Conditional Moment Restrictions and Triangular Simultaneous Equations

By

Jinyong Hahn and Geert Ridder

April 2007

IEPR WORKING PAPER 07.3

INSTITUTE OF ECONOMIC POLICY RESEARCH

UNIVERSITY OF SOUTHERN CALIFORNIA

http://www.usc.edu/iepr

Conditional Moment Restrictions and Triangular Simultaneous Equations*

Jinyong Hahn† and Geert Ridder‡

April 3, 2007

Abstract

We show that in the nonparametric triangular simultaneous equations model, the mean independence conditional moment restriction (CMR) identifies a causal relation between the dependent variable and an endogenous covariate, only if the model is structurally separable in observable covariates and unobservable random errors. If the CMR is used in a nonseparable population model, the average structural function (ASF) is not recovered. However, the ASF is recovered from the average CMR (ACMR) based on independence of the endogenous covariate and the random error given a control variate. We also provide a condition under which the nonseparable model is nonparametrically just identified from the population distribution of the observables, so that under this assumption the nonseparable triangular simultaneous equations model does not restrict their population distribution.

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

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

†Department of Economics, UCLA

‡Department of Economics, University of Southern California

unobservable “errors ” that are treated as random variables, because their realized value is not determined by the observed independent variables. In the absence of prior knowledge of the nature of the relation, we can ask what features of the relation can be identified from the joint distribution of Y, X.

In the sequel Y is a continuous dependent variable. The most general model for the relation between such a dependent variable and a vector of independent variables is

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

with f monotone in ε for all values of X. If we assume that ε and X are independent and we normalize because f is unspecified the distribution of ε as uniform on $[0, 1]$, then as shown by Matzkin (2003)

$$f(x, \varepsilon) = F^{-1}(\varepsilon|x) \tag{2}$$

with $F(y|x)$ the conditional cdf of Y given $X = x$.

The model (1) can be ‘constructed’ from the joint distribution of Y, X by defining the random variable ε as

$$\varepsilon = F(Y|X) \tag{3}$$

Because

$$\Pr(\varepsilon \le e|X = x) = \Pr(F(Y|X) \le e|X = x) = \Pr(Y \le F^{-1}(e|x)|X = x) = e \tag{4}$$

the error ε has a uniform distribution that is independent of X. Upon inversion we obtain (1). This construction implies that there is a one-one correspondence between the model (1) with ε independent of X and the observed joint distribution of Y, X. In other words, the model with ε independent of X is nonparametrically just identified.

As argued by Blundell and Powell (2004) and others, the main function of interest is often the Average Structural Function (ASF) $\phi(x)$ defined by

$$\phi(x) = \mathbb{E}[f(x, \varepsilon)] \tag{5}$$

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. This indicates that we can thinks 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)

$$\mathbb{E}[Y - \phi(x)|X = x] = 0 \tag{6}$$

This follows because by independence of X and ε

$$\mathbb{E}[Y|X=x] = \mathbb{E}[f(X,\varepsilon)|X=x] = \mathbb{E}[f(x,\varepsilon)] = \phi(x) \tag{7}$$

A special case of model (1) is the separable nonparametric regression model

$$Y = m(X) + \varepsilon \tag{8}$$

In this model we need not assume independence of X and ε. Mean independence $\mathbb{E}[\varepsilon|X=x] = 0$ suffices. In this model the ASF is $m(x)$, and we obtain the ASF from the CMR

$$\mathbb{E}[Y - m(x)|X=x] = 0 \tag{9}$$

Now assume that we use this CMR in the case that the model is nonseparable. As noted the alternative is the nonseparable model (1) with X and ε independent, a model that is always consistent with the population joint distribution of Y, X. By (7), we conclude that even if the model is nonseparable the CMR in (6) still recovers the ASF. We conclude that although the CMR in (6) makes a weaker mean independence assumption, the stronger independence assumption holds as well and this ensures that we always recover the ASF, even if the model is misspecified as additively separable.

The contribution of this paper is to show that the a similar result holds if the independent variables X are possibly correlated with the unobservable(s) ε and we use an instrumental variable Z to recover the relation between Y and X. The analysis is more complicated in this case. In particular, we must be careful in the choice of the CMR. We show that the obvious CMR does not recover the ASF, but an integrated CMR that is based on a control variate does. Our results clarify the relation between two strands of the nonparametric instrumental variable (IV) literature: that based on inversion of linear integral equations (e.g. Newey and Powell (2003) and Hall and Horowitz (2005)) and that based on control variables (e.g. Imbens and Newey (2003) and Blundell and Powell (2004)). We show that the latter is more 'robust' against misspecification, because the former requires that the relation is structurally separable in the observable independent variables and the unobservable random error.

