

Lecture 2

Introduction to Structural Models

Tyler Ransom

ECON 6343, University of Oklahoma

Today and next class

- What is a structural model?
 - Examples
- What steps are required to write a research paper that uses a structural model?
 - Computational steps

Attribution

Some of today's material is based on **these excellent slides** by Phil Haile (Yale)

Causality: the goal of econometrics

- In any econometric endeavor, the goal is to uncover causal relationships
- Causality is crucial because it's the only way to know the effects of a policy
 - e.g. what is the effect of reading to my child on her cognitive development?
 - we can't answer this by simply looking at *Corr*(read to child, cog. test score)
 - this correlation is contaminated by omitted variable bias:
 - parents who regularly read to their children likely also regularly do other things favorable to child development

Causality requires a counterfactual

- Causality is defined in terms of a **counterfactual**
 - what would've been the outcome if everything were the same except the policy?
 - this is the notion of ceteris paribus in principles of economics
 - theory says quantity demanded will decrease if price increases, *ceteris paribus*
 - what is child's test score if everything were the same *except* regular reading time?
- **"Causal effect"**: difference between reality and the most plausible counterfactual
- There are many ways to estimate a causal effect

Two types of empirical work in econ (Haile, 2019)

1. **Descriptive**

- Document facts about observable quantities
- These facts sometimes suggest a causal relationship
- e.g. raw earnings difference between college graduates and HS graduates

2. **Structural**

- Estimate parameters of a *data generating process* (DGP) which are assumed to be invariant to policy changes or other counterfactuals
- Once we know the DGP, we can make causal inferences
- e.g. estimate a DGP that specifies how cognitive ability and family background relate to the decision to enroll in college and to post-graduation earnings

Brief history of the term "structural"

- Dates back to Hurwicz (1950) and Koopmans and Reiersol (1950)
- A **structure** is a data generating process
- i.e. a set of functional or probabilistic relationships between observable and latent variables which implies a joint distribution of the observables
- The goal of structural estimation, then, is to estimate the parameters of the DGP
- This allows us to make counterfactual comparisons, i.e. perform causal inference
- Note that "structural" here refers to basically all of modern econometrics

Brief history of the term "reduced form"

- The term **reduced form** refers to solving a structural model
- The structural model may have endogenous variables on both sides of the equation
- But the reduced form puts all endogenous variables on the left hand side
- All exogenous variables and error terms are on the right hand side
- Classic example: supply and demand
 - Two equations, two endogenous variables (price, quantity)
 - In equilibrium, reduced form has P and Q as respective LHS variables
 - RHS contains observable and unobservable determinants of S and D

How these terms tend to be used today

- Reduced form tends to refer to linear models estimated by RCT, IV, DID, RDD, etc.
 - Methods that try to exploit randomization (or quasi-randomization)
 - Synonymous with the phrase "identification strategy"
- Structural tends to refer to non-linear models that are more difficult to estimate
 - Methods that make explicit the (typically large) set of maintained assumptions
 - Methods that focus on settings where RCTs would be infeasible
- Both of these terms are misnomers, but this is how they are used today
- For the rest of these slides, I'm going to misuse them (like most others today)
- See Haile (2019) for more semantic details

Structural vs. Reduced Form Methods

- There is a lot of animosity between structural and reduced-form practitioners
- Keane (2010) calls reduced-form methods "atheoretic"
 - others refer to reduced-form or behavioral approaches as "cuteconomics"
- Angrist and Pischke (2009) titled their book *Mostly Harmless Econometrics*
 - implying that structural methods include "harmful" econometrics (Lewbel, 2019)
- In fact, **we need both** methodological approaches to answer policy questions

Structural vs. Reduced Form Methods

- Blundell (2010) outlines three broad concerns determining what method to use:
 1. The nature of the question to be answered
 2. The type and quality of data available
 3. The mechanism by which individuals are allocated or receive the policy
 - i.e. random assignment or obeying rules of economic theory (like utility max.)

