Endogeneity in Non-Additive Nonparametric Models

Thomas Wiemann
University of Chicago

TA Discussion # 4

October 27, 2021

Outline

- Triangular System of Equations
- Imbens and Newey (2009)
- Florens et al. (2008)

Triangular System of Equations

Consider the system of equations given by

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

where $Z \perp\!\!\!\perp (U, V)$. Here, Y is an outcome, X is an endogenous variable, and Z is an instrument. The object of interest is, for example, the average structural function defined by $\mathrm{ASF}(x) := E_U[g(x, U)]$.

Empirical example: Suppose that Y is lifetime earnings, X is education, Z is some cost-shifter, U is ability that is imperfectly observed by the agent, and V is a noisy signal of ability that is observed by the agent. Suppose further that the agent chooses X according to maximize lifetime earnings minus the associated cost of education, i.e.,

$$X = \arg\max_x E[g(x, U)|V, Z] - c(x, Z), \tag{2}$$

where c is the cost function. This example leads to a system as in (1).

Triangular System of Equations

The approaches we will discuss today are based on control functions: find a function $C(X, Z)$ such that $X \perp\!\!\!\perp U \,|\, C(X, Z)$. In some contexts, control functions may also be defined using conditional mean-independence.

Recall our discussion of Newey et al. (1999) in TA Discussion #2, which considered a triangular system similar to (1) but assumed additive error terms in the first and second stage. In their setting, a valid control function is given by

$$C(X, Z) := V = X - E[X|Z]. \tag{3}$$

For easy reference, see the slides here.

Today, we will discuss the approaches of Florens et al. (2008) and Imbens and Newey (2009), which do not rely on additivity in the first and second stage.

Imbens and Newey (2009)

Imbens and Newey (2009) consider the triangular system

$$\begin{aligned} Y &= g(X, U), \\ X &= h(Z, V), \end{aligned} \tag{4}$$

where $Z \perp\!\!\!\perp (U, V)$, X and V are scalar, and h is strictly monotonic in V. U is of unknown dimension.

This model seems quite flexible: Although not a generalization of the NPIV approaches discussed in TA Discussion # 2 strictly speaking (because of the independence assumption), the system of Imbens and Newey (2009) allows for very flexible specifications of heterogeneity. Note the, for example, $\frac{\partial}{\partial x} g(x, U)$ are allowed to vary with unobserved heterogeneity.

Notice, however, that the assumption of scalar V is ruling out some key economic settings. A nonseperable supply and demand model with one disturbance per equation, for example, is ruled out by (4) because the reduced form $h(Z, V)$ would generally have a two-dimensional V.

Imbens and Newey (2009) (Contd.)

In the triangular system of Imbens and Newey (2009), a control function is given by $C(X, Z) := F_{X|Z}(X|Z)$ where $F_{X|Z}$ is the conditional distribution of X given Z.

Theorem 1 of Imbens and Newey (2009):

In the triangular system given in (4), suppose

1. $Z \perp\!\!\!\perp (U, V)$, and
2. V is a continuously distributed scalar with CDF that is strictly increasing on the support of V and $h(Z, t)$ is strictly monotonic in t almost surely.

Then $X \perp\!\!\!\perp U \mid F_{X|Z}(X|Z)$.

A scalar reduced form disturbance V and monotonicity of $h(Z, V)$ in its second argument are crucial for $F_{X|Z}$ to be a control variable. Notice that assumption 2. is trivially satisfied if the first stage is additive in V.

Imbens and Newey (2009) (Contd.)

Sketch of the proof:

1) By A.2, $h^{-1}(x,z)$ exists.

2)
$$\begin{aligned} F_{X|Z}(x|z) &= P(X \le x \mid Z = z) \\ &= P(h(z,V) \le x \mid Z = z) \\ &= P(V \le h^{-1}(x,z) \mid Z = z) \\ &= P(V \le h^{-1}(x,z)) \\ &= F_V(h^{-1}(x,z)) \end{aligned}$$

So $F_{X|Z}(X|Z) = F_V(h^{-1}(X,Z)) = F_V(V)$

3) By A.2, V is continuously distributed with CDF F_V that is strictly monotonic on the support of V.
It follows that conditional expectations given V are the same as those given $F_V(V)$.

