The Identification Zoo - Meanings of Identification in Econometrics: PART 2

Arthur Lewbel

Boston College

original 2015, heavily revised 2018

The Identification Zoo - Part 2 - sections 4 and 5.

(These are notes to accompany the survey article of the same name in the Journal of Economic Literature).

Well over two dozen types of identification appear in the econometrics literature, including (in alphabetical order):

Bayesian identification, causal identification, essential identification, eventual identification, exact identification, first order identification, frequentist identification, generic identification, global identification, identification arrangement, identification at infinity, identification by construction, identification of bounds, ill-posed identification, irregular identification, local identification, nearly-weak identification, nonparametric identification, non-robust identification, nonstandard weak identification, overidentification, parametric identification, partial identification, point identification, sampling identification, semiparametric identification, semi-strong identification, set identification, strong identification, structural identification, thin-set identification, underidentification, and weak identification.

1. Introduction

Econometric identification really means just one thing:

Model parameters or features uniquely determined from the observable population that data are drawn from.

Goals:

1. Provide a new general framework for characterizing identification concepts

2. Define and summarize, with examples, the many different terms associated with identification.

3. Show how these terms relate to each other.

4. Discuss concepts closely related to identification, e.g., observational equivalence, normalizations, and the differences in identification between structural models and randomization based reduced form (causal) models.

Table of Contents:

1. Introduction

2. Historical Roots of Identification

3. Point Identification

4. Coherence, Completeness and Reduced Forms

Continued next slide

Table of Contents - continued:

Continued next slide

Table of Contents - continued:

Part 1 had sections:

1. Introduction

2. Historical Roots of Identification

3. Point Identification

Part 3 Will have sections:

6. Identification of Functions and Sets

7. Limited Forms of Identification

8. Identification Concepts that Affect Inference

9. Conclusions

4. Coherence, Completeness, and Reduced Forms

Often ignored in practice, but analyzing coherence and completeness logically precedes the study of identification.

Most point identification proofs (e.g., all of Matzkin's 2005, 2007, 2012 identification survey examples) assume a unique reduced form, and so implicitly or explicitly assume both coherence and completeness.

In contrast, incompleteness often results in θ only being set identified.

Completeness here extends the concept of statistical completeness.

Statistical completeness in parametric models is associated with sufficient statistics, and is discussed in Newey and Powell (2003) for identification of nonparametric IV models.

Let $Y \in \Omega_y$ be a vector of endogenous variables.

Let $V \in \Omega_v$ be a set of observables and unobservables that determine Y. V can contain unknown parameters, exogenous observed covariates, and error terms.

Assume the model M is $Y = H(Y, V)$

This means each model value $m \in M$ implies a DGP in which V and Y satisfy $Y = H(Y, V)$.

M is *coherent* if for each $v \in \Omega_v$ there exists at least one $y \in \Omega_y$ satisfying $y = H(y, v)$.

M is *complete* if for each $v \in \Omega_v$ there exists at most one $y \in \Omega_y$ satisfying $y = H(y, v)$.

Having both coherence and completeness guarantees existence of a unique *reduced form* $y = G(v)$, defined by $G(v) = H[G(v), v]$.

Note: above terminology is due to Tamer (2003).

Gourieroux, Laffont, and Monfort (1980) defined the model to be coherent if, in Tamer's terminology, the model is both coherent and complete. Heckman (1978) referred to the combination of coherence and completeness, as the *principal assumption* and as *conditions for existence of the model*.

Tamer (2003) shows incoherent or incomplete models can arise in simultaneous games, e.g., the industry entry game of Bresnahan and Reiss (1991).
Here incoherency corresponds to the game having no Nash equilibrium Incompleteness to the case of multiple equilibria.

Aradillas-Lopez (2005) removes the incompleteness in these games by showing that a unique Nash equilibrium exists when the player's have some private information.

Coherency and completeness in limiting dependent variable systems of equation are studied by Blundell and Smith (1994), Dagenais (1997), and Lewbel (2007).

Consider the simple, harmless looking model

$$Y_1 = I(Y_2 + U_1 \geq 0), \qquad Y_2 = \theta Y_1 + U_2$$

where I is the indicator function, θ is a coefficient, U_1 and U_2 are unobserved error terms, and $V = (\theta, U_1, U_2)$.

These could be game reaction functions, where player one makes a discrete choice Y_1 (such as whether to enter a market), and player two makes a continuous decision Y_2 (such as the quantity to produce).

Estimate this model by standard methods, e.g., maximum likelihood assuming U_1 and U_2 normal? Big mistake!

