

Structural Estimation

2010 FMA Meetings, New York

Toni M. Whited

October 2010

What is Structural Estimation?

Structural estimation is an attempt

- to fit economic models directly to data
- to assess the quality of the fit
- to identify parameters that govern technology, preferences, and (thus far in corporate finance) largely time-invariant institutional features.

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) → dynamic
 - Albuquerque and Schroth (2010, JFE) → static

What Kinds of Econometrics

- GMM
- MLE
- SMM
- SMLE
- Indirect Inference

What Kinds of Econometrics

- **GMM**
- MLE
- **SMM**
- SMLE
- **Indirect Inference**

What is Good about Structural Estimation

- “Better data and computing facilities, have made sensible things simple.¹”
- Good empirical work starts out with simple summary correlations.
- They provide a summary of what an economic model has to explain.

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

What is Good about Structural Estimation

- In empirical corporate finance the connection between an economic model and observed correlations can be tenuous.
- More often there exists no model that can explain the correlations, only vague stories.
- Recent advances in the econometrics of simulation estimators have made transparent the relationship between economic models and the equations used to estimate them.

What is Good about Structural Estimation

- 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.

What Kinds of Questions Can These Models Answer?

- Structural estimation tends to ask different types of questions than either reduced form regressions or quasi-experimental techniques.
- Reduced form techniques examine the sign of the magnitude of an effect of one variable on another.
 - Is higher cash flow associated with higher or lower leverage?
 - Does CEO succession within a family help or hurt firm performance?

What Kinds of Questions Can These Models Answer?

- Structural estimation is better at examining economic mechanisms.
 - How exactly do finance constraints affect corporate investment?
 - Why does leverage adjust slowly to target?
- It can identify the effects of unobservables
- Parameter estimates can be used to analyze counterfactuals

Closed-Form Structural Estimation: Whited (1992, JF)

- Notation

I_t/K_t	=	investment to capital ratio
MPK_{t+1}	=	marginal product of capital
a	=	adjustment cost parameter
d	=	depreciation rate
β	=	discount factor
e_{t+1}	=	expectational error

- Investment Euler equation

$$a \frac{I_t}{K_t} = \beta E \left[\left(MPK_{t+1} + (1-d) a \frac{I_{t+1}}{K_{t+1}} \right) \right]$$

$$a \frac{I_t}{K_t} = \beta \left(MPK_{t+1} + (1-d) a \frac{I_{t+1}}{K_{t+1}} \right) + e_{t+1}$$

Closed-Form Structural Example

- Pro: very close connection between theory and what you want to measure.
- Con: the model is very simple. The parameter a can capture all sorts of frictions besides adjustment costs.
- With a more complicated model we can deal with this problem, but we cannot figure out a closed form equation to estimate.
- This is where simulation estimators come in.

Economic Model

- Before you can use simulation estimators, you need an economic model that has a numerical solution.
- The solution is a rule.
- If I am in state “A” and get hit with shock “z,” I move to state “B”.
- You can put the rules in Excel sheets and send them to your coauthors.
- You can use a random number generator to generate shocks, and then get simulated data.
- Use a seeded random number generator!!!!

Setup

- Let x_i be an *i.i.d.* data vector, $i = 1, \dots, n$.
- Let $y_{is}(b)$ be an *i.i.d.* simulated vector from simulation s , $i = 1, \dots, N$, and $s = 1, \dots, S$.
- 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 many applications one can calculate the weight matrix as the inverse of the variance covariance matrix of $h(x_i)$.
- This weight matrix has an exact analogy with GMM.
- This makes SMM easier than you would imagine. You calculate the weight matrix using real data, and you calculate it only once!

Inference

- The asymptotic distribution of b is given by

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

in which

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

- 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 .

Identification

- The success of this procedure relies on picking moments h that can identify the structural parameters b .
- The sensitivity of the moments to the parameters is high.
- Picking good moments is analogous to picking good instruments in a standard IV estimation.

