IDENTIFICATION IN GENERAL TRIANGULAR SYSTEMS

Maximilian Kasy*
May 28, 2012

This paper discusses identification in continuous triangular systems without restrictions on heterogeneity or functional form. In particular, we do not assume separability of structural functions, restrictions on the dimensionality of unobservables, or monotonicity in unobservables. We do maintain monotonicity of the first stage relationship in the instrument. We show that under this condition alone, and given rich enough support of the data, we can achieve point identification of potential outcome distributions, and in particular of the average structural function. If the support of the continuous instrument is not large enough potential outcome distributions are partially identified. If the instrument is discrete identification fails completely.

The setup discussed in this paper covers important cases not covered by existing approaches such as conditional moment restrictions (c.f. Newey and Powell, 2003) and control variables (c.f. Imbens and Newey, 2009). It covers, in particular, random coefficient models, as well as models arising as the reduced form of a system of structural equations.

Keywords: triangular systems, simultaneous equations, instrumental variables, identification.

1. introduction

A large literature in econometrics studies identification of triangular systems of the form

$$Y = g(X, U)$$
$$X = h(Z, V)$$

where it is assumed that $Z \perp (U, V)$. The variable Y is usually called the outcome variable, X is the treatment and Z is an instrumental variable. U and V are unobservable. In this paper, we show point identification of the average structural function and related objects in continuous triangular systems, without restricting functional forms or the dimensionality of unobservables. We do assume monotonicity of the function h in the instrument Z. Point identification further requires that the instrument has rich enough support. Our setup generalizes those considered in the literature on conditional moment restrictions (c.f. Newey and Powell, 2003), which restricts heterogeneity in the structural equation of interest. Our setup generalizes those considered in the literature on control variables (c.f. Imbens and Newey, 2009), which restricts heterogeneity in

Assistant professor, Department of Economics, UCLA, and junior associate faculty, IHS Vienna. Address: 8283 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095. E-Mail: maxkasy@econ.ucla.edu.

*I'd like to thank Gary Chamberlain, Bryan Graham, Jinyong Hahn, Guido Imbens, Zhipeng Liao, Rosa Matzkin, James Powell, Geert Ridder, Alexander Rothenberg and James Stock for valuable discussions and comments.

the first stage relationship. To the best of our knowledge, this paper is the first to show point identification in continuous triangular systems without restrictions on heterogeneity. Our results imply validity of the control function approach proposed by Imbens and Newey (2009) in this context, despite failure of conditional independence. We show, furthermore, that the triangular systems discussed here arise naturally in the context of simultaneous equations with exclusion restrictions, where h describes a "reduced form" relationship.

Identification of triangular systems with unrestricted heterogeneity for binary (discrete) treatment X and instrument Z is well understood, in particular since the contributions of Imbens and Angrist (1994) and Manski (2003). Identification of triangular systems with continuous treatment X is the subject of more recent contributions to the literature.

There are two main approaches in the literature on triangular systems with continuous treatment. One strand of the literature assumes an additively separable structural relationship of the form $Y = g^1(X) + U$. Under this assumption the average structural function g^1 satisfies the conditional moment restriction $E[Y - g^1(X)|Z] = 0$ (c.f. Newey and Powell, 2003; Horowitz, 2011). Hahn and Ridder (2011) show that, absent additivity, the function identified by the conditional moment restriction has no structural interpretation.[1] Chernozhukov et al. (2007) generalize the conditional moment restriction approach to structural equations monotonic in unobservables. Their approach also relies on the assumption of one-dimensional heterogeneity U, however.

Another strand of the literature uses control functions (c.f. Newey et al., 1999; Imbens and Newey, 2009). Imbens and Newey (2009) propose the control function $V^* = F(X|Z)$. They show that conditional independence $X \perp U|V^*$ holds if h is monotonic in V, where V is one-dimensional. Kasy (2011) shows that conditional independence fails in general if V is of higher dimension. It turns out, however, that the Imbens and Newey (2009) approach does identify the average structural function under the assumptions maintained in this paper. This is quite surprising given the failure of conditional independence.

There is, finally, the literature on nonparametric identification in systems of simultaneous equations, see in particular the important contribution by Matzkin (2008). This literature requires that the dimensionality of unobserved heterogeneity is no larger than the number of endogenous variables. The simultane-

[1] For a general triangular system, the function g^1 identified by the conditional moment restriction satisfies

$$E[g^1(X)|Z] = E[Y|Z] = \int g(h(z,v),u)dP(u,v),$$

whereas the average structural function $\overline{g}$ satisfies

$$E[\overline{g}(X)|Z] = \int g(h(z,v),u)dP(u)dP(v).$$

ous systems considered in the present paper, in contrast, allow for unrestricted heterogeneity. Blundell and Matzkin (2010) provide conditions under which simultaneous systems have a triangular reduced form with one-dimensional heterogeneity in the first stage.

The rest of this note is structured as follows. Section 2 introduces the formal setup analysed in this paper, and discusses the assumptions maintained. Section 3 presents our central positive result, characterizing identification under the assumption of monotonicity of h in Z. Section 4 discusses the relationship of this identification result to the control function approach. Section 5 shows that the triangular system we consider arises naturally as reduced form of a system of simultaneous equations. Section 6 shows non-identification for the case of discrete Z. Section 7 interprets partial- and non-identification in our context as arising from support problems. Section 8 concludes. Appendix A sketches a reweighting estimator based on the identification result of section 3. Appendix B provides an illustration in the context of a simple random coefficient model, and appendix C discusses two examples of simultaneous systems which naturally satisfy the assumptions maintained in this paper. Appendix D contains all proofs.

