

# AGEC 652 - Lecture 6.1

## Introduction to Structural Estimation Methods

Diego S. Cardoso

Spring 2023

# Course roadmap

1. Intro to Scientific Computing
2. Numerical operations and representations
3. Systems of equations
4. Function approximation (Skipped)
5. Optimization
6. Structural estimation
  1. **Introduction** ← You are here
  2. A review of estimation methods
  3. Estimation of single-agent models
  4. Estimation of multiple-agent models

\*These slides are loosely based on Reiss & Wolak (2007), Timmins & Schlenker (2009) and course materials by Matt Woerman (UMass) and Simon Quinn (Oxford).

**So, where have we got to so far?**

# Example 1

Suppose we have a constant elasticity demand function:

$$q(p) = p^{-0.2}$$

In equilibrium, quantity demanded is  $q^* = 2$

**Q: What price clears the market in equilibrium?**

# Example 1

A: Just invert the demand function

$$2 = p^{-0.2}$$

$$p^* = 2^{-5} \checkmark$$

Your calculator can do the rest

## Example 2

Suppose the demand function is now

$$q(p) = 0.5p^{-0.2} + 0.5p^{-0.5},$$

a weighted average of two CE demand functions

In equilibrium, quantity demanded is  $q^* = 2$

**Q: What price clears the market in equilibrium?**

**How would you solve it now?**

## Example 3

Consider a two period ag commodity market model

- Period 1: Farmer makes acreage decisions for planting
- Period 2: Per-acre yield realizes, equilibrium crop price clears the market

The farmer's policy function is:

$$a(E[p]) = \frac{1}{2} + \frac{1}{2}E[p]$$



## Example 3

Yield  $\hat{y}$  is stochastic and only realized after planting period. It follows  $\hat{y} \sim \mathcal{N}(1, 0.1)$

Crop production is

$$q = a\hat{y}$$

Demand is given by

$$p(q) = 3 - 2q$$

**Q: How many acres get planted?**

## Example 3

**Q: How many acres get planted?**

A: We plug in the stochastic quantity and take expectations

$$p(\hat{y}) = 3 - 2a\hat{y}$$

$$a = \frac{1}{2} + \frac{1}{2}(3 - 2aE[\hat{y}])$$

Rearrange and solve:

$$a^* = 1$$

## Example 3

Now suppose the government implements a price floor such that  $p > 1$ .

We now have that

$$p(\hat{y}) = \max(1, 3 - 2a\hat{y})$$

**Q: How many acres get planted with a price floor?**

## Example 3

**Q: How many acres get planted with a price floor?**

This is analytically intractable

The max operator is non-linear so we can't pass the expectation through

$$E[\max(1, 3 - 2a\hat{y})] \neq \max(1, E[3 - 2a\hat{y}])$$

**How would you solve it now?**

# So, where have we got to so far?

We've seen many tools that can help us solve many of the well-defined problems in economics

By now, we can:

- Numerically calculate difficult integrals and derivatives
  - E.g.: calculate expectations of weird functions, evaluate FOCs and SOC
- Solve linear and nonlinear systems
  - E.g.: solve for equilibria in single and multiple-agent models

# So, where have we got to so far?

By now, we can:

- Approximate unknown functions and solve functional equations
  - E.g.: interpolate between data points and solve for best-response functions
- Unconstrained and constrained optimization
  - E.g.: solve utility maximization and cost minimization problems

And, hopefully, you picked up useful programming skills and learned Julia and GitHub along the way

# So, where have we got to so far?

In most of these problems, we have (or assumed we have) all the necessary parameters and we calculated a solution

**But, for practical purposes, where do model parameter values come from?**

We need to estimate them!

- Sometimes, we are lucky and someone else has estimated them for us
- Sometimes, the parameters themselves are not the focus of a paper and we can "roughly" find a set of parameters that reproduce a stylized fact or illustrate a mechanism (aka, calibration)
- If those are not the case, we need data and econometric tools to estimate parameters

# Time to bring data in

**Key idea: economic agents behave as if they are solving some economic problem**

1. We conceptualize the problem and formulate its respective model and parameters
2. We observe agents solving the economic problem and use observations to infer the value of parameters that best describe their behavior

This is a rough description of **structural econometric models**



# Readings for this unit

- (Sections 1 to 3) Reiss, P. C., & Wolak, F. A. (2007). Structural econometric modeling: Rationales and examples from industrial organization. *Handbook of econometrics*, 6, 4277-4415.
- Timmins, C., & Schlenker, W. (2009). Reduced-form versus structural modeling in environmental and resource economics. *Annual Review of Resource Economics*, 1(1), 351-380.