Substituting out Y_2 get $Y_1 = I(\theta Y_1 + U_1 + U_2 \geq 0)$.
If $-\theta \leq U_1 + U_2 < 0$ then both $Y_1 = 0$ and $Y_1 = 1$ satisfy the model: incomplete.
If $0 \leq U_1 + U_2 < -\theta$ then neither $Y_1 = 0$ nor $Y_1 = 1$ satisfy the model: incoherent.

Model is coherent and complete iff $\theta = 0$ or if $U_1 + U_2$ is constrained to not lie between zero and $-\theta$.

$\theta = 0$ makes this system triangular, not fully simultaneous.

Lewbel (2007) shows simultaneous systems containing a dummy endogenous variable generally need to either be triangular or to restrict the supports of the error terms to be coherent and complete.

But the direction of triangularity can vary, e.g., are coherent and complete if

$$Y_1 = I(DY_2 + U_1 \geq 0), \quad Y_2 = (1 - D)\,\theta Y_1 + U_2$$

where D is binary (indicates who moves first in the game).

Typically, incoherence arises in fully simultaneous models but not in triangular systems.

Incoherence can often indicate model misspecification: for some v, there is no possible y, whereas in reality some y is observed.

Incompleteness is a model that is not fully specified. For example, structural restrictions that take the form of inequalities rather than equalities will often be incomplete, since they can imply multiple possible values of y for the same v.

Parameters of incoherent or incomplete models can sometimes be point identified and estimated (see Tamer 2003), but not by using a reduced form, and typically can only make limited predictions about y.

Model incompleteness often leads to set identification instead of point identification.

Incompleteness or incoherency can arise in models with multiple decision makers.

Models of a single optimizing agent are usually coherent though sometimes incomplete (if, e.g., the same utility or profit level can be attained in more than one way).

Equilibrium selection mechanisms or rules for tie breaking in optimization models are techniques for resolving incompleteness.

5. Causal Reduced Form vs Structural Model Identification

Recent rapid rise in so-called reduced form or causal inference methods, often based on randomization. The so-called "credibility revolution." Example proponents: Angrist and Pischke (2008), Levitt and List (2009), Banerjee and Duflo (2009).

Most identification theory was developed for structural models, Came before the recent rise of randomized/causal modeling in econometrics.

Most surveys of identification in econometrics (e.g. Matzkin 2007, 2012) don't mention this literature's use of identification

Whats in a name?

Proponents, sometimes perjoratively called "randomistas," call their methodology "reduced form."
But without model structure, what are these estimates reduced forms of?

Sometimes the methods are called "causal modeling" or "causal inference."
But causal just refers to the estimand: how the change in treatment changes an outcome. Many structural models identify causal effects.
Noted by Heckman (2008), some causal effects cannot be identified by randomization based methods

Relevant notions of causality go back to Hume (1739) and Mill (1851)
Temporal based definitions of causality are Granger (1969) and Sims (1972).
See Hoover (2006) for a survey of alternative notions of causality in economics and econometrics.

Another term used is "mostly harmless econometrics," title of Angrist and Pischke's (2008) book (from a satirical science fiction novel that also features "infinite improbability").

Finally, these methods are sometimes just called the "treatment effects" or "program evaluation" literature,
But randomization is neither necessary nor sufficient for identification of treatment effects in general.

For simplicity will just call these methods just "causal."

Further potential confusion: this literature also defines "local identification" and "instrument validity" differently from the earlier structural identification literature.

Causal methods are generally characterized by:

1. focus on identification and estimation of treatment effects, and

2. emphasize experimental or natural randomization as their primary source of identification.

Exceptions
a. numerous structural analyses (e.g. Roy 1951 model), also identify treatment effects.
b. some reduced form methods, like diff-in-diff, are not based on random assignment.
c. some literatures, e.g Pearl (2000, 2009), use structural type assumptions (like causal diagrams) to help identify causal effects.
d. growing number of empirical structural analyses use randomized data.

The Causal model ideal is randomized treatment (RCT - randomized control trial or natural experiment).

Commonly in economic, treatment is not random; at best a variable that correlates with treatment (the instrument) is randomly assigned.

Much causal literature is devoted to designing and interpreting RCTs (e.g., popular in development economics).

Much of the rest entails searching for and exploiting instruments that can argue are randomly determined (thereby satisfying some conditional independence assumptions).

Much structural identification (like Cowles) is for simultaneous systems.

In contrast, consider Rubin (1974) counterfactual notation:
$Y(t)$, the outcome Y that would have occurred if treatment T had equaled t. This notation generally assumes T affects Y, not the reverse.

Causal inference doesn't care about identifying structural parameters.
Causal estimands to be identified are summary measures.
Example: Average treatment effect (ATE): $\theta = E\left(Y\left(1\right) - Y\left(0\right)\right)$.