2. SETUP

We consider the following setup.

ASSUMPTION 1 (Triangular System)

$$\begin{aligned} Y &= g(X, U) \\ X &= h(Z, V) \end{aligned} \tag{1}$$

where $X, Y, Z \in \mathbb{R}$, the unobservables U, V have their support in an arbitrary space of unrestricted dimensionality, and

$$Z \perp (U, V). \tag{2}$$

ASSUMPTION 2 (Continuous treatment)
Treatment X is continuously distributed in $\mathbb{R}$ conditional on Z. The structural function g is continuous in x with probability one.

ASSUMPTION 3 (First stage monotonic in instrument)
The first stage relationship h is strictly monotonic in Z.

ASSUMPTION 4 (Continuous instrument)
The instrument Z is continuously distributed in $\mathbb{R}$, with support $[z_0, z_1]$. The first stage relationship h is continuous in z with probability one, and $P(X \leq x | Z = z)$ is continuous in z for all x.

Under these assumptions we can introduce the following potential outcome notation.[2]

DEFINITION 1 (Potential outcomes)
We denote

$$Y^x = g(x, U)$$
$$X^z = h(z, V) \tag{3}$$

We define furthermore

$$Z^x = \begin{cases} h^{-1}(x, V) & \text{if} \quad h(z_0, V) \leq x \leq h(z_1, V) \\ -\infty & \text{if} \quad x < h(z_0, V) \\ \infty & \text{if} \quad h(z_1, V) < x. \end{cases} \tag{4}$$

The notation Y^x and X^z for potential outcomes is standard. The variable Z^x is a slightly different object. It denotes the value of Z which would attain treatment level x given unobserved heterogeneity V. In equation (4), Z^x is well defined under assumption 3; h^{-1} is to be understood as the inverse of h given V. We set Z^x to equal $\pm\infty$ if there is no $z \in [z_0, z_1]$ that achieves $h(z, V) = x$.

The main object of interest in this paper is the average structural function $\overline{g}$ (c.f. Blundell and Powell, 2003), where

$$\overline{g}(x) := E[Y^x] = E[g(x, U)]. \tag{5}$$

Our results discuss, more generally, identification of the (marginal) potential outcome distribution $P(Y^x)$.

REMARK: We have not imposed any of the restrictions on heterogeneity common in the literature. In particular we have not assumed that V can be replaced by a one-dimensional index, as necessary for achieving conditional independence using the control function approach (c.f. Imbens and Newey, 2009; Kasy, 2011). We have not assumed, either, that the structural function of interest is separable in U, as necessary for the "nonparametric instrumental variable" approach (c.f. Newey and Powell, 2003; Horowitz, 2011; Hahn and Ridder, 2011). The assumptions maintained in this paper thus cover, in particular, the case of a generic random coefficient model, which is not covered by either the control function or the conditional moment restriction literature. Appendix B discusses an example of such a random coefficient model.

[2]In this paper, "structural functions" and "potential outcomes" are considered to be essentially equivalent notations for random functions, where the arguments are either written as function arguments or as superscripts. The structural function notation additionally makes explicit the unobserved heterogeneity, which is left implicit in the potential outcome notation.

REMARK: We are imposing one substantive restriction, monotonicity of h in Z. This assumption is potentially testable, since it implies monotonicity of $F(X|Z)$ in Z for all X. I would argue, also, that this assumption is much weaker than monotonicity of h in U, as required for the control function approach if conditional independence is to be achieved. The latter assumption requires heterogeneity to be effectively one-dimensional. The assumption of monotonicity of h in Z, as maintained here, is similar to the "no defiers" assumption in the discrete Imbens and Angrist (1994) setup.

The general triangular system model of assumptions 1, 2 and 4 induces a probability distribution on the space of structural functions mapping Z to X. A realization $h(.,V)$ from this distribution describes the schedule of potential treatment levels X^z for a given unit of observation. Without further restrictions, this is a distribution on the infinite-dimensional space $\mathscr{C}([z_0,z_1])$. The assumption that h is monotonic in V implies that the support of this distribution is restricted to a one-dimensional sub-manifold of $\mathscr{C}([z_0,z_1])$! In contrast, the assumption that h is monotonic in z restricts the support of this distribution to a proper subset of $\mathscr{C}([z_0,z_1])$, but it does not restrict it to lie in a subset of lower dimension.

3. IDENTIFICATION WITH A CONTINUOUS INSTRUMENT

The following theorem presents our central identification result. It shows how to achieve identification of the distribution of the potential outcome Y^x (for a given x) in three steps.

1. The distribution of Y^x given Z^x is identified from the distribution of Y given Z, X.
2. The distribution of Z^x is identified from the conditional distribution of X given Z.
3. Integrating $P(Y^x|Z^x)$ over the distribution of Z^x identifies the unconditional distribution of Y^x.

Point identification requires rich enough support of Z, i.e., the support of Z^x must be a subset of the support of Z.

