The Art of Structural Modelling

Michael Keane
University of New South Wales

Summer School in Structural Dynamic Modelling
University of Chicago
July 12, 2019

What is this talk about ?

- This talk deals with the “Art” of structural modelling
- What do I mean by that?
- By analogy, it takes a great deal of technical skill just to play Beethoven’s late piano sonatas (badly)
- But it takes “artistry” to play them <u>well</u>
- So I will not be talking about “techniques,” except in a very high level sense

Structural Model Development

- Structural estimation has several key stages:
- 1) Theoretical Model Development
- 2) Practical Specification Issues
- 3) Solving the Model
- 4) Understanding How the Model Works
- 5) Estimation
- 6) Validation
- 7) Policy Experiments

(1) Theoretical Model Development

- Assume you want to build a structural model to address an ***economic or policy question of interest***:
- A good starting point is to ask what ingredients your model must have to credibly address the question:
- The model cannot be so simple (or stylized) that it essentially imposes answers *a priori*
- It must be plausible to assume parameters of the model are invariant (or at least not too sensitive) to policy experiments that you plan to do

(1) Theoretical Model Development

- For example, say the question is "How important is the moral hazard effect of health insurance?"
- Then your model better have private information about health, and allow for adverse selection.
- Otherwise, you rig the results to finding that moral hazard is important:
- It is the only thing that the model can use to explain the correlation between health spending and insurance

(1) Theoretical Model Development

- Say the question is "Why have married women's <u>relative</u> wages increased?" (in the last 50 years)
- You will need to account for:
  - Sorting of who gets married
  - Human capital investment by married women
  - Work decisions of married women
  - Fertility
  - Divorce
- And perhaps other factors/mechanisms as well

(1) Theoretical Model Development

- How to develop a good model is very difficult to teach!!
- Fortunately, some other aspects of structural estimation are easier to teach ….

(2) Practical Specification Issues

- There is an inherent conflict between:
  - A) A Model that is rich enough to credibly address the question of interest
  - B) A Model that it feasible to solve and estimate
- As we make models richer (i.e., incorporate more key mechanisms or features), we tend to get more <u>state variables</u>
- Too many state variables makes solution infeasible

(2) Practical Specification Issues

- The Art of Structural Estimation is largely about how to develop “rich” or “realistic” models that are still feasible to solve and estimate

- How to capture complex behaviors with as few state variables as possible?

(2) Practical Specification Issues

- A very good example of this problem arises in modelling labor supply and fertility.
- If the disutility of work depends on number and ages of children, then the state is a vector that lists the number of children of each age
- This means literally trillions of possible state points, a problem that is not practical to deal with

Example: Married Person Value Function

$$V_t^{jM}(\Omega_{jt}) = \frac{1}{\alpha}(\psi C_t^M)^\alpha + L(l_t^j) + \theta_t \; + \pi_t^M p_t + A_j^M Q(l_t^f, l_t^m, Y_t^M, N_t)$$

Value of Leisure

$+\delta EV(\Omega_{j,t+1})$ where *j=m,f*

From Eckstein, Keane, Lifshitz (Econometrica, 2019) where:

- Value of Leisure = L$(l_t^j) = \frac{\beta_{jt}}{\gamma}(l_t^j)^\gamma + \boldsymbol{\mu_{jt}} l_t^j$

- β_{jt} - tastes for leisure – "deterministic part" (depends on health, education)
- $\boldsymbol{\mu_{jt}}$- tastes for leisure – "stochastic" part – This is an AR(1) process that is shocked upwards by the arrival of a new born ($\boldsymbol{p_{t-1}}$)