# Structural econometric models

# Structural econometric models

*(It is) the use of (statistical) models based in economic theory. Structural modeling attempts to use data to identify the parameters of an underlying economic model, based on models of individual choice or aggregate relations derived from them. (Nevo and Whinston, 2010)*

This approach combines theoretical and statistical models to study an economic phenomenon

- Economic theory tell us about how variables relate to each other
- Statistical properties rationalize why theory does not perfectly explain the data

This is a different from **reduced-form** or "non-structural" estimation (more on the differences later)

# Steps in structural modeling

## 1. Economic model

The economic model describes the economic environment

- Who are the decision-makers (agents)? What information is available to them?
  - What are their objectives? What choices can they make?
- What are the primitives?
  - Technologies available, preference structure, endowments
- How are markets and institutions organized?
- Given the economic environment, what is an equilibrium in this market?
  - E.g.: Perfect competition? Nash equilibrium?

# Steps in structural modeling

## 1. Economic model

Formally, economic theory formulates mathematical statements about how observable *endogenous* variables  $y$  relate to observable *explanatory* variables  $x$

- It may also tell us about how  $y$  relates to *unobservable* variables  $\xi$

Statements are usually in the form of equalities

$$y = g(x, \xi, \Theta)$$

where  $g$  is a known function and  $\Theta$  is a set of unknown parameters or functions

# Steps in structural modeling

## 2. Statistical (or econometric) model

The economic model is expanded into a statistical model to rationalize the fact that observed data does not fit the economic model perfectly

- This is done by including **unobservable** variables

Types of unobservables:

- Unobservable to the econometrician only: agents make decisions based on information that is not observed in the available data
- Unobservable to agents
- Optimization errors by economic agents
- Measurement errors in observables

# Steps in structural modeling

## 2. Statistical (or econometric) model

Formally, the statistical model establishes a joint distribution of  $x$ ,  $\xi$ , and possibly other unobservables

$$f(x, \xi)$$

Combined, the theoretical and statistical models give us an empirical model that can *rationalize* observable outcomes

- This is, for any set of observable variables, there some  $\Theta$  that can explain the observed behavior under the theoretical and statistical assumptions

# Steps in structural modeling

## 3. Estimation procedure

In this step, we first parameterize the model  $\Rightarrow$  we define what is in  $\Theta$

This is done by

- Specifying functional forms and their parameters
- Making distributional assumptions and selecting distribution parameters



# Steps in structural modeling

## 3. Estimation procedure

With  $\Theta$  defined, we construct a statistical object that relates data and the empirical model

The two most common objects are

1. A likelihood function  $L(y, x|\Theta)$ 
    - Use the Maximum Likelihood Estimator to find  $\Theta$
  2. Conditional moments  $E[y|x, \Theta]$ 
    - Use Generalized Method of Moments to find  $\Theta$
- In this step, we can also formulate specification and other hypothesis tests

# A simple structural model

Suppose we want to estimate output elasticities of capital and labor

We observe output  $Y_t$  and input use of capital  $K_t$  and labor  $L_t$

**Step 1:** An economic model with Cobb-Douglas production with parameters  $A, \alpha, \beta$

$$Q_t = AL_t^\alpha K_t^\beta$$

# A simple structural model

**Step 2:** A statistical model that specifies an unobservable productivity component  $\epsilon_t$

$$Q_t = AL_t^\alpha K_t^\beta \exp(\epsilon_t)$$

And we assume that  $\epsilon_t$  are independently drawn, with values following a mean-zero normal distribution, and independent of firm input choices (i.e., unobservable by agents)

$$\epsilon_t \sim N(0, \sigma^2)$$

$$E[\epsilon_t | K_t, L_t] = 0$$

P.S.: In Unit 4's optimization tutorial, we kinda did structural estimation

# A simple structural model

**Step 3:** We can rewrite the statistical equation with logs

$$\log Q_t = \log A + \alpha \log L_t + \beta \log K_t + \epsilon_t$$

Given our assumptions on  $\epsilon$ , we can estimate  $\alpha$  and  $\beta$  with OLS by regressing  $\log Q$  on a constant,  $\log L$ , and  $\log K$

But if, in reality,  $\epsilon_t$  includes differences in productivity that are *observable to the firm when it makes decisions* but *not observable to the econometrician*, then  $\epsilon_t$  is correlated with inputs

- I.e.,  $E[\epsilon_t | K_t, L_t] \neq 0 \Rightarrow$  we made a bad assumption for  $\epsilon_t$
- We can't use OLS; but we could use IV estimation instead if we have instruments for  $\log L$  and  $\log K$

# Structural estimation procedures

In this simple example, the model was simple enough that we could get a **estimating equation** that was linear in the parameters

- In those cases, we can use OLS (if assumptions hold), 2SLS, and other **linear estimation methods**

