Identification in Triangular Systems using Control Functions

Maximilian Kasy

Department of Economics, UC Berkeley

Introduction

- There is a lively literature on nonparametric IV, control functions, e.g., Newey, Powell, and Vella (1999), Imbens and Newey (2009).
- These papers discuss identification under assumptions on the first stage relationship (additive residual/monotonicity in one-dimensional residual).
- Question: Generalizability? What are necessary and sufficient conditions for the existence of control functions?
- Answer: Dimensionality restrictions on unobserved heterogeneity/family of conditional distributions.
- No control function exists in the context of a generic random coefficient model.

The nonparametric, continuous triangular system setup

$$Y = g(X, \epsilon) \tag{1}$$
$$X = h(Z, \eta) \tag{2}$$

where we assume

$$Z \perp (\epsilon, \eta) \tag{3}$$

with Z, X, Y each continuously distributed in $\mathbb{R}$.

- Z is the exogenous instrument,
- X is the treatment,
- Y is the outcome variable.

The object of interest is the structural function g.

Control functions

Idea: find a function C ("control function") of X and Z such that, for $V = C(X, Z)$,

$$X \perp \epsilon | V. \tag{4}$$

Compare Newey, Powell, and Vella (1999), Imbens and Newey (2009).

In this talk, we will discuss:

- Conditions which are both necessary and sufficient
- for the existence of control functions
- that satisfy conditional independence and support requirements.

Roadmap

- Review
- Counterexample with random coefficient first stage, failure of conditional independence
- Characterization of triangular systems, for which a control function C exist such that $V = C(X, Z)$ is a function of first stage unobservables η alone
- Characterization of triangular systems, for which C exist such that V satisfies conditional independence $X \perp \epsilon | V$
- Proof that no control function exists in the random coefficient model
- Conclusion

Why care

Recall the definition of the average structural function by Blundell and Powell (2003),

$$ASF(x) := E_\epsilon[g(x, \epsilon)].$$

Given a control function, the ASF is identified by

$$ASF(x) = E_V[E[g(X, \epsilon)|V, X = x]] = E_V[E[Y|V, X = x]]. \tag{5}$$

The first equality requires conditional independence.
Identification of $E[Y|V, X = x]$ for all V requires full support of V given X.
Under the same conditions, the quantile structural function (QSF) is identified.

Control functions proposed in literature

Newey, Powell, and Vella (1999):

$$V = C(X, Z) = X - E[X|Z]. \tag{6}$$

Justified by an additive model for h, $h(Z, \eta) = \tilde{h}(Z) + \eta$.

Imbens and Newey (2009):

$$V = C(X, Z) = F[X|Z]. \tag{7}$$

Justified by a first stage h that is strictly monotonic in a one-dimensional η.

In either case conditional independence follows from V being a function of η alone.

Counterexample - random coefficient first stage

Assume

$$X = \eta_1 + \eta_2 Z = \eta \cdot (1, Z) \tag{8}$$

$$(\eta_1, \eta_2, \epsilon) \sim N(\mu, \Sigma) \tag{9}$$

$$Z \perp (\eta, \epsilon), \tag{10}$$

and let

$$V = F(X|Z) = \Phi\left(\frac{(X - \mu_{\eta_1} - Z\mu_{\eta_2})}{\sqrt{Var(X|Z)}}\right). \tag{11}$$

Then

$$E[\epsilon|V, X] = E[\epsilon|V, Z] = E[\epsilon|X, Z] =$$

$$= \mu_\epsilon + \Phi^{-1}(V)\frac{\Sigma_{\eta_1,\epsilon} + Z\Sigma_{\eta_2,\epsilon}}{\sqrt{\Sigma_{\eta_1,\eta_1} + 2Z\Sigma_{\eta_1,\eta_2} + Z^2\Sigma_{\eta_2,\eta_2}}}. \tag{12}$$

$\Rightarrow$ Conditional independence is violated.

Q: Is there another function C for this model, such that conditional independence holds?

More generally: Under what conditions does a valid control function exist?

First characterization

Sufficient condition for conditional independence:

Proposition

If $V = C(h(Z,\eta), Z)$ does not depend on Z given η,
then conditional independence $Z \perp \epsilon | V$ holds.

Proof: By assumption, $Z \perp (\eta, \epsilon)$. As we can write V as a function of η,

$$Z|(V(\eta), \epsilon) \sim Z. \quad \square \tag{13}$$

$Z \perp \epsilon | V$ is equivalent to $X \perp \epsilon | V$ if there exists a mapping $(Z, V) \to (X, V)$, which is true if C is invertible.

