Overview of Structural Estimation

University of Texas — Austin

Toni M. Whited

October 2015

Outline

- What is it?
- Brief review of SMM
- Why do it?
- Tips on how to do it.
- One nice paper.

First, some terminology

- It makes no sense to say “structural model.”
- All economic models are “structural.”
- Usually when people say “structural model,” they really mean “dynamic model.”
- It makes a lot of sense to talk about “structural” versus “reduced-form” estimation.

Statistical and Economic Models

- A statistical model describes the relation between two or more random variables:
$$y = x\beta + u$$
- An economic model starts with assumptions about
  - agents' preferences
  - constraints
  - firms' production functions
  - some notion of equilibrium, etc.
- Then it makes predictions about the relation between observable, often endogenous variables.

Structural Estimation

- Structural estimation is an attempt to estimate an economic model's parameters and assess model fit.

- Parameters to estimate often include
  - Preference parameters (e.g., risk aversion coefficient)
  - Technology parameters (e.g. production function's curvature)
  - Other time-invariant institutional features (e.g. agents' bargaining power, financing frictions)

What is Structural Estimation?

- Structural estimation ascertains whether optimal decisions implied by a model resemble actual decisions by firms.

- Structural estimation may or may not require a dynamic—as opposed to a static—model.
  - Hennessy and Whited (2005, JF) $\longrightarrow$ dynamic
  - Albuquerque and Schroth (2010, JFE) $\longrightarrow$ static

What Ktnds of Econometrics

- GMM
- MLE
- SMM
- SMLE
- Indirect Inference
- All of the two-step methods used by the structural IO folks.

Moments and Likelihoods

- The moment estimators ascertain whether model-implied moments in the data resemble real-data moments.
- The likelihood estimators use economic models to construct the likelihoods for MLE.
- In both cases
  - The simulation estimators are used with models without closed-form estimating equations.
  - GMM and MLE are used with models with closed-form estimating equations.

What Kind of Model to Use

- The model has to be an economic rather than a statistical model
- Should produce realistic magnitudes and distributions
  - No two-state, "profits-are-either-hi-or-lo" models
  - Usually no two- or three-period models
  - Model should usually be fully dynamic
- The goal of the model is usually not to present "new" theory, but to present a formal structure through which to view data.

A brief poet's guide to dynamic optimization models

- The goal is to maximize the expected present value of some cash flows.
- The cash flows are functions of
  - stochastic state variables (demand shock)
  - non stochastic state variables (capital stock)
  - choice variables (investment)
- The solution has two parts
  - value function: maps the state variables into the expected present value
  - policy function: maps the state variables into the choice variables
- Then you can use a random number generator and the policy function to simulate data.

Calibration versus Structural Estimation

- Calibration:
  - Take many parameter values from other papers
  - Usually have more parameters than moments—model isn't identified, can't put standard errors on parameters
  - Mainly a theoretical exercise
- Structural estimation:
  - Infer parameter values from the data
  - Get standard errors for parameters
  - An empirical exercise

Calibration versus Structural Estimation

- Both:
  - Can assess how well model fits the data—no statistical tests with calibration
  - Can use model to ask counterfactual questions:
  - What would happen if we shocked this variable?
  - How would world look if we changed this parameter's value?

Structural versus Reduced-Form Estimation

- Reduced-form:
  - What is the (causal) effect of X on Y?
- Structural
  - Why does X affect Y?
  - What are the magnitudes of the parameters?
  - How well does theory line up with the data?
  - How would the world look if one of the parameters (counterfactually) changed?
  - What would happen if you (counterfactually) shocked the system?

Structural versus Reduced-Form Terminology

- Structural models often imply a "reduced-form," meaning a statistical model describing the relation between the observables generated by the model.

- Example from "Debt Dynamics." One reduced-form prediction from the structural model:

$$\text{Leverage}_{it} = \beta_0 + \beta_1 Q_{it} + \beta_2 \pi_{it} + u_{it}$$

  The regression slopes β are nonlinear functions of the model's structural parameters.

- The true (no u_{it}) reduced-form may actually be nonlinear in π_{it} and Q_{it}.