THEOREM 1 (Identification with a continuous instrument)
Under assumptions 1, 2, 3, and 4

$$P(Y^x \leq y|Z^x = z) = P(Y \leq y|X = x, Z = z) \tag{6}$$

and

$$F_{Z^x}(z) := P(Z^x \leq z) = P(X \geq x|Z = z) = 1 - F_{X|Z}(x|z). \tag{7}$$

Let $p^x := P(Z^x \in [z_0,z_1]) = F_{Z^x}(z_1) - F_{Z^x}(z_0)$. *If* $p^x = 1$, *then*

$$P(Y^x \leq y) = \int_{z_0}^{z_1} P(Y \leq y|X = x, Z = z) dF_{Z^x}(z). \tag{8}$$

If the distribution of Z^x is absolutely continuous relative to the distribution of Z, we can define the weighting function

$$w(x,z) := dF_{Z^x}(z)/dF(z|x). \tag{9}$$

For this definition of w we get

$$\begin{aligned} P(Y^x \leq y) &= E[\mathbf{1}(Y \leq y) \cdot w(X,Z)|X=x], \text{ and} \\ \overline{g}(x) &= E[Y^x] = E[Y \cdot w(X,Z)|X=x]. \end{aligned} \tag{10}$$

If $p^x < 1$ then we can bound the distribution of Y^x by

$$P(Y^x \leq y) \in \int_{z_0}^{z_1} P(Y \leq y|X=x, Z=z) dF_{Z^x}(z) + [0, 1-p^x], \tag{11}$$

where these bounds are sharp.

The proof of this theorem, and of all further results, can be found in appendix D.

REMARK: Theorem 1 shows identification of the marginal distribution of Y^x. It does not show identification of the joint distribution of (Y^{x_1}, Y^{x_2}) for any x_1, x_2. In particular, it does not identify the distribution of the functions $g(., U)$. This is not surprising - even in the case of exogenous X (i.e., $X \perp U$) we can only identify the marginal distributions of Y^x if heterogeneity is left unrestricted.

REMARK: If Z and Z^x are continuously distributed, we can differentiate their distribution functions, so that we can rewrite the weight function w as

$$w(x,z) = -\frac{\frac{\partial}{\partial z} F(x|z)}{\frac{\partial}{\partial z} F(z|x)}. \tag{12}$$

Theorem 1 shows that the weight w effectively "exogenizes" X, i.e., reweighting the distribution of (X,Y) by w implies that for the reweighted distribution $X \perp U$. We can thus run mean regressions or quantile regressions for the reweighted distribution and get unbiased estimates. Appendix A briefly discusses estimation based on this idea.

4. THE RELATIONSHIP TO CONTROL FUNCTIONS

Imbens and Newey (2009) propose to identify the average structural function $\overline{g}(x)$ using a control function V^* which satisfies $X \perp U|V^*$. In particular, they consider

$$\tilde{g}(x) = \int_0^1 m(x,v) dv \tag{13}$$

where

$$m(x,v) := E[Y|X=x,V^*=v] \tag{14}$$

and $V^* = F(X|Z)$. This definition of V^* implies that $V^* \sim U[0,1]$, so that equation (13) integrates over the marginal distribution of V^*. Imbens (2007) noted that conditional independence of X and U given V^* does not hold in the case of a random coefficient first stage. In Kasy (2011) it was shown more generally that conditional independence cannot be achieved if $\dim(V) > 1$.

The following theorem shows that the control function approach is nonetheless justified in the setup considered in the present paper, since it is equivalent to the reweighting given by equation (10) in theorem 1.[3] The support requirement for point identification of $\tilde{g}(x)$ is that V^* has full support given $X = x$. This is equivalent to the requirement that $p^x = 1$ in theorem 1.

THEOREM 2 (Validity of the control function approach)
Under assumptions 1, 2, 3, and 4, if $F(X|Z)$ is differentiable in X and Z, and if $V^* = F(X|Z)$ has full support given $X = x$, then

$$P(Y^x \leq y) = \int_0^1 P(Y \leq y|X=x,V^*=v)dv, \tag{15}$$

even though in general

$$P(Y^x \leq y|V^*=v) \neq P(Y^x \leq y|X=x,V^*=v).$$

In particular, $\tilde{g}(x) = \overline{g}(x)$, where $\tilde{g}(x)$ is defined as in equation (13) *and $\overline{g}(x) = E[Y^x]$.*

REMARK: This result is quite remarkable. The motivation for control functions in Imbens and Newey (2009) is based on conditional independence $X \perp U|V^*$, which implies $P(Y^x \leq y|V^*=v) = P(Y^x \leq y|X=x,V^*=v)$. This in turn implies that equation (15) holds. In general, however, conditional independence cannot be achieved. Theorem 2 implies that we can nonetheless achieve independence of X and Y^x by conditioning on the control function V^* and averaging over its marginal distribution. Equation (15) does hold, even though its usual justification - conditional independence - does not hold. Appendix 4 illustrates this point in the context of a random coefficient model.

5. TRIANGULAR SYSTEMS AS REDUCED FORM OF SIMULTANEOUS EQUATIONS

So far, we have taken the triangular system setup as given. In this section, we show that it arises as the reduced form of a system of structural equations without restrictions on functional form or heterogeneity.

[3] In a working paper version of (Hahn and Ridder, 2011), a related result was shown which implies that the ASF is identified using the control function approach, as long as it is point identified. Their proof proceeds by constructing an observationally equivalent triangular system with one-dimensional first stage heterogeneity, given any continuous distribution of (Y,X,Z).