The sufficient condition implies a one dimensional first stage:

Proposition

If $V = C(h(Z,\eta), Z)$ does not depend on Z given η for a $C(X,Z)$ that is smooth and almost surely invertible in X,
then $\{h(\cdot,\eta)\}$ is a one-dimensional family of functions in Z.

Sketch of proof:

- Since C is smooth and invertible with range independent of Z, V must have one dimensional range.
- Inverting C gives a function $\tilde{h}$ such that $X = \tilde{h}(Z, V)$.
- The assumption that V does not depend on Z given η (!) makes the first stage "structural" in the sense that we can write

$$h(Z,\eta) = \tilde{h}(Z, V(\eta)). \quad \square \tag{14}$$

Remarks

- Identification of the ASF requires additionally that V has full support given $X = x$, i.e., the range of $C(X, Z)$ must be independent of X.
- The family of functions $\{h(., \eta)\}$ is one-dimensional
  if and only if
  it is possible to predict the counterfactual X under manipulation of Z from knowledge of X and Z. – a much stronger requirement than identification of the ASF for the first stage relationship.

The reverse of the last proposition holds as well:

Proposition

If $\{h(.,\eta)\}$ is a one-dimensional family of functions in Z and almost surely $h(Z,\eta_1) \neq h(Z,\eta_2)$ for independent draws Z,η_1,η_2 from the respective distributions of Z and η,
then there exists a control function $V = C(h(Z,\eta),Z)$ which does not depend on Z given η.

Sketch of Proof: Choose $C(X,Z) = h(z_0, h^{-1}(Z,X))$.
Then $C(h(Z,\eta),Z) = h(z_0,\eta)$, which is a function of η alone. $\square$

Application to the random coefficient model

Here no control function satisfying the sufficient condition of proposition 1 and invertibility in X can exist. The family of functions

$$h(Z, \eta_1, \eta_2) = \eta_1 + \eta_2 Z \tag{15}$$

is two-dimensional.

This implies that we cannot predict the counterfactual X under a manipulation setting $Z = z$, $h(z, \eta)$, for a given observational unit from X and Z alone.

Second characterization

Conditional independence can hold if and only if $P(\epsilon|X,Z)$ is a one dimensional family of distributions:

Proposition

There exists a control function $V = C(X,Z)$ such that conditional independence $X \perp \epsilon | V$ holds and which is invertible in Z
if and only if
$P(\epsilon|X,Z)$ is an at most one-dimensional family of distributions that is not constant in Z if it is not constant.

Sketch of Proof:

- By invertibility, $P(\epsilon|X,Z) = P(\epsilon|X,V)$.
- By conditional independence, $P(\epsilon|X,V) = P(\epsilon|V)$.
- By invertibility of C, $dim(V) = dim(Z) = 1$.
- Reversely, let θ parametrize $P(\epsilon|X,Z)$. Take $C = \theta$. $\square$

No control function in the random coefficient model

Corollary

There exists no control function invertible in Z in the generic random coefficient model discussed before, such that conditional independence $X \perp \epsilon | V$ holds.

Sketch of proof:

$$\epsilon | X, Z \sim N\left(\mu_\epsilon + (X - \mu_{\eta_1} - \mu_{\eta_2} Z)\frac{Cov(X, \epsilon|Z)}{Var(X|Z)}, Var(\epsilon) - \frac{Cov^2(X, \epsilon|Z)}{Var(X|Z)}\right), \quad (16)$$

This is is a two-dimensional family for generic Σ. $\square$

Conclusion

- No control function exists in the random coefficient model.
- Examples of models for first stage structural relationships, where control functions do exist: First stage relationships
  - that are monotonic in unobserved heterogeneity,
  - of the form $X = h(|Z - \eta|)$, which could describe the loss from missing an unknown target η,
  - of the form $X = h(Z) \cdot \eta$, where h is of non-constant sign.

- Impossible to fully identify structural features such as the ASF or the QSF without assumptions which are hard to justify.
- Maybe more promising to look for identification of features that have some interpretable dependence on first stage parameters, e.g. the LATE as introduced in Imbens and Angrist (1994).
- Alternatively: partial identification approach pioneered by works such as Manski (2003) $\Rightarrow$ set identification of fully structural features under similarly weak assumptions.

Thanks for your time!