See the Borel-Kolmogorov paradox!

Imbens and Newey (2009) (Contd.)

4) Take an arbitrary bounded function $a(\cdot)$,

$$E[a(X)|V,U] = E[a(h(Z,V))|V,U]$$
$$= E[a(h(Z,V))|V] \quad Z \perp\!\!\!\perp (U,V) \Rightarrow Z \perp\!\!\!\perp U | V$$
$$= E[a(X)|V]$$

5) Take an arbitrary bounded function $b(\cdot)$,

$$E[a(X)b(U)|F(X|Z)] = E[a(X)b(U)|V] \quad \text{by step 3)}$$
$$= E[b(U)E[a(X)|V,U]|V]$$
$$= E[b(U)E[a(X)|V]|V]$$
$$= E[b(U)|V]E[a(X)|V]$$

6) B/c $a(\cdot)$, $b(\cdot)$ were arbitrary bounded functions, $X \perp\!\!\!\perp U | F_{X|Z}(X|Z)$.

(See, e.g., Resnick (2005) Exercise 4.15.)

Imbens and Newey (2009) (Contd.)

Notice that because $Y = g(X, U)$ is non-additively separable, it is not possible to simply run a regression of Y on X and the control variable $F_{X|Z}(X|Z)$, as done in the approach of Newey et al. (1999).

To identify $\mathrm{ASF}(x) := E_U[g(x, U)]$, Imbens and Newey (2009) make a an additional assumption.

Common Support: For all $X \in \mathcal{X}$, the support of $F_{X|Z}(X|Z)$ conditional on X equals the support of $F_{X|Z}(X|Z)$.

Intuitively, this says that for each value of the endogenous variable, the instrument must be such that (1) it is relevant – that is, it affects $F_{X|Z}(x|Z), \forall x \in \mathcal{X}$ – and (2) it varies sufficiently. In that sense, the common support assumption encompasses both a rank condition and a full support assumption.

The common support assumption is considered to be very strong in practice.

Imbens and Newey (2009) (Contd.)

Under the assumptions of Theorem 1 of Imbens and Newey (2009) and the common support assumption, Blundell and Powell (2006) show identification of the average structural function.

Sketch of the proof:

$$E[Y \mid X, F_{X|Z}(X|Z)] = E[Y \mid X, V] \quad \text{by A2.}$$

$$= E[g(X,U) \mid X, V]$$

$$= \int g(X,U)\, dF_{U|X,V}$$

$$= \int g(X,U)\, dF_{U|V} \qquad X \perp\!\!\!\perp U \mid V$$

$$\Rightarrow \int E[Y \mid X, V]\, dF_V = \int \left[\int g(X,U)\, dF_{U|V} \right] dF_V$$

w/ $E[Y|X,V]$ is defined on the support of V due to the common support assumption.

$$= \int g(X,U)\, dF_U$$

$$= E_U[g(X,U)]$$

Imbens and Newey (2009) (Contd.)

Imbens and Newey (2009) suggest a three-step procedure for estimating the average structural function.

1. Obtain an estimate $\hat{F}_{X|Z}$ of $F_{X|Z}$.
2. Obtain an estimate $\tilde{g}$ of $E[Y|X, F_{X|Z}]$ using $\hat{F}_{X|Z}$ inplace of the conditional density.
3. Calculate the average structural function at a point x via sample averages:

$$\widehat{\mathrm{ASF}}(x) = N^{-1} \sum_{i \in \mathcal{D}} \tilde{g}\left(x, \hat{F}_{X|Z}(x_i|z_i)\right). \tag{5}$$

The first step may leverage kernel density estimators. The second step may leverage kernel or sieve estimators.

Empirical applications appear scarce. A key reason for this is the unpopularity of the common support assumption.

Florens et al. (2008)

An alternative approach that does not rely on a common support assumption is considered in Florens et al. (2008). Instead of leaving $g(X, U)$ entirely unspecified, the authors impose a stochastic polynomial restriction on the form of the latent heterogneity:

$$Y = g(X, U) = \varphi(X) + \sum_{j=0}^{K} X^j h_j(U), \tag{6}$$

where X is the endogeneous variable, U is the disturbance of unknown dimension, and h_j are unknown functions normalized such that $E[h_j(U)] = 0, \forall j$.