ASSUMPTION 5 (System of simultaneous equations)

1. *(Observables) We observe the joint distribution of* (X, Y, Z)*, where* $X, Y, Z \in \mathbb{R}$.
2. *(Equilibrium conditions)* X *and* Y *satisfy the system of simultaneous equations*

$$\mathbf{g}(Y, X, Z, U) = 0, \tag{16}$$

   where $\mathbf{g}$ *is an* $\mathbb{R}^2$ *valued function. The unobservable* U *has its support in an arbitrary space of unrestricted dimensionality.*
3. *(Uniqueness) For all* z *and* u*, there is a unique solution* (y, x) *to the equilibrium conditions. For all* x *and* u*, there is a unique solution* (y, z).
4. *(Exclusion restriction)* $\mathbf{g}_1$*, the first component of* $\mathbf{g}$*, is constant in* Z.
5. *(Randomization)* Z *is statistically independent of* U.

REMARK: The setup given by assumption 5 is similar to the one considered by Matzkin (2008), with the important distinction that we allow for heterogeneity U of unrestricted dimensionality. Matzkin (2008) assumes that $\dim(U) = \dim(g)$. We only discuss the case of two endogenous variables, however, whereas Matzkin (2008) considers systems of arbitrary dimension. Appendix C discusses two examples which naturally satisfy assumption 5.

DEFINITION 2 (Potential outcomes and reduced form)
Under assumption 5, define (Y^x, Z^x) *as solution to the system of equations*

$$\mathbf{g}(Y^x, x, Z^x, U) = 0, \tag{17}$$

and (Y^z, X^z) *as solution to the system of equations*

$$\mathbf{g}(Y^z, X^z, z, U) = 0. \tag{18}$$

Define the corresponding functions

$$g(x, U) := Y^x, \tag{19}$$

$$h(z, U) := X^z. \tag{20}$$

REMARK: The random variables (Y^x, Z^x) and (Y^z, X^z) are well defined by the assumption of uniqueness. The exclusion restriction implies that Y^x is pinned down by the first equation, $\mathbf{g}_1(Y^x, x, U) = 0$, so that Y^x is a function of x and U alone. This justifies the definition of g in equation (19). Letting $V := U$, equations (19) and (20) have the same form as the triangular system of assumption 1.

The following theorem implies that the reduced form of the system of simultaneous equations given by assumption 5 indeed satisfies the assumptions of the triangular system we consider in this paper. In particular, the identification result of theorem 1 applies. To guarantee monotonicity of h in z we have to impose some additional restrictions on the signs of the derivatives of $\mathbf{g}$.

THEOREM 3 (Triangular reduced form)
Suppose assumption 5 holds, Z is continuously distributed, and X is continuously distributed given Z. Assume furthermore that $\mathbf{g}$ is differentiable in (x, y, z), and that $\text{sign}(\partial \mathbf{g}_1/\partial y \cdot \partial \mathbf{g}_2/\partial z)$, *as well as* $\text{sign}(\det(\partial \mathbf{g}/\partial(x, y)))$ *are constant.*

Then the reduced form functions g and h of definition 2 satisfy assumptions 1, 2, 3, and 4.

6. DISCRETE INSTRUMENTS

This section considers again the triangular system setup of section 2 with one modification. We drop the assumption of continuity of Z and assume instead that Z has finite support. It turns out that in this setup the average structural function is completely un-identified.

THEOREM 4 (Non-identification with discrete instruments)
Suppose that assumptions 1 and 2 hold, that Z has finite support $z = 1, \ldots, \overline{z}$, and that X is continuously distributed with bounded density conditional on Z.

Then the data are completely uninformative about the distribution of Y^x. That is, for any y the identified set for $P(Y^x \leq y)$, as a function of x, is given by $\{\tilde{g}^y : \tilde{g}^y$ continuous and bounded by $[0, 1]\}$. This also holds if we additionally impose assumption 3.

Similarly, under the additional assumption that $Y \in [0, 1]$, the identified set for the average structural function $\overline{g}$ is given by $\{\overline{g} : \overline{g}$ continuous and bounded by $[0, 1]\}$.

REMARK: This theorem shows that, for the case of continuous treatment and discrete instrument, the data are completely uninformative about the distribution of Y^x, and in particular about the average structural function. There is a "discontinuity" of identifiability, going from large but finite support of Z to a continuous support. This is counter-intuitive and owed to the fact that we have not precluded structural functions with arbitrarily high slopes or curvatures. An a-priori bound on slopes or curvatures would lead to "continuity" of identifiability, but seems hard to justify in most cases. A more attractive alternative might be the imposition of a Bayesian prior which incorporates the assumption that high curvatures are unlikely. This alternative is explored in a working paper by the author of the present paper.

7. SUPPORT PROBLEMS, INTERPOLATION, AND EXTRAPOLATION

Theorem 1 shows that the average structural function at x is partially identified if the support of a continuous instrument Z is a subset of the support of Z^x, so that treatment level x is never observed for some sub-population. Theorem 4 shows that the average structural function is not identified if the support of Z is discrete. In this section we discuss these results and provide a unified perspective on the support problems arising in (i) non-parametric finite sample inference, (ii)