Other common causal estimands:
average treatment effects on the treated (ATT),
local average treatments effects (LATE),
marginal treatment effects (MTE),
quantile treatment effects (QTE).

Typical structural model obstacles to identification are problems like simultaneity, nonuniqueness, multiple solutions.

The main obstacle to identification of causal parameters is that they are defined in terms of unobservable counterfactuals.

However, causal estimands can also be expressed in structural terms - identification issues then recast as endogeneity issues. See, e.g., Pearl (2000, 2009, 2015) and Heckman (2008, 2010).

5.1 Randomized Causal or Structural Identification? Do Both!

Heated debates on relative merits of causal vs structural methods.

Angrist and Pischke's (2008) "Mostly Harmless Econometrics" title implies that structural modeling includes harmful econometrics.

Opponents refer to many reduced form studies as cuteonomics.

Economics journals are known for being either friendly or hostile to structural methods.

These debates have even spilled over into the popular press, e.g., Scheiber (2007) in the New Republic: "Freaks and Geeks; How Freakonomics is Ruining the Dismal Science."

Researchers do NOT need to choose between either searching for randomized instruments or building structural models.

For identification, the best strategy is to exploit both approaches.

1. Causal inference based on randomization can be augmented with structural econometric methods to cope with data problems like attrition, sample selection, measurement error, and contamination bias.

Example Conlon and Mortimer (2016):
Field experiment randomly removed a popular brand from vending machines.
But outcomes only observed when machines are serviced.
Combine causal analysis with a simple structural model of purchase timing.

2. Structural Identification often depends on independence assumptions, which are more likely to hold under randomization. Good causal instruments are also generally good structural instruments

Example: Ahlfeldt, Redding, Sturm, and Wolf (2015), uses a natural experiment (the partition of Berlin) to identify sa tructural model of gains associated with people living and working near each other in cities.

The 2018 Frisch Medal committee said this paper, "provides an outstanding example of how to credibly and transparently use a quasi-experimental approach to structurally estimate model parameters."

3. Causal effects can provide useful benchmarks for structure.

Example: Estimate a structural model from a large survey
Calculate an ATE from small randomized trials drawn from the same
underlying population.
Check if the ATE implied by estimated structural parameters equals the
causally estimated ATE.

Example: Andrews, Gentzkow and Shapiro (2017, 2018), construct
summary statistics based on the estimated joint distribution of reduced
form parameters (like LATE) and structural model parameters.
Use these statistics to assess how structural results depend on intuitively
transparent identifying information.

4. Economic theory and structure can provide guidance regarding the external validity of causally identified parameters.

Example: In regression discontinuity design (RDD) the cutoff or threshold is often a policy relevant variable, e.g., the grade at which one qualifies for a scholarship.

Dong and Lewbel (2015): Assume a mild structural assumption, local policy invariance. Then can identify how RDD LATE would change if threshold changed, even when no such threshold change is observed.

Their estimator also measures the stability of RDD LATE, see Cerulli, Dong, Lewbel, and Poulsen (2017).

Example: Frölich and Huber (2017) use structural assumptions to separate direct from indirect effects of treatment on outcomes.

Example: Rosenzweig and Udry (2016) use structure to model how ATE (returns from policy interventions) estimated from randomized control trials, varies with macro shocks like weather.

5. Can use causal methods to link randomized treatments to observable variables, then use structure to relate these observables to more policy relevant treatments and outcomes.

Example: Middle aged and older women in India have high mortality rates. Why?

Calvi (2016) uses a causal analysis (a change in inheritance laws) to link changes in women's household bargaining power to their health outcomes.

She then constructs structural estimates of women's relative poverty rates based on their bargaining power, measured by estimated household resource shares.

Shows estimated relative poverty rates almost perfectly match women's higher than expected mortality rates by age (correlation of .96).

Most causal analyses include informal speculation regarding the wider implications of estimated treatment effects.

Calvi could have done that with just the causal part of her analysis that linked power to health, speculating on the likely connection to mortality.

Such speculation is nothing more or less than crude structural modeling.

More convincing is the rigor imposed by real structural identification and estimation, as in Calvi's demonstration that such estimates really can explain the observed excess mortality.

Another, related example is Calvi, Lewbel, and Tommasi (2017):

Estimates LATE: treatment is women's control of most resources within a household, outcomes are family health measures, instrument is changes in inheritence laws.

Treatment can't be directly observed, and so is estimated using a structural model of household behavior.

Structural models can be misspecified and have estimation errors. The estimated treatment will be mismeasured for some households.

Propose and apply an alternative estimator, called MR-LATE (mismeasurement robust LATE), that accounts for the potential measurement errors.

Here structure allows identification of a more policy relevant LATE than would otherwise be possible.