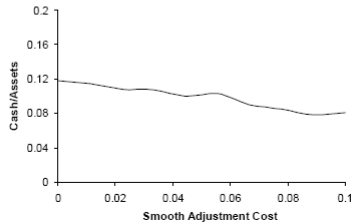
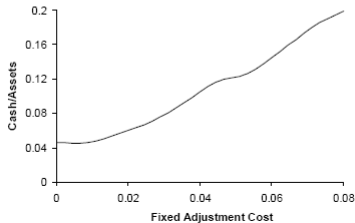
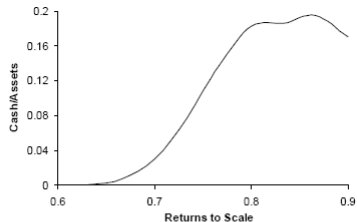
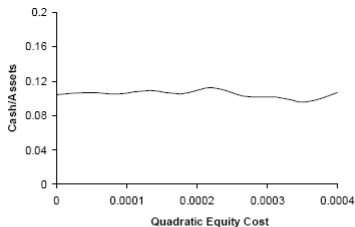
Identification

- How do you ensure that the model is identified?
- Check the standard errors:
 - The precision of the estimates is related to the sensitivity of the moments to movements in the structural parameters through $\partial h_b (y_{i.s} (b)) / \partial b$
 - 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.

Examples²

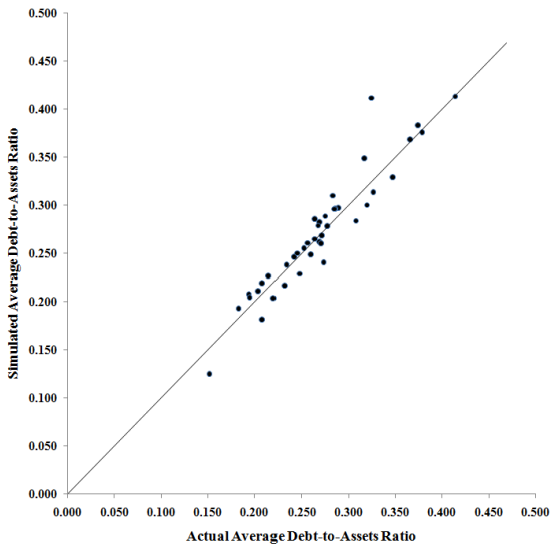


²Riddick and Whited (2009, JF)

Firm Heterogeneity

- It is really hard to address the issue of firm heterogeneity using SMM.
- This is perhaps the biggest drawback of this technique.
- The models we simulate are usually of a single firm.
- Heterogeneity comes from random draws of shocks.
- So you have to suck as much heterogeneity out of your data as you can before you can have any hope of fitting the model to the data.
 - Firm and time fixed effects
- You can also do sample splits. This used to be computationally infeasible, but . . .

Sample Splits³



³DeAngelo, DeAngelo, and Whited (2009)

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.

Proxy 1	\iff	Hypothesis 1
Proxy 2	\iff	Hypothesis 2
	⋮	

- 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 to 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. [...] evaluating their magnitudes requires estimating or calibrating an economic model.

Model assumptions:

In each period t :

- Board decides whether or not to fire CEO
- CEO quits / retires with probability $f(\tau)$
- Firm generates profitability 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 $\alpha = \text{CEO's skill}$:

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

$$\phi = \text{persistence parameter} \quad \epsilon_t \sim N(0, \sigma_\epsilon^2)$$

Model assumptions: Learning

- New CEOs drawn from talent pool:

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

- Board's prior beliefs:

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

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

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

Model assumptions: Board's preferences

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

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

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

$c^{(pers)}$ 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:

$$\theta \equiv \left\{ \mu_0, \sigma_0, \sigma_z, \sigma_\epsilon, \phi, c^{(firm)}/\kappa, c^{(pers)} \right\}$$

$$\hat{\theta} \equiv \arg \min_{\theta} \left(\widehat{M} - \frac{1}{S} \sum_{s=1}^S \widehat{m}^s(\theta) \right)' W \left(\widehat{M} - \frac{1}{S} \sum_{s=1}^S \widehat{m}^s(\theta) \right)$$

\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:

Firm cost	=	\$57M
Personal cost	=	\$197M
Total cost	=	\$254M

Parameter estimates: Turnover costs

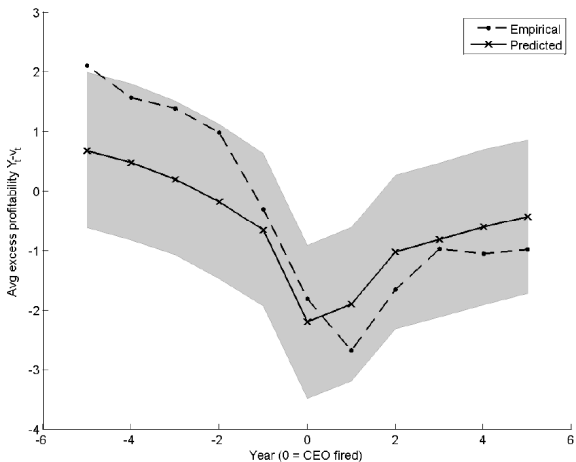
Firm turnover cost	$c^{(firm)}$	1.33 (0.61)
Personal turnover cost	$c^{(pers)}/\kappa$	4.61 (0.58)
Prior mean skill	μ_0	0.88 (0.34)
Prior stdev. skill	σ_0	2.42 (0.06)
Persistence parameter	ϕ	0.125 (0.004)
Profitability stdev.	σ_ϵ	3.43 (0.09)
Additional signal stdev.	σ_z	5.15 (0.33)

- 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.
- 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.