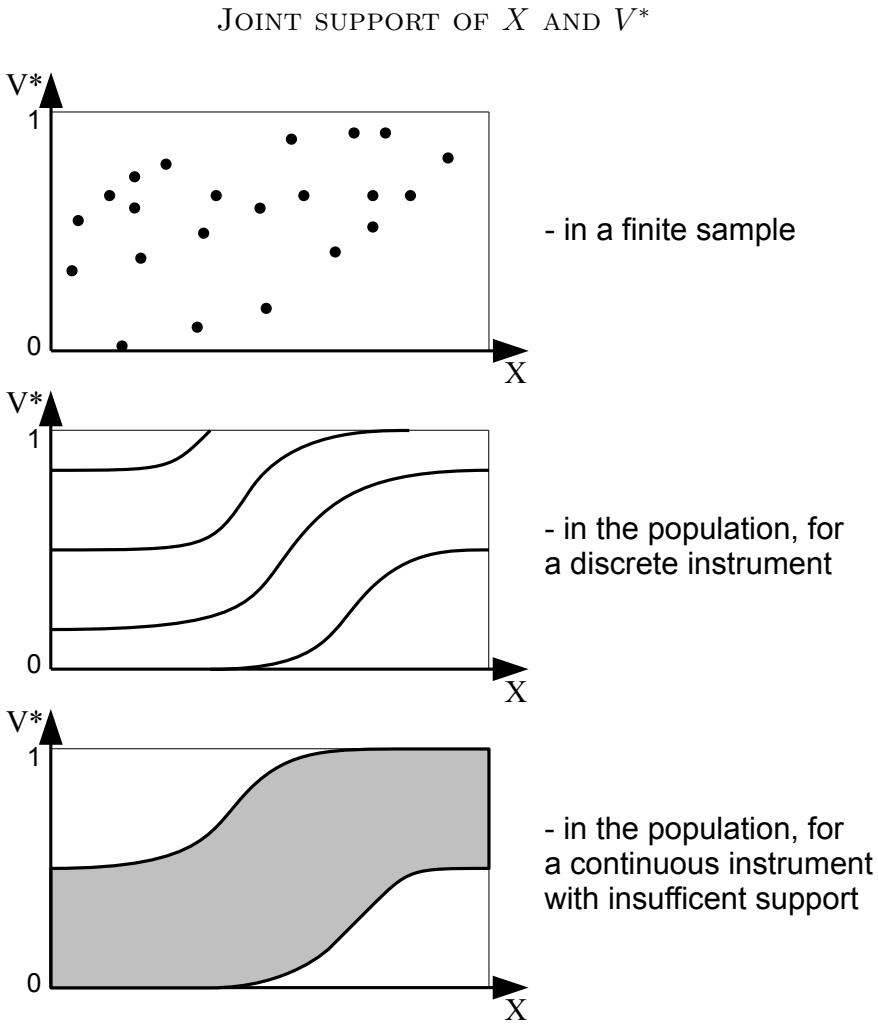

FIGURE 1.— Regressor support and inference, interpolation, and extrapolation

triangular systems with discrete instruments, and (iii) triangular systems with continuous, bounded instruments.

Suppose that assumptions 1, 2, and 3 hold. Suppose furthermore that $h(z, V)$ is well defined for all $z \in \mathbb{R}$; even for values z outside the support of Z. Suppose finally that x is in the range of $h(., V)$ almost surely. Under these assumptions, the potential outcome distributions $P(Y^z, X^z) = P(g(h(z, V), U), h(z, V))$ are well defined for all $z \in \mathbb{R}$, and we can set $P(X, Y|Z = z) = P(Y^z, X^z)$ for all z. We can thus apply the results of sections 3 and 4 to get

$$\overline{g}(x) = E[Y^x] = \int_0^1 m(x, v)dv,$$

where, as before, $m(x, v) := E[Y|X = x, V^* = v]$ and $V^* := F(X|Z)$.

(i) In finite samples, we only observe $Y = m(x, v) + \epsilon$ for a finite number of points (x, v); the first graph in figure 1 illustrates. Here ϵ is a statistical residual satisfying $E[\epsilon|X, V^*] = 0$. Suppose Z is continuously distributed with large enough support, so that V^* has full support given X, and $\overline{g}$ is point-identified. If we want to estimate m and $\overline{g}$, in finite samples we still have to interpolate and extrapolate using some (non-parametric) estimation method.

(ii) If Z has only finite support $z_1 < z_2 < \ldots < z_k$, then V^* has only finite support given X, too:

$$\text{supp}(V^*|X = x) = \{v_1^x > v_2^x > \ldots > v_k^x : \ v_j^x = P(X \leq x|Z = z_j)\}.$$

This is illustrated in the second graph in figure 1. The data identify $m(x, v_j^x)$ for $j = 1 \ldots k$, but they are uninformative about m for other values of v. Knowledge of $m(x, v)$ for a finite number of points v does not constrain $\overline{g}(x) = \int_0^1 m(x, v)dv$ in any way. This implies that the data do not impose any bounds on $\overline{g}(x)$, as shown in theorem 4.

(iii) If we have a continuous instrument Z, identification problems arise if the support of Z^x is a strict superset of the support of Z. This happens if and only if the support of V^* given $X = x$ is a strict subset of $[0, 1]$, $\text{supp}(V^*|X = x) \subsetneq [0, 1]$. In this case, the distribution of Y^x is only identified for the sub-population for which $Z^x \in \text{supp}(Z)$, i.e., $v \in \text{supp}(V^*|X = x)$, and unidentified for the sub-population for which $Z^x \notin \text{supp}(Z)$. The average structural function $E[Y^x]$ is therefore only partially identified, as shown in theorem 1. The third graph in figure 1 illustrates this case.

8. CONCLUSION

This paper discusses identification in continuous triangular systems without any restrictions on heterogeneity or functional form. We do assume monotonicity of the first stage relationship in the instrument. We show that, under this

condition alone, the average structural function (ASF) is identified if the instrument has rich enough support. Furthermore, the control function approach of Imbens and Newey (2009) does identify the ASF despite the fact that conditional independence does not hold if heterogeneity is multidimensional. We show that the ASF is partially identified if the instrument is continuous with insufficient support. The ASF remains completely unidentified if the instrument is discrete. These results immediately apply to simultaneous systems with unrestricted heterogeneity, since such systems give rise to a reduced form satisfying our assumptions.