In other models, we might not have the same luck but still get closed-form expressions that are nonlinear in parameters. Then, we can resort to **nonlinear estimation methods**

- If we assume specific distributions for unobservables, we can calculate log-likelihoods  $\Rightarrow$  MLE
- If we assume moment conditions, we can calculate empirical conditional moments  $\Rightarrow$  GMM

# Structural estimation procedures

In more complicated cases, we do not have closed-form statistical objects to perform the estimation and need to calculate them numerically at every iteration of the estimation procedure

- When we can't get closed-form expressions because there is a difficult integral in the statistical object, we can use **simulation-based estimators**
  - The "simulation" part comes from Monte Carlo integration: we simulate/draw samples from a distribution
  - If the integral is to calculate log-likelihood, we get the **Maximum Simulated Likelihood**
  - If the integral is to calculate an expectation, we get the **Method of Simulated Moments**

# Structural estimation procedures

In other complicated cases, theoretical models might not have closed-form solutions either. These give rise to what is generally referred to as "**nested estimation**"

1. We start with some initial guess of the parameter value
2. **Inner loop:** given current parameter values, solve the model numerically for those parameters and calculate statistics we want to optimize over
  - E.g., minimize the sum of squared residuals or maximize log-likelihood
  - This step might involve rootfinding, optimization, solving functional equations, etc
3. **Outer loop:** based on the statistics generated in the inner loop, an estimation algorithm picks the next set of parameters
4. Repeat steps 2 and 3 until convergence

# Why add structure to an econometric model?

Structural models are useful to

1. Estimate unobserved values of economic interest
  - Marginal costs, returns to scale, risk preferences, discount rates, switching costs, etc
2. Perform counterfactual or policy simulations
  - What would happen if we introduced a new policy or changed the economic environment?
  - These analyses require primitives that are invariant to policy changes
3. Calculate welfare
  - Welfare usually requires structural modeling of utility and profit functions
4. Compare competing economic theories



# Reduced-form models

**Reduced-form models** focus on estimating key "high-level" response parameters rather than primitives of a model

- These include from simple summary statistics (means, conditional means, and higher moments) to the estimation of average treatment effects and time series models

These models generally use economic theory to guide the selection of  $y$  and  $x$ . Their goals are usually to estimate joint population density  $f(x, y)$  or some related statistical objects such as

- Conditional density  $f(y|x)$  or expectation  $E[y|x]$
- Conditional covariance (or correlation)  $Cov(y|x)$
- Conditional  $\alpha$  quantiles  $Q_\alpha(y|x)$

# Structural vs. reduced-form models

## THERE IS NO SINGLE APPROACH THAT IS BETTER FOR ALL PURPOSES

There might be the right approach for your *specific research question* given the data available and the institutional details

Each approach has its *pros* and *cons*

# Structural vs. reduced-form models

**Reduced-form models** are a great tool to find out "**what works**"

They can be used flexibly to estimate the effects of policy interventions, aggregate responses, etc

- They require few theoretical and statistical assumptions
- With good research designs, quasi-experimental methods can deliver credible identification of effects without imposing an untestable economic model
  - Though statistical assumptions can imply some constraints on the unstated theoretical model

# Structural vs. reduced-form models

**Reduced-form models** have more limitations when studying "**why it works**"

- This is done by isolating mechanisms and might require large amounts of detailed data

These models are less useful to study **how it would work** under different circumstances

- Counterfactual analysis requires parameters that are invariant to changes in the economic environment
  - This is rarely the case of high-level parameters and statistics

# Structural vs. reduced-form models

In contrast, **structural models** are a useful tool for counterfactual analysis, testing hypothesis on mechanisms, and quantifying unobservable values of economic interest (costs, welfare, etc)

However, these models require

- Existing economic theory that addresses the empirical context
- Many assumptions to formulate an estimable empirical model
  - Unrealistic assumptions deliver non-credible results
  - Modeling assumptions are difficult to test formally

# Structural vs. reduced-form models

There is no trade-off between structure and credibility

- Once again, it all depends on your research question and data

In some cases, adding structure is the only way to deliver credible answers

- E.g.: counterfactual questions and non-marginal welfare effects

# Structural vs. reduced-form models

These approaches are *complements rather than substitutes*

When studying a specific policy:

- Reduced-form methods can give a credible estimate of treatment effects
- Structural methods can investigate the underlying mechanisms that generate the treatment effects

Regardless the method: **Good data + identification = less restrictive assumptions**

- The better the data, the less we need to assume about the data generating process

# Homework

In this unit, we will work with data

In Julia, the `DataFrames` package offers interface to work with data tables

- It's quite similar to R

Read and follow the steps in this tutorial:

- <https://julia.school/julia/dataframes>