6. Big data analyses on large data sets can uncover promising correlations.

Structural analyses of such data might be used to uncover possible economic and behavioral mechanisms that underlie these correlations.

Randomization can be used to investigate the causal direction of these correlations.

Those who argue that machine learning, natural experiments, and randomized controlled trials are replacing structural economic modeling and theory are wronger than wrong.

As ML and experiments uncover ever more previously unknown correlations and connections, the desire to understand these newfound relationships will rise, thereby increasing, not decreasing, the demand for structural economic theory and models.

7. Structural type assumptions can clarify when and how causal effects may be identified.

Examples: Pearl (2000, 2009) and Pearl and Mackenzie (2018): structural causal models and causal diagrams, directed acyclic graphs,

Another line of research that formally unifies structural and randomization based approaches to causal modeling is Vytlacil (2002), Heckman, Urzua and Vytlacil (2006), Heckman and Vytlacil (2007), and Heckman (2008, 2010).

5.2 Randomized Causal or Structural Identification: An Example

Let Y_i be an observed outcome for individual i.
Let T_i be a binary endogenous regressor (treatment indicator).
Let Z_i be a binary covariate (potential instrument), correlated with T.

ϕ includes the first and second moments of (Y, T, Z).
In practice the DGP is such that these moments can be consistently estimated by sample averages.

We will compare a structural vs a causal model. Other comparisons: Imbens and Angrist (1994), Angrist, Imbens, and Rubin (1996), Vytlacil (2002) and Heckman (1997, 2008, 2010).

Goal here: illustrate differences between a popular structural and a popular causal model, in terms of their assumptions and notation.

Structural model: linear regression $Y = a + bT + e$ for error e and constants a and b, under the standard instrumental variables identifying assumption that $E(eZ) = 0$.

Causal model: the local average treatment effect (LATE) model of Imbens and Angrist (1994).

Key difference: Structural model makes a behavioral (response to treatment) assumption: any heterogeneity of the impact of T on Y is in the error term e, assumed uncorrelated with Z. Identifies ATE.

LATE model drops this behavioral restriction, and instead has a "no defiers" assumption, and identifies the average effect of T on Y for a subpopulation called compliers.

What makes one model or analysis structural and the other causal?

Structural models usually have:
fixed "deep" parameters (like the coefficient b),
and behavioral restrictions (like $E(eZ) = 0$).

Structural models: generally models of economic behavior, ideally derived from (and identified by) economic theory.

Causal models usually have:
As few behavioral assumptions as possible,
randomization as the primary source of identification,
a focus on identifying treatment effects: the average (in a subpopulation) change in an outcome from changing a covariate value (the treatment).

NOTE: what we are here calling structural vs. causal restrictions are just common examples in the literature.

They do NOT define what makes a model structural or causal.

Not all structural models are linear regressions
not all linear regressions are structural
not all causal analyses areLATEs.

These are just typical examples of the kinds of models each literature uses and the kinds of restrictions that each literature imposes.

Y is outcome, T is a binary endogenous regressor (treatment indicator), Z is a binary covariate (potential instrument), correlated with T.

ϕ includes the first and second moments of (Y, T, Z).

Let $c = cov(Z, Y) / cov(Z, T)$, which is therefore identified (by construction).

No model has yet been specified, but this c would be the limiting value of the estimated coefficient of T in a linear instrumental variables regression of Y on a constant and on T, using a constant and Z as instruments.

Start with the most general possible model for Y and T:

$$Y = G\left(T, Z, \widetilde{U}\right) \text{ and } T = R\left(Y, Z, \widetilde{V}\right)$$

$\widetilde{U}$ and $\widetilde{V}$ are vectors of unobservable errors.
G and R are arbitrary, unknown functions.

Assume G does not depend directly on Z (exclusion).
and R does not depend directly on Y (triangular).

Model becomes $Y = G\left(T, \widetilde{U}\right)$ and $T = R\left(Z, \widetilde{V}\right)$.

Have $Y = G\left(T, \widetilde{U}\right)$ and $T = R\left(Z, \widetilde{V}\right)$

Let $U_0 = G\left(0, \widetilde{U}\right), \quad U_1 = G\left(1, \widetilde{U}\right) - G\left(0, \widetilde{U}\right),$
$\quad V_0 = R\left(0, \widetilde{V}\right), \quad V_1 = R\left(1, \widetilde{V}\right) - R\left(0, \widetilde{V}\right).$

Since both T and Z are binary, this lets us without loss of generality rewrite the model as

$Y = U_0 + U_1 T$ and $T = V_0 + V_1 Z.$

This is a linear random coefficients model.

Also: Since T and Z are binary, we also have that V_0 and $V_1 + V_0$ are binary.