APPENDIX A: ESTIMATION USING REWEIGHTING

Following theorem 2, we can estimate the average structural function by regressing Y on X and V^*, and averaging predicted values for $X = x$ over the marginal distribution of the control V^*. This is the approach proposed by Imbens and Newey (2009).

The identification result of theorem 1 suggests an alternative estimation approach, based on reweighting. In particular, it was shown in section 3 that, given sufficient support of the instrument, the ASF is identified by

$$\overline{g}(x) = E[Y \cdot w(X, Z)|X = x],$$

where

$$w(x, z) = \frac{f_{z^x}(z)}{f(z|x)} = -\frac{\frac{\partial}{\partial z}F(x|z)}{\frac{\partial}{\partial z}F(z|x)}.$$

Reweighting by w effectively makes X exogenous relative to the distribution of heterogeneity U. An estimator for w can be constructed as follows Use a kernel density estimator for the denominator, i.e.,

$$\widehat{f}(z|x) = \frac{1}{\tau}\frac{\sum_i K\left(\frac{Z_i - z}{\tau}\right) K\left(\frac{X_i - x}{\tau}\right)}{\sum_i K\left(\frac{X_i - x}{\tau}\right)}.$$

Use local linear regression, as well as smoothing of the indicator function, to estimate the numerator, i.e.,

$$\widehat{f}_{z^x}(z) = \underset{b}{\operatorname{argmin}} \ \min_a \sum_i \left(L\left(\frac{x - X_i}{\tau}\right) - a - b \cdot (Z_i - z)\right)^2 \cdot K\left(\frac{Z_i - z}{\tau}\right).$$

In these expressions, K is an appropriate kernel function integrating to one with support $[-1, 1]$, L is the cumulative distribution function of a smooth symmetric distribution with support $[-1, 1]$, and τ is a bandwidth parameter. Form an estimator for w from these two estimates,

$$\widehat{w}(x, z) = -\frac{\widehat{f}_{z^x}(z)}{\widehat{f}(z|x)}.$$

Assuming that the support conditions for point identification are satisfied, an estimator of the average structural function can be formed using weighted least squares with $\widehat{w}(x, z)$ as weights.

Application of this approach has to carefully consider support issues. In particular, if the ASF is only partially identified, there will be values of x, z for which the denominator of w equals 0. Even if the ASF is point identified, identification will in general be "weak" or "irregular", c.f. Khan and Tamer (2010). In either case, consistent estimators will have to apply trimming for observations with large estimated weights $\widehat{w}(x, z)$. As a consequence, the conditional expectation of weights given x will be less than 1, and it might be useful to report bounds, as in the partially identified case in theorem 1.

APPENDIX B: A RANDOM COEFFICIENT EXAMPLE

Theorem 2 showed validity of the Imbens and Newey (2009) control function approach under our assumptions, which cover in particular the case of random coefficient models. This result stands in surprising contrast to the results of Kasy (2011), where it was shown that conditional independence can in general not be achieved if first stage heterogeneity is multidimensional, as for instance in random coefficient models. This section illustrates.

Consider the following model.

$$\begin{aligned} Z &\sim N(0,1) \\ X &= V_1 + V_2 Z \\ V_1|Z &\sim N(0,1) \\ V_2|V_1, Z &\sim 1 + Ber(.5) \\ Y &= \beta_1 + \beta_2 X + U \\ U|V, Z &\sim N(V_1, 1) \end{aligned} \tag{21}$$

This model clearly satisfies assumptions 1 through 4, and $F(X|Z)$ has full support for all X. As a consequence the implications of theorem 1 and 2 hold. However, as shown by the following calculations, conditional independence of X and U given $V^* = F(X|Z)$ does not hold.

To show this, let us first characterize $P(U|X,Z)$, which immediately follows from $P(V_1|X,Z)$. By construction $P(X|V_1,Z) = .5 \cdot \mathbf{1}(X = V_1 + Z \text{ or } X = V_1 + 2Z)$ and $P(V_1|Z) = \phi(V_1)$, so that V^1 has only two points of support given X, Z and

$$P(V_1|X,Z) = \frac{1}{\phi(X-Z) + \phi(X-2Z)} \cdot \phi(V_1) \cdot \mathbf{1}(V_1 = X - Z \text{ or } V_1 = X - 2Z).$$

The distribution $P(U|X,Z)$ is thus given by

$$U|X,Z \sim (X-2Z) + Z \cdot Ber\left(\frac{\phi(X-Z)}{\phi(X-Z)+\phi(X-2Z)}\right) + N(0,1),$$

which clearly is a 2 dimensional family of distributions. (To see this, re-parametrize in terms of $X-2Z$ and $X-Z$.) The control function is given by

$$V^* = F(X|Z) = .5 \cdot (\Phi(X-Z) + \Phi(X-2Z)).$$

It is thus immediate that $E[U|X,V^*]$ is not a function of V^* alone, so that conditional mean independence is violated.