Structural Estimation Buys You Three Things

From least to most interesting

- Estimates of interesting economic primitives
- Deep tests of theory:
  - Formal, joint tests of multiple predictions (e.g., test of overidentifying restrictions in GMM/SMM)
  - Testing quantitative, not just directional, predictions
  - “Seeing where models fail opens doors to future research”
    Example: Mehra and Prescott (1985), equity premium puzzle
- Can answer interesting counterfactual questions

Pros and Cons

- Reduced-form
  - "Fewer" assumptions? Results more convincing?
  - Easier to do
  - Easier to understand: larger audience
- Structural
  - Often the only feasible option for answering certain important questions
  - Tough to find good instruments
  - The connection between theory and tests of theory is extremely tight, which allows more transparent interpretation of any results. In structural, we "put the model first" and make it explicit.
  - Results generalize better.
  - For job market: Makes you look smart

Example: How does equity market misvaluation affect firm policies?

	Reduced-form Baker, Stein, Wurgler (2003, QJE)	Structural Warusawitharana and Whited (2015, RFS)
Approach	Regress investment on a proxy for misvaluation – Q	Estimate structural parameters by SMM. Use counterfactual analysis to measure effects of misvaluation on policies
Data challenges	Difficult to measure misvaluation	Use observed data on firm decisions viewed through the lens of a model
Identifying assumptions	Exogenous variation in equity market misvaluation Proxies for misvaluation are "good"	Model captures the important determinants of relevant firm policies

The structural approach complements existing reduced-form research by:

- overcoming certain data challenges
- imposing a different type of identifying assumption.

Pros and Cons: Bottom Line

- Choose the approach that lets you answer your question most easily and convincingly.
- If structural and reduced-form will both get the job done, go reduced-form!!

Why use complicated simulation estimators?

- "Better data and computing facilities, have made sensible things simple.[1]"
- Simulation estimators make transparent the relationship between economic models and the equations used to estimate them.
- It seems odd to call computationally intensive econometric techniques "simple," especially given the all-too-frequent criticism that they are a "black box."
- However, there exists a tension between realism and the sorts of models that can produce closed-form estimating equations.
- Better models that can explain more phenomenon may not lend themselves to closed-form solutions.

[1] Ariel Pakes, Keynote address delivered at the inaugural International Industrial Organization Conference in Boston, April 2003

Setup

- Let Let x_i be an $i.i.d.$ data vector, $i = 1, \ldots, n$.
- Let $y_{is}(b)$ be an $i.i.d.$ simulated vector from simulation s, $i = 1, \ldots, N$, and $s = 1, \ldots, S$.
- The simulated data vector, $y_{is}(b)$, depends on a vector of structural parameters, b.
- The goal is to estimate b by matching a set of *simulated moments*, denoted as $h(y_{is}(b))$, with the corresponding set of actual *data moments*, denoted as $h(x_i)$.
- The simulated moments, $h(y_{is}(b))$ are functions of the parameter vector b because the moments will differ depending on the choice of b.

Moment Matching

- The first step is to estimate $h(x_i)$ using the actual data.
- The second step is to construct S simulated data sets based on a given parameter vector.
- For each of these data sets, estimate a simulated moment, $h(y_{is}(b))$.
- Note that you have to make the *exact* same calculations on the simulated data as you do on the real data.
- SN need not equal n.
- Michaelides and Ng (2000, *Journal of Econometrics*) find that good finite sample performance requires a simulated sample that is approximately ten times as large as the actual data sample.

Moment Matching

- Now let's figure out how to match the moments:
- Define

$$g_n(b) = n^{-1} \sum_{i=1}^{n} \left[h(x_i) - S^{-1} \sum_{s=1}^{S} h(y_{is}(b)) \right].$$

- The simulated moments estimator of b is then defined as the solution to the minimization of

$$\hat{b} = \arg\min_b Q(b, n) \equiv g_n(b)' \hat{W}_n g_n(b),$$

- $\hat{W}_n$ is a positive definite matrix that converges in probability to a deterministic positive definite matrix W.