Now write the same model causally.

Identification involving outcomes based on treatment selection goes back to Neyman (1923), Wright (1925), Haavelmo (1943), Wald (1943), Roy (1951), and Heckman (1978). But typically use the counterfactual causal notation of Rubin (1974):

$Y(t)$ is the random variable denoting outcome Y if $T = t$.
$T(z)$ is the random variable denoting treatment T if $Z = z$.

How do the notations relate? $Y(t) = G\left(t, \widetilde{U}\right)$, $T(z) = R\left(z, \widetilde{V}\right)$
Get $Y(0) = U_0$, $Y(1) = U_0 + U_1$, $T(0) = V_0$ and $T(1) = V_0 + V_1$.

Exclusion: $Y(0)$ and $Y(1)$ do not depend on Z.
If started with $Y(t, z)$, the exclusion is then $Y(t, 1) = Y(t, 0)$.
Triangular: $T(0)$ and $T(1)$ do not depend on Y.

So far, structural and causal are identical; both assumed nothing except exclusions/triangular structure.

Now make a common causal inference assumption: SUTVA (Stable Unit Treatment Value Assumption):

SUTVA means that any one person's outcome is unaffected by the treatment that other people receive.

The term SUTVA was coined by Rubin (1980), concept goes back at least to Cox (1958), and perhaps implicit in Fisher (1935) and Neyman (1923, 1935).

SUTVA, Stable Unit Treatment Value Assumption, formal definition:

For any two different people i and j, $Y_i(0)$ and $Y_i(1)$ are independent of T_j.

SUTVA essentially rules out social interactions, peer effects, most general equilibrium effects.

SUTVA is behavioral. But commonly accepted in causal, since enforce-able in many experimental settings (by, e.g., physically separating experimental subjects).

Randomized natural or field experiments can violate the SUTVA, due to people interacting with each other, either directly or via market effects.

If SUTVA is violated one must usually either
1. make behavioral (structural) assumptions to gain point identification, or
2. construct complicated experiments to identify the spillover effects, or
3. settle for set identification of causal effects.
4. adjust inference to account for the failure of point identification
Examples: See Rosenbaum (2007), Manski (2013), Lazzati (2015), Angelucci and Di Maro (2016), and Laffers and Mellace (2016)

Structural analogs to SUTVA are restrictions on correlations between $\{U_{0i}, U_{1i}\}$ and $\{V_{0j}, V_{1j}, Z_j\}$.

For simplicity, here assume $\{U_{0i}, U_{1i}, V_{0i}, V_{1i}, Z_i\}$ independent across i. Is sufficient, stronger than necessary for SUTVA.

A structural alternative to SUTVA would be to model dependence across individuals, e.g., a social interactions model as in Blume, Brock, Durlauf, and Ioannides (2011).

$Y = U_0 + U_1 T$ and $T = V_0 + V_1 Z$. Same model as $Y(T)$ and $T(Z)$.

Have assumed exclusions, and SUTVA type independence across people. What else?

By construction T depends on V_0 and V_1.
Endogenous T could also correlate with U_0 and U_1.

Want Z to be an instrument, so add the causal assumption:
$\{Y(1), Y(0), T(1), T(0)\}$ is independent of Z.
This is called unconfoundedness. Is equivalent to Z randomly assigned.

In the structural notation, unconfoundedness is
$\{U_1, U_0, V_0, V_1\}$ is independent of Z.

$Y = U_0 + U_1 T$ and $T = V_0 + V_1 Z$. Same model as $Y(T)$ and $T(Z)$.

Assumptions so far: Exclusions/triangular, SUTVA/independence, unconfoundedness.

Assume also structurally that $E(V_1) \neq 0$, or equivalently in the causal notation, that $E(T(1) - T(0)) \neq 0$. This assumption ensures that the instrument Z is relevant.

Given just these assumptions, what does the instrumental variables estimand c equal?

$$c = \frac{cov\,(Z, Y)}{cov\,(Z, T)} = \frac{cov\,(Z, U_0 + U_1 T)}{cov\,(Z, (V_0 + V_1 Z))} = \frac{cov\,(Z, U_0 + U_1\,(V_0 + V_1 Z))}{cov\,(Z, (V_0 + V_1 Z))}$$

$$= \frac{cov\,(Z, U_1 V_1 Z)}{cov\,(Z, V_1 Z)} = \frac{E\,(U_1 V_1)\,var(Z)}{E\,(V_1)\,var\,(Z)} = \frac{E\,(U_1 V_1)}{E\,(V_1)}$$

For this problem, the only difference between structural and causal approaches will be that different additional assumptions are made to interpret the equation $c = E\,(U_1 V_1)\,/\,E\,(V_1)$.