To get an explicit representation of the weight function in this model, note that

$$-\frac{\partial}{\partial z} F(X|Z) = .5 \cdot (\phi(X-Z) + 2\phi(X-2Z)).$$

Furthermore, we have

$$\frac{\partial}{\partial z} F(Z|X) = \phi(Z) \cdot .5(\phi(X-Z) + \phi(X-2Z))/P(X).$$

The weight function of theorem 1, as represented in equation (12), is thus given by

$$w(x,z) = -\frac{\frac{\partial}{\partial z}F(x|z)}{\frac{\partial}{\partial z}F(z|x)} = \frac{P(X)}{\phi(Z)} \cdot \frac{\phi(X-Z)+\phi(X-2Z)}{\phi(X-Z)+2\phi(X-2Z)}.$$

APPENDIX C: EXAMPLES OF SIMULTANEOUS SYSTEMS

The following two examples naturally satisfy the assumptions we have imposed in section 5. In both examples we focus on the case where we have an instrument excluded from one of the structural equations. Our results immediately extend to the case where we have an instrument

for both equations. In that case, both average structural functions are identified.

Example: Supply and Demand
Take $Y = Q$ as quantity and $X = P$ as price in a partial equilibrium market, and let Z be some random factor shifting supply but excluded from demand. Let

$$\mathbf{g}(Q, P, Z, U) = \begin{pmatrix} D(P, U) - Q \\ S(P, U, Z) - Q \end{pmatrix}.$$

In this example, the average structural function $\overline{g}(x) = E[Y^x]$ corresponds to expected demand at a given price p, $E[D(p, U)]$.

Example: Two player complete information Nash equilibrium with real valued actions
Consider a two player Nash game where Y is the action of player 1, X is the action of player 2 and Z is a random factor that enters player 2's utility without affecting player 1 directly. Let

$$\mathbf{g}(Y, X, Z, U) = \begin{pmatrix} BR^1(X, U) - Y \\ BR^2(Y, Z, U) - X \end{pmatrix},$$

where BR^1 and BR^2 are the best response functions of either player. In this example, the average structural function $\overline{g}(x) = E[Y^x]$ corresponds to the average best response of player 1 to a given action x of player 2, $E[BR^1(x, U)]$.

APPENDIX D: PROOFS

Proof of Theorem 1:
Equation (6) holds because

$$\begin{aligned} & P(Y \leq y | X = x, Z = z) \\ &= P(Y^x \leq y | X = x, Z = z) \\ &= P(Y^x \leq y | Z^x = z, Z = z) \\ &= P(Y^x \leq y | Z^x = z). \end{aligned} \tag{22}$$

The first equality follows from the definition of Y^x. The second equality follows because $X = x$ and $Z = z$ holds if and only if $Z^x = z$ and $Z = z$. The last equality follows from exogeneity of Z; $(U, V) \perp Z$ implies $(Y^x, Z^x) \perp Z$.

To see that equation (7) holds, note that the monotonicity assumption 3 implies that $Z^x \leq z$ if and only if $X^z \geq x$. Exogeneity of the instrument implies

$$P(X^z \geq x) = P(X \geq x | Z = z) = 1 - P(X \leq x | Z = z),$$

where the last equality follows from continuity of the distribution of X given Z.

In order to show the reweighting representation of the average structural function given in equation (10), note that

$$\begin{aligned} \overline{g}(x) &= \int E[Y | X = x, Z = z] dF_{Z^x}(z) \\ &= \int E[Y | X = x, Z = z] \frac{dF_{Z^x}(z)}{dF(z|x)} dF(z|x) \\ &= \int E\left[Y \cdot \frac{dF_{Z^x}(z)}{dF(z|x)} \,\middle|\, X = x, Z = z \right] dF(z|x) \\ &= E[Y \cdot w(X, Z) | X = x]. \end{aligned} \tag{23}$$

The reweighting representation for $P(Y^x \leq y)$ follows by the same argument, once we replace Y by $\mathbf{1}(Y \leq y)$.

To derive the bounds on $P(Y^x \leq x)$ for the case that $p^x < 1$, note that

$$P(Y^x \leq x) = P(Y^x \leq x | Z^x \in [z_0, z_1]) \cdot p^x + P(Y^x \leq x | Z^x \notin [z_0, z_1]) \cdot (1 - p^x),$$

where

$$P(Y^x \leq x | Z^x \in [z_0, z_1]) = \int_{z_0}^{z_1} P(Y \leq y | X = x, Z = z) dF_{Z^x}(z).$$

Under our assumptions, the data impose no restrictions on $P(Y^x \leq x | Z^x \notin [z_0, z_1])$, which implies sharpness of these bounds. □

Proof of Theorem 2:
By similar arguments as in the derivation of the reweighting result of theorem 1 (note that $dF(V^*) = 1$),

$$\begin{aligned}
&\int_0^1 P(Y \leq y | X = x, V^* = v) dv \\
&= \int E[\mathbf{1}(Y \leq y) | X = x, V^*] \frac{dF(V^*)}{dF(V^*|x)} dF(V^*|x) \\
&= \int E\left[\mathbf{1}(Y \leq y) \cdot \frac{dF(V^*)}{dF(V^*|x)} \;\middle|\; X = x, V^* = v\right] dF(V^*|x) \\
&= E[\mathbf{1}(Y \leq y) \cdot \tilde{w}(X, Z) | X = x], \qquad (24)
\end{aligned}$$

where

$$\tilde{w}(X, Z) = \frac{dF(V^*)}{dF(V^*|x)} = \frac{1}{f(V^*|x)}.$$

To calculate this weight we need the conditional density of $V^* = F(X|Z)$ given X. Recall that assumption 3 implies that the the mapping from Z to V^* given X is invertible. The conditional density of V^* given X thus equals

$$f(v|x) = f(z|x) / \left|\frac{\partial v}{\partial z}\right| = \frac{dF(z|x)}{-\frac{\partial}{\partial z} F(x|z)}$$

by a change of variables, so that $\widetilde{w}(X, Z) = w(X, Z)$. This implies that the control function approach is equivalent to the reweighting approach in theorem 1.