"Just as an experiment needs to be carefully designed, the identification of a structural economic model needs to be carefully argued."

"Poorly designed quasi-experiments have little to offer, but so too do poorly focused structural estimations."

Structural vs. Reduced Form Methods

- As an observation, our profession under-invests in structural methods
- I believe this is mainly because their implementation can be difficult
- For every structural paper produced, 5 or 6 RF papers could be produced
- This matters for publishing, tenure, and training the next generation
- Nonetheless, structural methods can be immensely useful
- But they take longer, so their science progresses relatively more slowly

RCT as structural estimation

- All causal inference is structural in nature (as correctly defined)
- Structural estimation need not be difficult; models need not be complex
- An RCT is a structural model that can be evaluated *descriptively*
 - No fancy econometrics needed: just compute $\bar{y}_{\text{Treatment}} - \bar{y}_{\text{Control}}$
- This is because great effort was expended at the randomization step
- The experimenters had a (structural) model in mind when defining treatment

Not every DGP can be evaluated by RCT

- Some causal questions of interest can't be answered with an RCT
- For a variety of reasons, experimentation may be too costly, unethical, etc.
 - e.g. can't randomize a merger of two large firms, or a person's height
- Without randomization, we have to rely on observational data
- This requires more complex econometric methods to estimate the DGP
 - Need to explicitly specify how unobservables relate to other parts of model

In the example of wage related to education, unobservables could include family background, personality, quality of education, genetics, ability (cognitive and non-cognitive) which can create issues in identifying the causal effect of education on wages.

Key parameters of any economic model

- As mentioned above, we assume that DGP parameters are policy-invariant
- These parameters tend to be related to **economic fundamentals**:
 - commodities (land, capital, labor, oil, etc.)
 - demographics
 - preferences
 - production technology
 - information and expectations
 - space (includes networks & social interactions)

Reading-to-Children Example: Reduced-form example

- A **reduced-form** (as misused today) approach would look like the following:
 1. recruit a group of families to participate in a reading study
 2. randomize into "no-read" and "read" groups
 3. after some period of time, give their children a cognitive test
 4. compare the average scores of children across each of the groups

Reading-to-Children Example: Structural approach

- A **structural** (as misused today) approach would look like the following:
 1. write a model of child skill formation (Cunha, Heckman, and Schennach, 2010)
Assumptions are key!
 2. gather data on parental and child time use and child test scores
 3. estimate the parameters of the child skill formation model
 4. use model to simulate counterfactual policies (e.g. where reading is set to 0)
 5. compare average scores in counterfactual and status quo

Reading-to-Children Example: Hybrid approach

- A hybrid approach would do the following:
 1. estimate the skill formation parameters
 2. leverage randomization to better estimate/validate the model
 - e.g. by allowing for identification of a parameter previously not identifiable
 - e.g. recover randomization-implied ATE using structural parameter estimates
 3. use the validated structural model to explore other counterfactuals
- A great example of this hybrid approach is Delavande and Zafar (2019)

What is identification? (Lewbel, 2019)

- **Identification:** model parameters being *uniquely determined* from the *observable population* that generates the data
- identification is never a question about a sample of data
- it is a question about the population from which the sample is drawn
- there are many different terms for identification in econometrics
- but the unifying definition is the one given above
- Lewbel (2019) lists 33 different terms from the econometrics literature
(I include all of the terms on the penultimate slide of this deck)

More formal definition (Lewbel, 2019)

- Let θ denote a set of unknown parameters that we would like to learn about, and ideally, estimate
- e.g. regressor coefficients, average treatment effects, or error distributions
- identification asks what could be learned about parameters θ from observable data
In the education example, its the identification of beta1 which is problematic given that the error term includes a host of omitted variables.
- if we knew the population that data are drawn from, would θ be known?
- if not, what could be learned about θ ?