Weight Matrix

- In most applications one can calculate the weight matrix as the inverse of the variance covariance matrix of $h\left(x_i\right)$.
- The nice part about this type of weight matrix is that you can estimate it **before** you start minimizing the SMM objective function.
- It does not depend on any parameters, so you do not have to iterate on it, as one does in many GMM applications.

Inference

- The simulated moments estimator is asymptotically normal for fixed S!! (This is not the case for SMLE.)

- The asymptotic distribution of b is given by

$$\sqrt{n}\left(\hat{b}-b\right) \xrightarrow{d} \mathcal{N}\left(0, \mathrm{avar}(\hat{b})\right)$$

in which

$$\mathrm{avar}(\hat{b}) \equiv \left(1+\frac{1}{S}\right)\left[\frac{\partial g_n\left(b\right)}{\partial b} W \frac{\partial g_n\left(b\right)}{\partial b'}\right]^{-1}.$$

Inference

- As in the case of plain vanilla GMM, one can perform a test of the overidentifying restrictions of the model

$$\frac{nS}{1+S}Q(b,n)$$

- This statistic converges in distribution to a χ^2 with degrees of freedom equal to the dimension of g_n minus the dimension of b.

Pseudo Code

```
function SMM (in double parameters[numberParameters],
              out double objectiveFunctionValue)

    call function solveTheModel( in double parameters[numberParameters],
              out double valueFunction[stateSpaceSize],
              out int policyFunction[stateSpaceSize])

    read weightMatrix[numberMoments, numberMoments]

    read dataMoments[numberMoments]

    call function simulateFirms( in double valueFunction[stateSpaceSize],
              in int policyFunction[stateSpaceSize],
              out double simulatedFirms[numberOfFirms, numberOfVariables])

    call function calculateMoments(in double simulatedFirms[numberOfFirms,
              numberOfVariables],
              out double SimulatedMoments[numberMoments])

    momentError = dataMoments − SimulatedMoments

    objectiveFunctionValue = momentError * weightMatrix * momentError
```

Identification

- The success of this procedure relies on picking moments h that can identify the structural parameters b.
- The conditions for global identification of a simulated moments estimator are similar to those for GMM.
  - The expected value of the difference between the simulated moments and the data moments equals zero iff the structural parameters equal their true values.
  - A sufficient condition for identification is a one-to-one mapping between the structural parameters and a subset of the data moments of the same dimension.

Identification

- The moments h, are informative about the structural parameters, b.
- That is, the sensitivity of h to b is high.
- Picking good moments is analogous to picking strong instruments in a standard IV estimation.

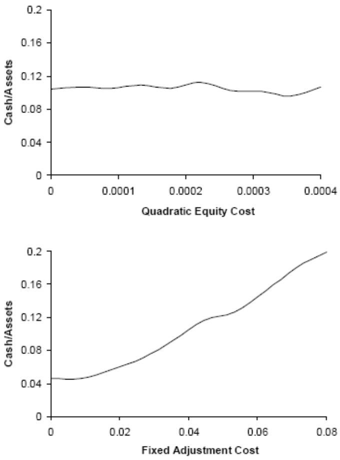
Identification

- How do you ensure that the model is identified?
- Check the standard errors:
  - The precision of the estimates, measured through the asymptotic variance above, is related to the sensitivity of the auxiliary parameters to movements in the structural parameters through $\partial h_b\left(y_{is}\left(b\right)\right)/\partial b$
  - If the sensitivity is low, the derivative will be near zero, which will produce a high variance for the structural estimates.

Use Economics

- **PLAY WITH YOUR MODEL UNTIL YOU UNDERSTAND HOW IT WORKS!!!!!!!!!**
- Do comparative statics: plot the simulated moments as functions of the parameters.
- You want to find steep, monotonic relationships.
- You want moments that move in different directions for different parameters.

Examples: Riddick and Whited (2009, JF)



What do reduced-form and structural folks even mean by identification?

- It's basically the same thing.
- But the means to the end is different.