The result that (in general) $P(Y^x \leq y | V^* = v) \neq P(Y^x \leq y | X = x, V^* = v)$ follows from the arguments in Kasy (2011), and is also confirmed by the example discussed in appendix B. □

Proof of theorem 3:
Consider the functions g and h of definition 2, and let $V = U$. Then:

Assumption 1 follows from the assumptions of randomization, and from the exclusion restriction in assumption 5.

The continuity of g in assumption 2 follows from differentiability of $\mathbf{g}_1$ in x and y (given U), and the implicit function theorem.

The monotonicity of h in assumption 3 follows again from differentiability of $\mathbf{g}$ in x, y and z (given U), and the implicit function theorem, since

$$\partial h / \partial x = \partial X^z / \partial z = -\left[(\partial \mathbf{g} / \partial (y, x))^{-1} \cdot \partial \mathbf{g} / \partial z\right]_2 = \frac{\partial \mathbf{g}_1 / \partial y \cdot \partial \mathbf{g}_2 / \partial z}{det(\partial \mathbf{g} / \partial (x, y))},$$

where the second equality exploits the exclusion restriction $\partial \mathbf{g}_1/\partial z = 0$.

Continuity of the distribution of Z as in assumption 4 holds by assumption. □

Proof of Theorem 4:
We will focus on the identified set for $\overline{g}(x) = E[Y^x]$, under the assumption of bounded Y. The claim for $P(Y^x \leq y)$ follows by the same argument.We will show that $\overline{g} \equiv 0$ is in the closure of the identified set for $\overline{g}$; the general claim follows immediately.

The proof is constructive, providing an explicit distribution of structural functions $P(g(.,U))$ consistent with the data. The proof has two steps: (i) first we need to construct a coupling of the marginal distributions $P((X,Y)|Z=z) = P(X^z, Y^z) = P(h(z,V), g(h(z,V),U))$. Coupled values of (X,Y) for different values of Z are those assumed to have the same realization of (U,V). (ii) In the second step we show, by filling in intermediate values of $g(.,U)$ for values of X that are not realized for any Z, that any ASF satisfying the a-priori bounds can be achieved.

(i) The data and assumptions identify the (marginal) distributions of (X^z, Y^z) for $z = 1 \dots \overline{z}$, since $P((Y^z, X^z)) = P((Y,X)|Z=z)$. They are completely uninformative about the joint distribution across different values of z. The coupling of these marginal distributions can be achieved in any number of ways, we can choose for instance the independence coupling, i.e. $(X^z, Y^z) \perp (X^{z'}, Y^{z'})$ for $z \neq z'$.

(ii) Now take any realized value from this joint distribution of $\{(X^z, Y^z) : z = 1, \dots, \overline{z}\}$. For this realized value construct $g(.,U)$ as follows:[4]

$$g(x,U) = Y^x = \sum_{z=1}^{\overline{z}} K\left(\frac{x - X^z}{\tau'}\right) Y^z,$$

where K is a continuous kernel function with support $[-1,1]$ and $K(0) = 1$, and τ' is a (random) "bandwidth" parameter, given by

$$\tau' = \min(\tau, \min_{z,z'}(|X^z - X^{z'}|)).$$

With this construction we get

$$\overline{g}(x) = E[Y^x] = \sum_{z=1}^{\overline{z}} E\left[K\left(\frac{x - X^z}{\tau'}\right) Y^z\right].$$

Since we assumed that X is continuously distributed with bounded density conditional on Z, choosing τ small enough we can make this expression arbitrarily small, completing the proof. □

REFERENCES

Blundell, R. and R. Matzkin (2010): "Conditions for the existence of control functions in nonseparable simultaneous equations models," *Unpublished manuscript*.

Blundell, R. and J. Powell (2003): "Endogeneity in nonparametric and semiparametric regression models," in *Advances in economics and econometrics: Theory and applications, eighth world congress*, vol. 2, 655–679.

Chernozhukov, V., G. Imbens, and W. Newey (2007): "Instrumental variable estimation of nonseparable models," *Journal of Econometrics*, 139, 4–14.

Hahn, J. and G. Ridder (2011): "Conditional Moment Restrictions and Triangular Simultaneous Equations," *The Review of Economics and Statistics*, 93, 683–689.

[4]The exact same construction works for $P(Y^x \leq y)$.

HOROWITZ, J. (2011): "Applied nonparametric instrumental variables estimation," *Econometrica*, 79, 347–394.

IMBENS, G. (2007): "Nonadditive Models with Endogenous Regressors," *Econometric Society World Congress Series*, 75, 17–46.

IMBENS, G. W. AND J. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–467.

IMBENS, G. W. AND W. NEWEY (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77, 1481–1512.

KASY, M. (2011): "Identification in triangular systems using control functions," *Econometric Theory*, 27, 1–9.

KHAN, S. AND E. TAMER (2010): "Irregular identification, support conditions, and inverse weight estimation," *Econometrica*, 78, 2021–2042.

MANSKI, C. F. (2003): *Partial identification of probability distributions*, Springer Verlag.

MATZKIN, R. (2008): "Identification in nonparametric simultaneous equations models," *Econometrica*, 76, 945–978.

NEWEY, W. AND J. POWELL (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1565–1578.

NEWEY, W., J. POWELL, AND F. VELLA (1999): "Nonparametric estimation of triangular simultaneous equations models," *Econometrica*, 67, 565–603.