Why is identification important? (Lewbel, 2019)

- The study of identification logically precedes estimation, inference, and testing
Identification -> estimation -> statistical inference -> hypothesis testing (Frequentist approach)
- For θ to be identified, alternative values of θ must imply different distributions of the observable data using Monte Carlo simulations
- If θ is not identified, then we cannot hope to find a consistent estimator for θ
- More generally, identification failures complicate statistical analyses of models, so recognizing lack of identification, and searching for restrictions that suffice to attain identification, are fundamentally important problems in econometric modeling

Reduced-form vs. Structural Identification

- How is "identification" used differently in reduced-form vs. structural econometrics?
- In reduced-form econometrics (a.k.a. causal modeling):
 - Typically talk of an "identification strategy" (i.e. randomization setup)
 - Focus is on estimation of treatment effects, not "deep parameters"
 - Relies on randomization from some kind of randomized or natural experiment
- In structural econometrics:
 - Typically talk of "establishing identification" (i.e. sufficient variation in data)
Are the parameters uniquely determined given the data variation?
 - In complex models, can be difficult to do without imposing more assumptions

In epidemiological literature, there is also a difference between reduced-form and structural model. For example, flattening the curve for COVID. Structural model may not be perfectly identified in this case. Heterogeneous agent model being incorporated in SIR model of infection.

The Credibility Revolution (Angrist and Pischke, 2009)

- What makes an identification strategy credible?
- Identification means separating selection from treatment
- This is best done when treatment is randomized
- Randomization is also how the natural sciences make discoveries
 - Typically done via controlled experiments
- The closer a reduced-form model is to an RCT, the better

(Note that controlled experiments are impossible to do with humans)

Examples of Identification Strategies

- Randomized experiments, field experiments, lab experiments
- Instrumental variables, regression discontinuity
- Difference in differences, synthetic control methods
- Matching methods (nearest neighbor, propensity score, ...)
- OLS that does not suffer from omitted variable bias
- These are almost exclusively estimated using linear econometric models
- Credibility is proportional to the "cleanliness" of randomization
 - You'll often hear about whether something is "cleanly" identified

Credible Structural Models

- What makes a structural model credible?
- At the very least, the model should **"fit the data"** (i.e. reproduce key patterns)
- But that is usually a low bar to clear, so additional criteria are required
- Results should also "make sense" (i.e. conform to economic theory)
 - e.g. An upward-sloping demand curve would violate this criterion
 - or a result that says agents prefer lower income or fewer profits
- Typically requires modeling heterogeneity in preferences or productivity
- Another difficulty: separating preferences from constraints (choice set)

Structural Methods

- Unlike reduced-form methods, there is not a set "toolkit" of techniques
- Rather, structural modeling is a bit *ad hoc* or a bit "Wild West"
- Whereas RF methods almost exclusively focus on linear econometric models,
- Structural methods overwhelmingly require use of non-linear econometric models
 - e.g. discrete choice models, fixed point mappings for equilibrium models, etc.
- Structural models are typically estimated by GMM or Maximum Likelihood
- Computational know-how helps speed up the process of estimating these models
- These topics will be the focus of this class

Structural papers in various fields of econ

- **Labor:** Keane and Wolpin (1997), education investment decisions
- **IO:** Berry, Levinsohn, and Pakes (1995), demand estimation using market-level data
- **Urban:** Ahlfeldt, Redding, Sturm, and Wolf (2015), estimation of spatial agglomeration
Spatial equilibrium models - location choice, rents, etc. Ahlfeldt et al (2015) uses falling of Berlin wall on spatial equilibrium (combination of reduced form and structural methods)
- **Environmental:** Rudik (2020), quantify uncertainty in environmental IAMs
- **Public:** Bayer, McMillan, Murphy, and Timmins (2016), dynamic Tiebout sorting model
Tiebout sorting model: how people sort into neighborhoods
- **Macro:** all DSGE models
- **International:** Jin and Shen (2020), coordination of sovereign debt

Internal and External Validity

- **Internal validity** refers to "how causal" an estimated parameter is
 - "This approach is internally valid" \Rightarrow no selection bias
- **External validity** refers to generalizability of estimates to new contexts
- Typically, RF approaches are very good at internal validity but not at external validity
- On the other hand, if economic agents behave similarly across contexts, structural models can be externally valid
- RF and structural methods used together can improve both internal and external validity