More on identification

- Reduced form work:
  - You want exogenous variation in a variable of interest so that you can **interpret** a regression slope coefficient in a meaningful way.
  - You want the natural experiment to produce an estimate of a parameter that answers a well-posed *economic* question.
  - Even purely random variation need not be exogenous, so all reduced form identification requires (usually implicit) assumptions.
  - With structural you get identification via very explicit modeling assumptions.

How do you get identification?

- “Endogeneity” is not a curse word here. Structural estimation accounts for and exploits endogeneity within the model to get parameter estimates.
  - Davis, Fisher, Whited (2014) Do agglomeration externalities affect aggregate growth?
  - The externality induces an *endogenous* correlation between *predicted* land rents and TFP.
  - This correlation is zero in the absence of the externality, even though *realized* land rents and TFP are always correlated.
- Typically impossible to prove whether model is identified, just as it is typically impossible to test an exclusion restriction.

Do Not Construct a Black Box

- More parameters $\neq$ a better model!!!!!!
- Different features of the data should change when underlying parameters change.
- If the author cannot clearly explain which features of the data identify each parameter, the paper / job market candidate is a "reject"
- Structural estimation should not be a black box.

The question comes first — Not the model

- Before going structural, convince yourself that a structural approach is absolutely necessary.
- The answer will usually be whether you have serious data limitations or not.

See whether your estimation can uncover your parameters under ideal conditions.

- Simulate a “fake” dataset off the model
- Estimate the model, treating the fake data as if it were real data
- Does the estimator recover the true, known parameter values?
- Are the standard errors accurate?

You are going to have to minimize an objective function:

- You can't use a gradient based method unless you have a closed-form GMM or MLE problem
- Use the simulated annealing (SA) or differential evolution algorithm (DE) to avoid local minima
- DE is easier to parallelize, but it only gets close to the minimum. So you have to use Nelder Meade at the end to hunt for the bottom.
- Use the same seed for the random-number generator each time you simulate data off the model.

Software

- Do not use Matlab, R, Numpy, Octave, Gauss, or any other *interpreted* language.
- They are too slow!
- To estimate a model, you usually have to solve it $\sim$50,000 times.
- Use a compiled language: C, C++, Fortran
- Learn how to exploit multiple processors, a graphics card, a supercomputer, ...

Get the standard errors right.

- The actual data are usually not i.i.d.
- When estimating the covariance matrix for empirical moments, you must take into account
  - Heteroskedasticity
  - Time-series autocorrelation
  - Cross-sectional correlation
  - Serial correlation, including correlation across moments.
- I usually stack influence functions, and then covary them in a way that deals with these issues. The other method is to estimate the moments as a big fat GMM system.

WHY GO STRUCTURAL? BECAUSE YOU GET TO DO IT ALL!

- Write down models, solve models numerically, gather data, do complicated econometrics, . . .
- Going structural may be right for you if...
  - . . .not much on your calendar for next few years
  - . . .emotionally robust

Introduction

- Question 1: Why are so few CEOs fired every year (2%)?
- Question 2: Is this number large or small?
  - How much firing should we expect from a well functioning board?
- If a 2% firing rate is suboptimally low, how much shareholder value is being destroyed?

Introduction

- Why are these questions well suited to structural estimation?
- It is hard to answer "why" questions from reduced form regressions.

$$\begin{array}{ccc} \text{Proxy 1} & \Longleftrightarrow & \text{Hypothesis 1} \\ \text{Proxy 2} & \Longleftrightarrow & \text{Hypothesis 2} \\ & \vdots & \end{array}$$

- With structural estimation, you replace questionable proxies with modeling assumptions.
- Questions 2 and 3 require calculating a counterfactual: you can ask what happens if you change an estimated parameter.
- A counterfactual is a "what if" question. Sometimes you can do this with reduced form regressions, if you have extremely clever identification, but mostly you cannot.

Why do CEOs get Fired?

- Four potential reasons:

1. Turnover cost to shareholders may be large.
2. If the next best CEO is as good as the current one, why bother?
3. Boards may learn slowly about CEO ability.
4. CEO entrenchment

Why to CEOs get Fired?