Consider structural identification first:

Model is still: $Y = U_0 + U_1 T$ and $T = V_0 + V_1 Z$.
Can rewrite this as: $Y = a + bT + e$,
where $b = E(U_1)$, $e = (U_1 - b)T + U_0 - a$.

Structural identification question: When does IV estimand
$c = cov(Z, Y) / cov(Z, T)$ equal the structural coefficient b?

Standard answer: Z being a valid structural model instrument requires
$cov(e, Z) = 0$.

And what does b mean?

Recall definition ATE $= E[Y(1) - Y(0)]$. Under our assumptions,
$U_1 = Y(1) - Y(0)$. Therefore $E[Y(1) - Y(0)] = E(U_1) = b$. So the
structural coefficient b is precisely the causal ATE.

With $Y = a + bT + e$, have Z is valid structural instrument if $cov(e, Z) = 0$, and then IV estimand $c = b$.

But what does $cov(e, Z) = 0$ mean? We showed $cov(e, Z) = cov(U_1, V_1) var(Z)$, so $cov(e, Z) = 0$ if $cov(U_1, V_1) = 0$. Note from above that

$$
\begin{aligned}
c &= \frac{E(U_1 V_1)}{E(V_1)} = \frac{E(U_1) E(V_1) + cov(U_1, V_1)}{E(V_1)} \\
&= E(U_1) + \frac{cov(U_1, V_1)}{E(V_1)} = b + \frac{cov(U_1, V_1)}{E(V_1)}
\end{aligned}
$$

So again $c = b$ if $cov(U_1, V_1) = 0$.

Summary: the structural identifying assumption is $cov(U_1, V_1) = 0$. Under this assumption, $c = b = $ ATE for the population.

Note: under other assumptions, can interpret c as the average treatment effect on the treated. See, e.g., Heckman (1997).

Now look at causal identification, and then compare.

Define a complier: A person i for whom Z_i and T_i are the same random variable. If a complier has $Z = 0$, then he has $T = 0$, and if has $Z = 1$, then has $T = 1$.

Define a defier: A person i for whom Z_i and $1 - T_i$ are the same random variable.

We can't know who the compliers and defiers are! Because they are defined in terms of counterfactuals.

If someone has $Z = 0$, we can see if they have $T = 0$ or $T = 1$, but we can't know what their T would have been if they had been assigned $Z = 1$.

Recall $T = V_0 + V_1 Z$.   All the possibilities relating T to Z are:

Compliers: $T = Z$: $V_0 = 0$ and $V_1 = 1$.
Defiers: $T = 1 - Z$: $V_0 = 1$ and $V_1 = -1$.
Always takers: $T = 1$ for any Z: $V_0 = 1$ and $V_1 = 0$.
Never takers: $T = 0$ for any Z: $V_0 = 0$ and $V_1 = 0$.

Compliers are the only type that have $V_1 = 1$.

Imbens and Angrist (1994) define the LATE (local average treatment effect) as the ATE just among compliers.

(note: their use of the word local is not the same as in local identification).

In our notation here, LATE $= E\left[Y\left(1\right) - Y\left(0\right) \mid V_1 = 1\right]$.

Now revisit the IV coefficient c. Let P_v be the probability that $V_1 = v$. Then, by definition of expectations,

$$c = \frac{E(U_1 V_1)}{E(V_1)} =$$

$$\frac{E(U_1 V_1 \mid V_1 = 1) P_1 + E(U_1 V_1 \mid V_1 = 0) P_0 + E(U_1 V_1 \mid V_1 = -1) P_{-1}}{E(V_1 \mid V_1 = 1) P_1 + E(V_1 \mid V_1 = 0) P_0 + E(V_1 \mid V_1 = -1) P_{-1}}$$

$$= \frac{E(U_1 \mid V_1 = 1) P_1 - E(U_1 \mid V_1 = -1) P_{-1}}{P_1 - P_{-1}}$$

Imbens and Angrist (1994): Assume that there are no defiers in the population. This rules out the $V_1 = -1$ case, making $P_{-1} = 0$.

Then above simplifies to: $c = E(U_1 \mid V_1 = 1) =$ LATE.

So causal assumes $V_1 \neq -1$ (no defiers) and identifies LATE for just compliers.

Comparing Structural and Causal:

Model is $Y = U_0 + U_1 T$ and $T = V_0 + V_1 Z$.
where V_1 and $V_1 + V_0$ are binary.

Given our assumptions: Exclusions/triangular, SUTVA/independence, unconfoundedness;

Structural makes the additional assumption $cov(U_1, V_1) = 0$. This restricts heterogeneity of the treatment effect U_1. Identifies ATE for the population.