Notice that the average structural function in the setting of Florens et al. (2008) is given by

$$\mathrm{ASF}(x) = E_U[g(x, U)] = E_U\left[\varphi(x) + \sum_{j=0}^{K} x^j h_j(U)\right] = \varphi(x). \tag{7}$$

Florens et al. (2008) (Contd.)

The following is an adapted version of Theorem 1 in Florens et al. (2008):

Assume equation (6) holds for finite $K \geq 1$ and suppose

1. φ is K times differentiable in X almost surely and the support of D does not contain any isolated points almost surely, and
2. there exists a (known or identifiable) control function $\tilde{V}$ s.t., $E[h_j(U)|X, Z] = E[h_j(U)|\tilde{V}] =: r_j(\tilde{V})$, and
3. X and $\tilde{V}$ are measurably separated, that is, any function of X almost surely equal to a function of $\tilde{V}$ must be equal to a constant almost surely.

Then the average structural function is identified.

Assumption 1. is a smoothness condition on the ASF. Assumption 2. is satisfied under the assumptions of Theorem 1 of Imbens and Newey (2009). Assumption 3. is a type of rank condition. A necessary condition for it to hold is that the instruments Z affect X. Theorem 2 of Florens et al. (2008) suggests that it is a rather weak condition. (In particular, the conditional support of V can depend on X.)

Sketch of the proof:

1) Suppose that $\exists (\varphi^i, r_0^i, \ldots, r_k^i)$, $i = 1, 2$, s.t.

$$E[Y \mid D = d, \tilde{V} = v] = \varphi^i(d) + \sum_{j=0}^{k} d^j E[h_j(U) \mid D = d, \tilde{V} = v]$$

$$= \varphi^i(d) + \sum_{j=0}^{k} d^j r_j^i(v)$$

by A2.

Then $[\varphi^1(d) - \varphi^2(d)] + \sum_{j=0}^{k} d^j [r_j^1(v) - r_j^2(v)] = 0$

2) By A1, $\frac{\partial^k}{\partial d^k}[\varphi^1(d) - \varphi^2(d)] + k! [r_k^1(v) - r_k^2(v)] = 0$

3) By A3, this implies $r_k^1(v) - r_k^2(v)$ is constant.

Hence, $r_k^1(v) - r_k^2(v) = E[r_k^1(\tilde{V}) - r_k^2(\tilde{V})] = 0$ b/c $E[r_j^i(\tilde{V})] = 0$

$\Rightarrow r_k^1(v) = r_k^2(v)$ a.s.

Florens et al. (2008) (Contd.)

4) Similar to step 2, consider the $(K-1)^{\text{th}}$ derivative:

$$\frac{\partial^{K-1}}{\partial d^{K-1}}\left[\varphi^1(d) - \varphi^2(d)\right] + K!\, d\, \underbrace{\left[r_K^1(v) - r_K^2(v)\right]}_{\overset{a.s.}{=}0} + (K-1)!\left[r_{K-1}^1(v) - r_{K-1}^2(v)\right] = 0$$

$$\Rightarrow \quad \frac{\partial^{K-1}}{\partial d^{K-1}}\left[\varphi^1(d) - \varphi^2(d)\right] + (K-1)!\left[r_{K-1}^1(v) - r_{K-1}^2(v)\right] = 0$$

$\Rightarrow \quad r_{K-1}^1(v) = r_{K-1}^2(v)$ a.s. by A3 and $E\left[r_j^i(\tilde{v})\right] = 0$.

5) Repeat this to show $r_j^1(v) \overset{a.s}{=} r_j^2(v)$, $j = 0, \dots, K$.