2 Creation of Nonparametric Triangular Simultaneous Equations Model

Often the observable X and unobservable ε are correlated. The model that has received most attention is the separable model with endogenous covariates

$$Y = m(X) + \varepsilon \tag{10}$$

with (components of) X and ε dependent. Newey and Powell (2003) and Hall and Horowitz (2005)) discuss identification and estimation under the assumption that ε is mean independent of Z. Under this assumption

$$\mathbb{E}[Y|Z = z] = \mathbb{E}[m(X)|Z = z] = \int_{\mathbb{X}} m(x)p(x|z)\mathrm{dz} \tag{11}$$

and m is the solution to a linear integral equation. Newey and Powell (2003) discuss completeness assumptions under which this equation has a unique solution. If the completeness assumption holds mean independence is sufficient for the identification of the ASF m.

The relation between Y and X need not be separable in the observable X and the unobservable ε. A general model in this case is the triangular nonparametric simultaneous equations model

$$Y = f(X, \varepsilon) \tag{12}$$
$$X = g(Z, V) \tag{13}$$

Model (12) is the same as model (1), but here we do not assume that X and ε are independent. Instead we have an instrumental variable Z that is related to the endogenous variable X in (13). We assume that Z is independent of both V and ε. This is a stronger assumption than mean independence of ε and Z[1]. However by

$$Y = \mathbb{E}[f(X, \varepsilon)|Z = z] + e$$

with

$$e = (f(X, \varepsilon) - \mathbb{E}[f(X, \varepsilon)|Z = z])$$

we obtain a model with a 'separable' error e that is mean independent of Z. The question then becomes what $\mathbb{E}[f(X, \varepsilon)|Z = z]$ tells us about the

[1] We argue below that the independence of V and Z is a free assumption.

causal relation between Y and X. Because any joint distribution of Y, X corresponds to exactly one model (1) with independent (of X) error, the function $\tilde{f}$ or the ASF $\mathbb{E}[\tilde{f}(x, \tilde{\varepsilon})]$ of that model will differ from the function f or the ASF $\mathbb{E}[f(x, \varepsilon)]$ that is identified from the joint distribution of Y, X, Z. The difference between $\tilde{f}$ and f and the corresponding ASF is the endogeneity bias due to the correlation between the error and the independent variables.

We ask two questions: (i) if the model is not separable in X and ε, is the nonseparable triangular simultaneous equations model in (12) and (13) a general alternative? (ii) if we estimate a model under the mean independence assumption, while in the population the relation is not separable, what do we learn about the ASF of the relation between Y and X?

The answer to (i) is that, if X and Z are both scalar, then under a monotonicity assumption on the relation between X and Z the system (12) and (13) with Z independent of ε, V is nonparametrically just identified from the joint distribution of Y, X, Z. It is easy to see that this is true if we start with the linear relations (in variables in deviation from their mean)

$$Y = \gamma_1 Z + \eta \tag{14}$$

$$X = \gamma_2 Z + V \tag{15}$$

If the joint distributions of Y, Z and X, Z are such that the conditional means of Y and X given Z are linear in Z, then both η and V are mean independent of Z. If $\gamma_2 \neq 0$ which is a requirement for Z to be a valid instrument, then we can solve for Z from (15) and substitute the solution in (14) to obtain the relation

$$Y = \beta X + \varepsilon \tag{16}$$

with $\beta = \frac{\gamma_1}{\gamma_2}$ and $\varepsilon = \eta - \frac{\gamma_1}{\gamma_2} V$. Obviously X is endogenous in (16), and Z is a valid instrument for X. Hence β satisfies the CMR

$$\mathbb{E}[Y - \beta X | Z = z] = 0 \tag{17}$$

Although βx is the ASF, it may or may not have a causal interpretation. If the population is such that (14) and (15) hold, then if we believe that there is a causal relation between Y and X it is given by (16) and the corresponding ASF is βx. In this case the validity of the instrument Z follows from the assumption of linear conditional means.