Causal instead makes the additional assumption $V_1 \neq -1$ (no defiers). This restricts heterogeneity of types of individuals. Identifies LATE for compliers, says nothing about the treatment effect on non-compliers.

Comparing Structural and Causal - continued:

Neither the structural nor causal assumptions above are a priori more plausible. Both require assumptions on unobservables, not just $cov(U_1, V_1) = 0$ or $V_1 = -1$, but also by assumptions like exclusion restrictions and SUTVA.

Structural $cov(U_1, V_1) = 0$ restricts assumed heterogeneity of the treatment effect U_1. Assumes an individual's type V_1 is on average unrelated to the size of personal treatment effect U_1. But delivers b, the population ATE.

Causal $V_1 \neq -1$ (no defiers) restricts heterogeneity of types of individuals. Compared to $cov(U_1, V_1) = 0$, has the big advantage of not imposing restrictions on the Y equation. But only delivers LATE.

Comparing Structural and Causal - continued:

Can show with binary Z and T and our other assumptions that the structural restriction $cov(U_1, V_1) = 0$ requires either that defiers exist, or that everyone is a complier, or that $E(U_1 \mid V_1 = 0) = E(U_1 \mid V_1 = 1)$. Pretty limiting compared to causal $V_1 \neq -1$.

But, this setup is very restrictive - what if have other covariates, or if Z or T has more than two values?

Then for causal analysis, the number of types (compliers, deniers, etc.) becomes large and assumptions about them gets complicated and unwieldy.

But structural stays just the same: $cov(U_1, V_1) = 0$ or equivalently, $cov(e, Z) = 0$.

Some Limits to LATE:

LATE says nothing about treatment effect on noncompliers. Maybe ok if Z is a policy variable.

People who seek treatment for reasons related to the outcome are by construction not compliers! Generally need structure like a Roy (1951) to identify causal relationships when treatment correlates with outcomes. See, e.g., Heckman (2008).

Compliers are essentially people for whom treatment was randomly assigned. Unreliable for policy if compliers are not representative. Stucture might also be unreliable, if the population doesn't approximately satisfy assumed behavioral restrictions.

We don't know who the compliers are. Though we can identify the probability someone is a complier, conditional on their observable characteristics (Angrist and Pischke 2008). Can also get bounds on ATE (Manski 1990, Balke and Pearl 1997).

Another property of LATE: the definition of a complier depends on Z.

Suppose we saw a different instrument $\widetilde{Z}$ instead of Z.
Let $c = cov\left(Z, Y\right) / cov\left(Z, T\right)$ and $\widetilde{c} = cov\left(\widetilde{Z}, Y\right) / cov\left(\widetilde{Z}, T\right)$.

If $\widetilde{Z}$ is a valid instrument in the structural sense, then $c = \widetilde{c} = b = $ ATE in the population.

Also could then test instrument validity. If reject $c = \widetilde{c}$, then either $\widetilde{Z}$ or Z is not a (structurally) valid instrument.

But even if both are causally valid, will ususally have $c \neq \widetilde{c}$. Both are LATEs, but for different (unknown) compliers.

Are there any testable implications of causal instrument validity? Yes, there exist a few weak inequalities one can test, e.g., P_1 can't be negative. See, e.g., Kitagawa (2015).

5.3 Causal vs Structural Simultaneous Systems

Assume now a simultaneous system of equations, say
$Y = U_0 + U_1 X$ and $X = H(Y, Z, V)$.
(U_0, U_1) and V are unobserved error vectors.
Observe Y, X, and instrument Z. Using X not T because might not be treatment (need not be binary).

As before, analyze meaning of $c = cov(Z, Y) / cov(Z, X)$.

Structural analysis is exactly the same as before: Have $Y = a + bX + e$ where $b = E(U_1)$ and $e = U_0 + (U_1 - b)X$. If $cov(e, X) = 0$ then $c = b =$ average marginal effect of X on Y.

In contrast, causal analysis is *much* more complicated. Angrist, Graddy and Imbens (2000) show under LATE type assumptions, c equals a complicated weighted average of U_1, with weights that depend on Z and some weights are zero.

Another limitation of applying causal methods to simultaneous systems:

Counterfactual notation implicitly assumes that reduced forms exist (indeed, are often called reduced form models).

Example: an incomplete model like in Section 3, endogenous Y and X have
$Y = I(X + U \geq 0), \quad X = Y + Z + V$, and $-1 \leq U + Z + V < 0$.

Observe Y, X, Z. Could be a game where Y and X are player choices/actions.

A reduced form $Y = m(Z, U, V)$ does not exist.

So assuming any $Y(z)$ imposes additional info that is not in the model.

Another limitation in applying causal analyses to simultaneous systems: SUTVA again.