Economic Models and Auxiliary Models

- Sometimes models can produce likelihood functions for the simulated data.
- Usually this is not the case or it is computationally intensive.
- You can still use what is called an **auxiliary model** to estimate the model parameters.

Economic Models and Auxiliary Models

- A likelihood is a description of the true data generating process.
- An auxiliary model is an approximation of the true data generating process.
 - DSGE model of the macroeconomy. The likelihood is impossible to solve for, but a VAR might describe the data approximately.
 - Microeconomic model of investment spikes or infrequent price adjustments. The auxiliary model could be a duration model.

Auxiliary Models

- The auxiliary model is itself characterized by a set of parameters.
- These parameters can be estimated using either the observed data or the simulated data.
- Indirect inference chooses the parameters of the underlying economic model so that these two sets of estimates of the parameters of the auxiliary model are as close as possible.
- You should be able to match exactly if you have as many parameters in the auxiliary model as you do in the economic model.
- But the number of auxiliary parameters can be greater than the number of economic parameters.

Estimation (GMM style)

- This is really just a special case of SMM.
- The parameters of the auxiliary model are the “moments.”
- There! We are done with this!

Estimation (ML style)

- All estimations require maximizing or minimizing some function of the data.
- What if the criterion function you maximize is a likelihood function.
- Let $\phi(x_i | \theta)$ be the density of x_i .
- Then you can estimate b by minimizing the following:

$$\hat{b} = \arg \min_b \left(\sum_{i=1}^N \ln \phi(x_i | \theta) - \sum_{i=1}^N \ln \phi(x_i | \theta(b)) \right)$$

- Note that the second term contains the parameters estimated with **simulated** data plugged into a likelihood function formed with the **real** data.

Estimation (ML style)

- Equivalently, because the first term does not depend on b :

$$\hat{b} = \arg \max_b \left(\sum_{i=1}^N \ln \phi(x_i | \theta(b)) \right)$$

- Inference is as for standard ML except that you always have the $\left(1 + \frac{1}{S}\right)$ term in front of variance estimates.
- This approach can be viewed as maximizing the approximate likelihood subject to the restrictions that the economic model imposes on the parameters of the auxiliary model.

Estimation (LM style)

- When you are doing standard ML, you maximize by setting the first derivative—the score—equal to zero.
- You can do the same with indirect inference.
- This works exactly backwards from ML style estimation:
 - Estimate θ in the real data.
 - Plug this estimate into the score based on the simulated data.
 - Set this score as close to zero as possible.

Estimation (LM style)

- Here is exactly what you do:

$$\hat{b} = \arg \min_b \Phi(b)' V \Phi(b)$$

in which

$$\Phi(b) = \sum_{s=1}^S \sum_{i=1}^N \frac{\partial}{\partial \theta} \ln \phi \left(x_i^s(b) \mid \hat{\theta} \right)$$

in which $\hat{\theta}$ sets the score in the observed data to zero.

- V is a positive definite weight matrix. What is the optimal matrix?

Comparison

- The three approaches are asymptotically equivalent under correct specification of the auxiliary model.
- (Nice theoretical result, but we use an incorrectly specified auxiliary model for a reason.)
- They are numerically identical if θ and b are of the same dimension.

Comparison

- If you are being agnostic on distributions and have a semiparametric auxiliary model (like a regression), then GMM-style is the only feasible option.
- If you know the likelihood of the auxiliary model, the two maximum likelihood methods are clearly more efficient.
- If you are guessing at the likelihood, then the tradeoffs between GMM-style and ML-style indirect inference are standard.

What is NOT Structural Estimation

- Running regressions on simulated data.
- Calibration is not structural estimation.
- Testing directional predictions off of dynamic models is not structural estimation.
- Using dynamic panel estimators is not structural estimation (mostly).

An Advertisement

- I hope I have piqued your interest in this stuff.
- Why should you do it.
 - You learn a lot. **a lot!!!!**
 - It is really fun when you finally get it to work.
 - It is orders of magnitude faster than it used to be.
 - It heats your office in the winter.