag{18}$$

$$X = g(Z, V) \tag{19}$$

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 (2003) 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  $\varepsilon$  is a scalar random variable. The construction in Section 5 gives a model in which  $\varepsilon$  is a vector and in that case we do not recover  $f$ , but the ASF. The key insight is that because  $Z$  and  $V$  are independent, the distribution of  $X$  given  $V = v$  is the same as that of  $g(v, Z)$ . Because  $\varepsilon, V$  are independent of  $Z$ , we have that

$$X \perp \varepsilon|V = v \tag{20}$$

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 (2003) 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  $\phi$  we integrate this expression with respect to the marginal distribution of  $V$  that is the uniform distribution on  $[0, 1]$ . Hence  $\phi$  satisfies the average conditional moment

restriction (ACMR)

$$\int_0^1 \mathbb{E}[Y - \phi(X)|X = x, V = v]dv \tag{21}$$

It is important to note that this works even if  $\varepsilon$  is a vector.

If the model is structurally separable we identify the ASF by the CMR using e.g. Newey and Powell (2003). This ASF can be compared to the ASF identified with the control variate approach of Imbens and Newey (2003), i.e. we compare  $\mathbb{E}_{\varepsilon_1}[k(x, \varepsilon_1)]$  and  $\mathbb{E}[f(x, \varepsilon_1, \varepsilon_2)]$ . The ASF can differ because  $f$  is not separable in  $\varepsilon_2$  and/or  $X$  and  $\varepsilon_1$  are correlated.

## 7 Implications for parametric models

The discussion in the previous section may appear to be a subtle problem and a consequence of our desire to avoid parametric models. We argue that it in fact suggests a problem with 2SLS for the linear model, which is commonly used in empirical application, if we allow for heterogeneous response. Random coefficient models are a natural choice if we are interested in the causal effect of a change in a continuous variable on an outcome. The treatment effect literature has made researchers aware of the importance of allowing for variation in individual responses. If the independent variables are exogenous least squares will recover the average response. The results in the previous section show that that is not true if the independent variables are correlated with the random coefficients. Only if the endogeneity is due to dependence of a separable error and the independent variables and in addition the random coefficients are independent of the covariates does 2SLS recover the average response<sup>8</sup>.

As an example we consider identification of an Engel curve, where  $Y$  is the budget share of a certain good 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<sup>9</sup>. 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. In that case the econometrician is interested in the average response to some exogenous change in total expenditure and this is the average structural function. If total expenditure were exogeneous, the (linear) regression coefficients

---

<sup>8</sup>If the independent variable is chosen by the agent, this amounts to the assumption that the agent does not act on his/her own value of the coefficient, but all agents choose according to a common value, e.g. the population average.

<sup>9</sup>A recent contribution to this literature is Blundell, Chen, and Kristensen (2007)

would be equal to this average response.

We reconsider the first example, and make small changes to the specification and notation. We write the second stage as  $Y = s^* + X^2\varepsilon$  maintaining the first stage specification. The  $\varepsilon$  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

$$Y = \alpha + \beta X + u \tag{22}$$

using e.g. 2SLS, as is often done in empirical practice, the econometrician identifies linear regression parameter  $\beta = 1$ . Obviously the causal parameter should be 0, because the ASF  $\phi(x) = s^*$  is constant. Therefore although 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-parameter  $\beta = 1$  is such that the linear specification is ‘correct’ in the statistical sense that, with  $\beta = 1$ , the  $u$  is mean independent of  $Z$ . The problem lies in the statistical model with the conditional moment restriction

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

which does not identify a causal but a reduced form parameter. In other words, instrumental variable methods as 2SLS that are based on the conditional moment restriction, identify a pseudo-parameter that has no clear interpretation (and certainly not a causal interpretation) in general.

## 8 Conclusion

Our main conclusion is that in nonseparable models, the usual CMR based on mean independence of the error from the instrument identifies a 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, nor is any other structural object.

We also provided a condition under which the nonseparable model is nonparametrically just identified from the population distribution, so that if that condition holds, the nonseparable

model does not impose any restriction on the population distribution and hence is the natural alternative to the separable model.

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. We suggest a control variate estimator that does recover the average response in that case.

## References

- 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.
- HALL, P., AND J. HOROWITZ (2005): “Nonparametric Methods for Inference in the Presence of Instrumental Variables,” *Annals of Statistics*, 33, 2904–2929.
- IMBENS, G., AND W. NEWEY (2003): “Identification and Estimation of Triangular Simultaneous Equations Models without Additivity,” Working Paper, Department of Economics, Berkeley.
- MATZKIN, R. (2003): “Nonparametric Estimation of Nonadditive Random Functions,” *Econometrica*, 71, 1339–1375.
- NEWEY, W., AND J. POWELL (2003): “Instrumental Variable Estimation of Nonparametric Models,” *Econometrica*, 71, 1565–1578.