Many simultaneous systems have treatment of one person affects outcomes of others. Examples: peer effect models, social interactions models, network models, general equilibrium models.

Example: Progresa and Oportunidades in Mexico. Widely cited example of randomized treatment.
But people move to communities for the program, or interact with treated individuals, or act expecting the program may expand to their own community (see, e.g., Behrman and Todd 1999).

As noted earlier, point identification if SUTVA is violated usually requires behavioral, structural type assumptions (see, e.g., Manski 2013).

These may be why causal methods popular in traditionally partial equilibrium analysis fields (e.g., labor economics and micro development), but not where general equilibrium models are the norm (e.g., industrial organization and macroeconomics).

5.4 Randomized Causal vs Structural Identification: Conclusions

Primary advantage of causal methods: History of success in the sciences.
Randomized controlled trials (RCT's) referred to as the gold standard for
empirical work in other fields.
Randomization in economic field and natural experiments strives to
approximate that gold standard.

Another virtue of causal methods: treatment effects are fundamentally
interpretable estimands.
As long as the framework is coherent and complete, so potential outcomes
are well defined, then causal estimands like ATE, ATT, etc., are
meaningful.

In contrast, when a structural model is misspecified, the deep policy
invariant parameters it attempts to identify can become meaningless, or at
best difficult to interpret.

An advantage of structural models:

They contain info about underlying behavioral structure.

Structural models can incorporate and test restrictions implied by economic theory

Can get identification from theory restrictions, without randomization.

Structural models can cope with data issues that cause difficulties for causal analyses.

Examples: Structures that account for self selection into treatment, or for measurement problems such as sample attrition, or for SUTVA violations like network effects and social interactions.

Features that are obstacles to causal inference can help identification in structural models.

Example: with random assignment, impossible to identify features of the joint distribution of potential outcomes, such as what fraction of the population would benefit from treatment (see, e.g., Heckman 2008).

In the Roy (1951) model and competing risks models with covariates, selection is based on maximization over potential outcomes.

This provides information about the joint distribution of potential outcomes that could not be uncovered by random assignment. See Heckman and Honoré (1989, 1990) and Lee and Lewbel (2013).

Common objection to the structural approach:
Where are the deep parameters structural models have uncovered?
What are their widely agreed upon values?

One answer: look at calibration.
Calibrated models in macroeconomics have parameters users treat as known.
Many parameters have values, mostly from prior structural work, that users largely agree upon.

Example: the calibrated value of the rate of time preference.

Structural model experience has lead to consensus among economists regarding reasonable ranges of values for many parameters, such as price and income elasticities.

Empirical structural analyses have found behavioral relationships, like Engel's (1857) law, that appear to hold up almost universally.

Main disadvantage of imposing behavioral, structural type restrictions for identification?

Reality is complicated, structural models oversimplify, so are misspecified.

"All models are wrong, but some are useful" - George Box (1979),

Generally don't know how much misspecification corrupts interpretation and applicability of structural model estimates.

But, causal models can also be misspecified:
Example: Even with randomized Z, in LATE the population may contain defiers or SUTVA may be violated.

More specification trouble for both:

Both structural and causal models generally require covariates:

Including covariates requires functional forms that may be parametrically misspecified (e.g, linear probability models) or are nonparametric, suffering the curse of dimensionality upon estimation.

Another issue for both structural and causal models:

External validity. If the environment changes even slightly, how would an identified parameter or treatment effect change?

Deep structural parameters are supposed to be constants across environments.
But behavioral restrictions may hold in one context and not another
Or correct specifications may change due to, e.g., the Lucas (1976) critique.

External validity is a larger problem with causal models
These have no underlying economic or behavioral restrictions that one can assess in a new environment.

Example: Rosenzweig and Udry (2016) document how macro conditions that one cannot control for, like weather, can dramatically affect estimated treatment effects obtained in randomized controlled trials.

Empirical methods of assessing external validity is an active area of research.

Economic policy is often concerned with characteristics that cannot be directly observed.

Examples: utility, risk aversion, noncognitive skills, bargaining power, expectations, or social welfare.

Structural, behavioral assumptions are usually required to link observables these elusive concepts, and hence to evaluate the impacts of treatment on them.

Summary:

Identification based on randomization in economics has been called "credibility revolution."

But in practice both causal and structural methods depend on a host of assumptions

Either can lead to invalid (or "incredible") inference when their identifying assumptions are violated.

Both sources of identification have advantages and disadvantages.

Best practice: Do both!

End of Part 2

Part 1 had sections:

1. Introduction

2. Historical Roots of Identification

3. Point Identification

Part 3 will have sections:

6. Identification of Functions and Sets

7. Limited Forms of Identification

8. Identification Concepts that Affect Inference

9. Conclusions