- Great paragraph:

> *It is a challenge to measure the importance of these four potential reasons why CEOs are rarely fired. The board's firing choices are endogenous, which generates endogenous patterns in firm performance. There are no obvious instruments. Several elements are unobservable, including a CEO's actual and perceived ability, the CEO talent pool, the board's additional signals of CEO ability, and the board's personal turnover cost.*

Model assumptions:

In each period t:

- Board decides whether or not to fire CEO
- CEO quits / retires with probability $f(\tau)$
- Firm generates profitablity Y_t :

$$\underbrace{Y_t}_{\text{firm profitability}} = \underbrace{v_t}_{\text{industry profitability}} + \underbrace{y_t}_{\text{firm-specific}} - \underbrace{\mathbf{1}\{fire_t\}\,c^{(firm)}}_{\text{CEO turnover cost}}$$

$c^{(firm)}$ includes separation pay, executive search fees...

- Firm-specific profitability reverts around $\alpha =$ CEO's skill:

$$y_t = y_{t-1} + \phi\left(\alpha - y_{t-1}\right) + \epsilon_t$$

$\phi =$ persistence parameter $\qquad \epsilon_t \sim N(0, \sigma_\epsilon^2)$

Model assumptions: Learning

- New CEOs drawn from talent pool:

$$\alpha \sim N\left(\mu_0, \sigma_0^2\right)$$

- Board's prior beliefs:

$$\alpha \sim N\left(\mu_0, \sigma_0^2\right)$$

- Board uses Bayes' Rule to update beliefs about α each period
- Receives two signals about CEO skill, y_t and z_t

$$z_t \sim N\left(\alpha, \sigma_z^2\right)$$

Model assumptions: Board's preferences

$$\max_{\{fire_{t+s}\}_{s=0}^{\infty}} U_t \equiv E_t\left[\sum_{s=0}^{\infty}\beta^s u_{t+s}\right]$$

$$u_t = \kappa \underbrace{B_t Y_t}_{\text{profits}} - \underbrace{B_t \mathbf{1}\{fire_t\}\, c^{(pers)}}_{\text{pers. turn. costs}}$$

$$B_t = \text{firm assets}$$

$c^{(pers)}$ includes loss of CEO as ally, search effort...

Predictions summary

- Board optimally fires CEO as soon as posterior mean skill drops below endogenous threshold
- Why are CEOs rarely fired? Potential reasons:
  - Entrenchment (high $c^{(pers)}/\kappa$)
  - Costly to shareholders (high $c^{(firm)}$)
  - CEO skill does not matter much (low σ_0)
  - Slow learning (low σ_0, high σ_ε, low ϕ, high σ_z)
- Goal: Measure reasons' importance
- How: Estimate parameters
- Notice how the model gets at the questions via specific model features. This is hard to do and exactly what you should do!

SMM Estimation

- Data:
  - 981 CEOs who left office between 1971-2006
  - Successions classified as either forced or voluntary
  - Profitability during each year CEO in office
- SMM estimator:

$$
\begin{aligned}
\theta &\equiv \left\{\mu_0, \sigma_0, \sigma_z, \sigma_\epsilon, \phi, c^{(firm)}/\kappa, c^{(pers)}\right\} \\
\widehat{\theta} &\equiv \arg\min_{\theta} \left(\widehat{M} - \frac{1}{S}\sum_{s=1}^{S} \widehat{m}^s(\theta)\right)' W \left(\widehat{M} - \frac{1}{S}\sum_{s=1}^{S} \widehat{m}^s(\theta)\right)
\end{aligned}
$$

$\widehat{M}$ =14 empirical moments, $\widehat{m}^s(\theta)$ =14 simulated moments

- CEO firing rates
- Average profitability around firings
- Variance in profitability across CEOs

Parameter estimates

Firm turnover cost	$c^{(firm)}$	1.33
		(0.61)
Personal turnover cost	$c^{(pers)}/\kappa$	4.61
		(0.58)
Prior mean skill	μ_0	0.88
		(0.34)
Prior stdev. skill	σ_0	2.42
		(0.06)
Persistence parameter	ϕ	0.125
		(0.004)
Profitability stdev.	σ_ϵ	3.43
		(0.09)
Additional signal stdev.	σ_z	5.15
		(0.33)