Can we 'construct' model (12) and (13) if the conditional means are not linear in Z? If $H(y|z) = \Pr(Y \leq y | Z = z)$ is the conditional cdf of Y given

Z and $G(x|z) = \Pr(X \leq x|Z = z)$ the conditional cdf of X given Z, then as in (4) if we define

$$U = H(Y|Z) \qquad V = G(X|Z)$$

U and V have a uniform distribution that is independent of Z. Hence if we define $h(z, u) = H^{-1}(u|z)$, we have

$$Y = h(Z, U) \qquad Z \perp U \tag{18}$$

In the same way

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

Note that by construction g, h are increasing in their second argument.

To construct a triangular simultaneous equations model we would like to invert (19) 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

$$Z = g^{-1}(X, V) \tag{20}$$

where g^{-1} is the inverse with respect to the first argument. Note that in (20) X and V are not independent. Substitution in equation (18) gives

$$Y = h(g^{-1}(X, V), U) = f(X, \varepsilon) \tag{21}$$

with ε the vector U, V. Hence we have constructed a triangular system (19) and (21) with errors ε, V that are independent of Z. Because monotonicity is equivalent to $G(x|z)$ being increasing in z for all x we can check whether this assumption holds.

We conclude that under monotonicity the nonseparable triangular simultaneous equations model is the natural alternative to the separable model

$$Y = m(X) + \varepsilon \tag{22}$$
$$X = g(Z, V) \tag{23}$$

Under monotonicity the separable model is obtained if $h(g^{-1}(X,V),U) = m(X) + k(U,V)$ and we define $\varepsilon \equiv k(U,V)$. Note that in this case Z is independent of ε, V by construction. If the model has an additive error, we call the model *structurally separable*. If monotonicity does not hold, then we can still consider the nonseparable model as the natural alternative to the separable model. However, in that case the nonseparable model imposes restrictions on the joint distribution of Y, X, Z, be it weaker assumptions than the separable model.

If the model is structurally separable, then Z and ε are independent. The nonparametric IV estimators of Newey and Powell (2003) and Hall and Horowitz (2005)) make the weaker assumption that ε is mean independent of Z and, as we have seen above, mean independence holds even if the model is not structurally separable. If both the mean independence and completeness assumptions hold and the model is structurally separable, then mean independence and full independence both are sufficient for the identification of the ASF m. However, if the population model is not structurally separable, then, as we will show in the next section, the CMR corresponding to mean independence does not recover the ASF while the average CMR on the assumption of full independence still identifies the ASF of the nonseparable model. Hence whether we assume full or mean independence has implications for the identification of the ASF by a conditional moment restriction.

These are counter-intuitive results given our experience with simpler models. Consider, for example, a parametric[2] model of the form $Y = q(X,\theta) + \varepsilon$. Consider two possible assumptions on ε: (i) $\mathbb{E}[\varepsilon|Z] = 0$; and (ii) $\varepsilon \perp Z$ and $\mathbb{E}[\varepsilon] = 0$. These two assumptions do not contradict each other. If there are two estimators that fully utilize (i) and (ii), respectively, they converge to the same probability limit. The only difference between the two estimators is that the estimator that fully utilizes the independence assumption is in general more efficient than the estimator that only uses the mean independence restriction. The question of identification may seem trivial even in the nonparametric context. If the model is additively separable as in (22), then the same kind of analysis as in parametric model can be conducted. In other words, we are tempted to say that the same m is identified whether ε has a conditional zero mean given Z or is independent of Z.

[2]This model is semiparametric, to be precise, because we do not impose any parametric assumption on the distribution of ε.

3 What Does the Conditional Moment Restriction Identify?

In this section we compare the conditional moment restriction for mean independence and the average conditional moment restriction for full independence. We first show that the CMR for mean independence of ε and Z does not identify a causal parameter, i.e. the ASF. Hence if we define ψ as the solution to

$$\mathbb{E}[Y - \psi(X)|Z = z] \tag{24}$$

then in general $\psi \neq \phi$ with ϕ the ASF. We begin with an example to illustrate this point.