6) $$\left[\varphi^1(d) - \varphi^2(d)\right] + \sum_{j=0}^{K} d^j \underbrace{\left[r_j^1(v) - r_j^2(v)\right]}_{\overset{a.s.}{=}0} = 0$$

$$\Rightarrow \varphi^1(d) \overset{a.s.}{=} \varphi^2(d)$$

Discussion

Of course, today has only been a brief introduction to the literature on endogeneity in non-additive nonparametric models. For example, see also Professor Torgovitsky's job market paper (Torgovitsky, 2015) and the corresponding estimator discussed in Torgovitsky (2017).

In my understanding, there seems to be noticeable divide between econometric theory and empirical practice in the context of non-additive nonparametric models with endogeneity. Consistent estimators seem to be rarely implemented for the models of Florens et al. (2008) and Imbens and Newey (2009), and empirical applications can seldom be found (if at all).

A more practice-oriented approach that is related to the papers discussed today is Masten and Torgovitsky (2016). The authors impose a linear random coefficients structure on $g(X, U)$. This approach allows for consistent estimation without the common support assumption. For a recent empirical application, see Gollin and Udry (2021).

References

Blundell, R. and Powell, J. L. (2006). Endogeneity in nonparametric and semiparametric regression models. *Advances in Economics and Econometrics*, pages 312–357.

Florens, J.-P., Heckman, J. J., Meghir, C., and Vytlacil, E. (2008). Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. *Econometrica*, 76(5):1191–1206.

Gollin, D. and Udry, C. (2021). Heterogeneity, measurement error, and misallocation: Evidence from african agriculture. *Journal of Political Economy*, 129(1):1–80.

Imbens, G. W. and Newey, W. K. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5):1481–1512.

Masten, M. A. and Torgovitsky, A. (2016). Identification of instrumental variable correlated random coefficients models. *Review of Economics and Statistics*, 98(5):1001–1005.

Newey, W. K., Powell, J. L., and Vella, F. (1999). Nonparametric estimation of triangular simultaneous equations models. *Econometrica*, 67(3):565–603.

Torgovitsky, A. (2015). Identification of nonseparable models using instruments with small support. *Econometrica*, 83(3):1185–1197.

Torgovitsky, A. (2017). Minimum distance from independence estimation of nonseparable instrumental variables models. *Journal of Econometrics*, 199(1):35–48.

Newey et al. (1999)

Consider the triangular system of equations of Newey et al. (1999):

$$\begin{aligned} Y &= g_0(X) + U \\ X &= \Pi_0(Z) + V \\ E[U|V] &= E[U|V,Z] \\ E[V|Z] &= 0, \end{aligned} \tag{8}$$

where X is a vector of endogeneous variables, Z is a vector of instruments, Π_0 is an unknown function, and V is a vector of disturbances. Then

$$\begin{aligned} & E[Y|X,Z] && = g_0(X) + E[U|X,Z] \\ & && = g_0(X) + E[U|V,Z] \\ & && = g_0(X) + E[U|V] \\ & && = g_0(X) + \nu(V) \\ \Rightarrow \quad & Y && = g_0(X) + \nu(V) + \varepsilon, \quad E[\varepsilon|X,V] = 0 \end{aligned} \tag{9}$$

Notice that $\Pi_0(Z) = E[X|Z]$ is identified so that V is identified.

Newey et al. (1999) (Contd.)

Newey et al. (1999) provide various sufficient assumptions under which g_0 is identified. These conditions ensure that V and X are sufficiently distinct.

Theorem 2.1 of Newey et al. (1999): g_0 is identified, up to an additive constant, if and only if $P(\delta(X) + \gamma(V) = 0) = 1$ implies there exists a constant c_g such that $P(\delta(X) = c_g) = 1$.

For intuition, suppose that identification fails. Then, by the theorem, there are functions $\delta(X)$ and $\gamma(V)$ such that $\delta(X) + \gamma(V) = 0$ and $\delta(X)$ nonconstant. This implies a degeneracy in the joint distribution of these two random variables. Absence of such an exact relationship will imply identification. A sufficient condition is given by the following theorem:

Theorem 2.3 of Newey et al. (1999): If g_0, ν, Π_0 are differentiable, the boundary of the support of (Z, V) has zero probability, and $\Pi_0(Z) = \dim X$, then g_0 is identified.

This is a nonparametric generalization of the rank condition in TSLS.