Parameter estimates: Turnover costs

Firm turnover cost	$c^{(firm)}$	1.33
		(0.61)
Personal turnover cost	$c^{(pers)}/\kappa$	4.61
		(0.58)
Prior mean skill	μ_0	0.88
		(0.34)
Prior stdev. skill	σ_0	2.42
		(0.06)
Persistence parameter	ϕ	0.125
		(0.004)
Profitability stdev.	σ_ϵ	3.43
		(0.09)
Additional signal stdev.	σ_z	5.15
		(0.33)

Parameter estimates: Turnover costs

Firm turnover cost	$c^{(firm)}$	1.33
		(0.61)
Personal turnover cost	$c^{(pers)}/\kappa$	4.61
		(0.58)
Prior mean skill	μ_0	0.88
		(0.34)
Prior stdev. skill	σ_0	2.42
		(0.06)
Persistence parameter	ϕ	0.125
		(0.004)
Profitability stdev.	σ_ϵ	3.43
		(0.09)
Additional signal stdev.	σ_z	5.15
		(0.33)

- In dollars for median firm:

$$\begin{aligned} \text{Firm cost} &= \$57M \\ \text{Personal cost} &= \$197M \\ \text{Total cost} &= \$254M \end{aligned}$$

Parameter estimates: Turnover costs

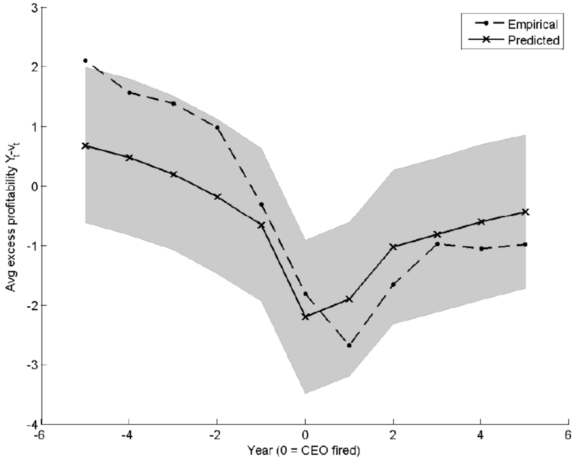
Firm turnover cost	$c^{(firm)}$	1.33
		(0.61)
Personal turnover cost	$c^{(pers)}/\kappa$	4.61
		(0.58)
Prior mean skill	μ_0	0.88
		(0.34)
Prior stdev. skill	σ_0	2.42
		(0.06)
Persistence parameter	ϕ	0.125
		(0.004)
Profitability stdev.	σ_ϵ	3.43
		(0.09)
Additional signal stdev.	σ_z	5.15
		(0.33)

- Not a 4.61% cost to directors
- Board is indifferent between firing CEO and seeing shareholders lose an extra 4.61% of assets
- Cannot determine whether
  - board has strong distaste for firing CEO (high $c^{(pers)}$)
  - board does not care about shareholder value (low κ)

Model fit: CEO turnover

	Empirical	Simulated
% of CEOs fired per year	2.29	2.16
% successions forced	17.1	16.2
Median spell length (years):		
Forced	4	4
Voluntary	7	7

Model fit: Profitability around CEO firings



Effect of entrenchment on shareholder value

	Baseline	Counter-factual
Personal turnover cost	4.6%	0.0%
% of CEOs fired per year	2	13
Mean profitability per year	15.5%	16.0%
Mean M/B	1.55	1.60

Conclusion: Great Paper!

- What is good about this paper?
  - It asks both a “why” and a “how much” question.
  - It uses the model estimates to conduct interesting counterfactual experiments.
  - The connection between theory and empirical work is very tight.

Conclusion: Great Paper!

- What is not so good about this paper? Not much!
  - The financing part of the model is nonexistent. With financing actions, it would not take as long to figure out if the CEO was bad. Higher costs needed to rationalize the 2% rate.
  - Production is CRS. With decreasing returns, CEO actions would have a dampened effect on profits, and it would take longer to figure out if the CEO was bad. Lower costs needed to rationalize the 2% rate.