Example: Internal vs. External Validity

- Suppose we want to measure earth's gravitational force, g
- We can measure g by timing how long it takes various objects to fall some distance
- We can do this with objects of varying mass and of varying fall distances
 - Using this data, we can estimate earth's g
- But what about the g on Mars? Or some other planet?
- For this we need a model of what exactly determines g
 - (A planet's mass and proximity to other large objects)
- This model will tell us what g is on planets we haven't yet visited

33 terms for identification

- 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
- weak identification

References

- Ahlfeldt, G. M, S. J. Redding, D. M. Sturm, et al. (2015). "The Economics of Density: Evidence From the Berlin Wall". In: *Econometrica* 83.6, pp. 2127-2189. DOI: [10.3982/ECTA10876](#).
- Angrist, J. D. and J. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press. ISBN: 0691120358.
- Bayer, P, R. McMillan, A. Murphy, et al. (2016). "A Dynamic Model of Demand for Houses and Neighborhoods". In: *Econometrica* 84.3, pp. 893-942. DOI: [10.3982/ECTA10170](#).
- Berry, S, J. Levinsohn, and A. Pakes (1995). "Automobile Prices in Market Equilibrium". In: *Econometrica* 63.4, pp. 841-890. URL: <http://www.jstor.org/stable/2171802>.
- Blundell, R. (2010). "Comments on: Structural vs. Atheoretic Approaches to Econometrics" by Michael Keane". In: *Journal of Econometrics* 156.1, pp. 25-26. DOI: [10.1016/j.jeconom.2009.09.005](#).
- Cunha, F, J. J. Heckman, and S. M. Schennach (2010). "Estimating the Technology of Cognitive and Noncognitive Skill Formation". In: *Econometrica* 78.3, pp. 883-931. DOI: [10.3982/ECTA6551](#).
- Delavande, A. and B. Zafar (2019). "University Choice: The Role of Expected Earnings, Nonpecuniary Outcomes, and Financial Constraints". In: *Journal of Political Economy* 127.5, pp. 2343-2393. DOI: [10.1086/701808](#).
- Haile, P. (2019). *Structural vs. Reduced Form" Language and Models in Empirical Economics*. Lecture Slides. Yale University. URL: <http://www.econ.yale.edu/~pah29/intro.pdf>.
- Hurwicz, L. (1950). "Generalization of the Concept of Identification". In: *Statistical Inference in Dynamic Economic Models*. Hoboken, NJ: John Wiley and Sons, pp. 245-257.
- Jin, H. and H. Shen (2020). "Foreign Asset Accumulation Among Emerging Market Economies: A Case for Coordination". In: *Review of Economic Dynamics* 35.1, pp. 54-73. DOI: [10.1016/j.red.2019.04.006](#).
- Keane, M. P. (2010). "Structural vs. Atheoretic Approaches to Econometrics". In: *Journal of Econometrics* 156.1, pp. 3-20. DOI: [10.1016/j.jeconom.2009.09.003](#).
- Keane, M. P. and K. I. Wolpin (1997). "The Career Decisions of Young Men". In: *Journal of Political Economy* 105.3, pp. 473-522. DOI: [10.1086/262080](#).
- Koopmans, T. C. and O. Reiersol (1950). "The Identification of Structural Characteristics". In: *The Annals of Mathematical Statistics* 21.2, pp. 165-181. URL: <http://www.jstor.org/stable/2236899>.
- Lewbel, A. (2019). "The Identification Zoo: Meanings of Identification in Econometrics". In: *Journal of Economic Literature* 57.4, pp. 835-903. DOI: [10.1257/jel.20181361](#).
- Rudik, I. (2020). "Optimal Climate Policy When Damages Are Unknown". In: *American Economic Journal: Economic Policy* 12.2, pp. 340-373. DOI: [10.1257/pol.20160541](#).