Example: Let

$$Y = X^2\varepsilon \quad \varepsilon \sim N(0, 1)$$

so that the ASF $\phi(x) \equiv 0$. In addition

$$X = Z + V \qquad Z \perp V$$

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

and $V \sim N(0, 1/2)$, $V^* \sim N(0, 1/2)$[3]. Under these assumptions $\mathbb{E}[Y|Z] = Z$ and $\mathbb{E}[X|Z] = Z$, so that $\psi(x) = x \neq \phi(x)$. If we "define" our error to be $e \equiv X^2\varepsilon - X$, we can easily see that the conditional moment restriction is satisfied, and we can conclude that (24) does not identify the ASF. Note that although we can write the model as $Y = X + e$ with $\mathbb{E}[e|Z = z] = 0$, the model is not structurally separable.

In order to understand this example, we consider the question what is identified by the CMR (24). In section 2 we showed that if the monotonicity assumption holds, we can represent the joint population distribution of Y, X, Z by a triangular simultaneous equations model (12) and (13) with the 'instrument' Z independent of both ε and V. Hence we have

$$\mathbb{E}[Y|Z = z] = \mathbb{E}[\psi(X)|Z = z] = \mathbb{E}[f(g(z, V), \varepsilon)] = \int\int f(g(z, v), \varepsilon)p(\varepsilon, v)\mathrm{d}v\mathrm{d}\varepsilon \tag{25}$$

[3] Under these assumptions Newey and Powell (2003) completeness assumption is satisfied and the example does not rely on a failure of that assumption.

The conditional expectation of the ASF in (5) given $Z = z$ is

$$\mathbb{E}[\phi(X)|Z = z] = \mathbb{E}\left[\mathbb{E}[f(g(Z, V), \varepsilon)]\right] = \int\int f(g(z, v), \varepsilon)p_V(v)p_\varepsilon(\varepsilon)\mathrm{d}v\mathrm{d}\varepsilon \tag{26}$$

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. The problem with the CMR (24) 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 is assigned independently of V as in (26).

Imbens and Newey (2003) propose an approach to the identification and estimation of the triangular simultaneous equation model in (12) and (13) that also applies if (12) is not structurally separable. Their method recovers the function f in the case that the random error ε is a scalar random variable. The construction in section 2 gives model in which ε 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 ε, V are independent of Z, we have that

$$X \perp \varepsilon | V = v \tag{27}$$

The variable V is called the control variate. The control variate approach seems to use the representation of the first-stage relation in (13). 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 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)

$$\int_0^1 \mathbb{E}[Y - \phi(X)|X = x, V = v]\mathrm{d}v \tag{28}$$

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

Example, continued: In the example we have if we define $\tilde{V} = G(X|Z)$ that $X = Z + \frac{1}{\sqrt{2}}\Phi^{-1}(\tilde{V}) = Z + V$, so that we can either use $\tilde{V}$ or V as the control variate. Because ε, V are jointly independent of Z, $\mathbb{E}[\varepsilon|X = x, V = v] = \mathbb{E}[\varepsilon|Z = x - v, V = v] = \mathbb{E}[\varepsilon|V = v]$. Hence $\mathbb{E}[Y|X = x, V = v] = x^2\mathbb{E}[\varepsilon|V = v] = x^2 v$ so that averaging over the marginal $N(0, 1/2)$ distribution of V gives 0 for all x which is the ASF.

If the model is structurally separable, we can use the ACMR to recover m. This is an alternative to the inversion of the integral equation in (11). The control variate method adds the control variate V to the set of regressors in the nonparametric regression of Y on X. The nonparametric regression is then averaged over the control variate.

If the model is not structurally separable, then the CMR based on mean independence does not identify the ASF. However if the monotonicity assumption holds, then the ACMR still recovers the ASF. Hence the ACMR is not only simpler than the inversion estimator, it is also more robust against misspecification of the model, because it will recover the ASF if the population relation between Y and X is not separable.

4 Conclusion

Our main conclusion is that in the nonparametric triangular simultaneous equations model, 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 in observable covariates and unobservable random error. 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. However, the ASF is recovered from the average CMR (ACMR) based on independence of the endogenous covariate and the random error given the control variate. The ACMR also applies if the model is structurally separable, so that in all cases the ACMR is preferred over the CMR derived from mean independence, if we are interested in recovering the causal relation between the dependent and endogenous independent variable. It is a bonus that estimators based on the ACMR avoid the ill-conditioned inverse that needs to be computed for the mean independence CMR.

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 restrictions on the population distribution and hence is the natural alternative to the separable model. We are currently working an relaxing the monotonicity assumption, so that the nonseparable model could be considered as the natural alternative in a larger class of population distributions.

References

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.