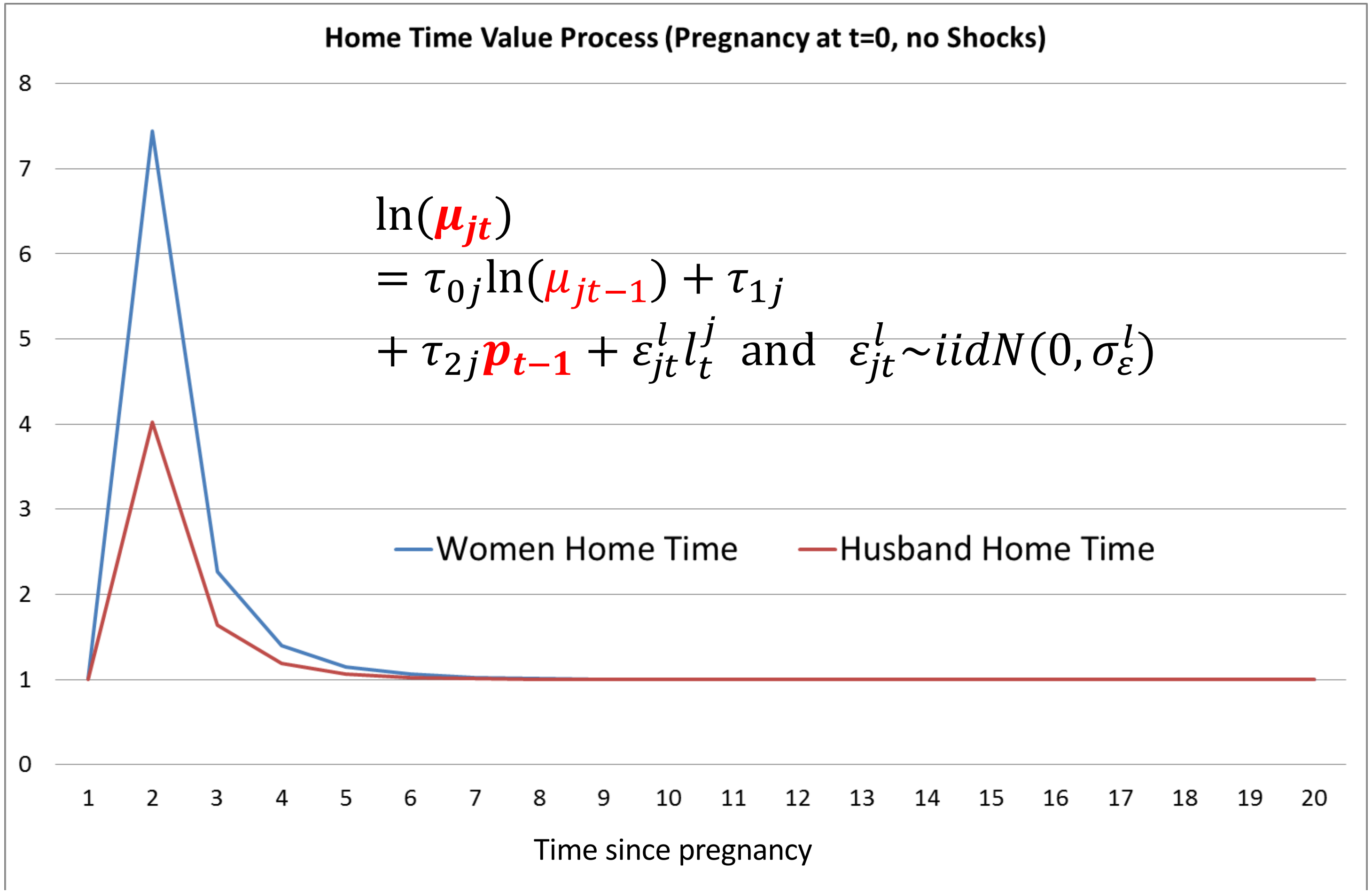
- $\ln(\boldsymbol{\mu_{jt}}) = \tau_{0j}\ln(\mu_{j,t-1}) + \tau_{1j} + \tau_{2j}\boldsymbol{p_{t-1}} + \varepsilon_{jt}^l l_t^j$ and $\varepsilon_{jt}^l \sim iid N(0, \sigma_\varepsilon^l)$
- It then decays towards steady state until another baby arrives

How Value of Leisure is Affected by Child

$$\mathrm{L}(l_t^j) = \frac{\beta_{jt}}{\gamma}(l_t^j)^{\gamma} + \boldsymbol{\mu_{jt}} l_t^j \qquad \beta_{jt} = \beta_{j1} E_{jt} + \beta_{j2} H_{jt} + \beta_{j3} P_t$$

Female: $\beta_{ft} = 0.01 E_{jt} + 0.026 H_{jt} + 0.059 P_t$

Male: $\beta_{mt} = 0.00 E_{jt} + 0.033 H_{jtt}$



$$\ln(\boldsymbol{\mu_{jt}}) = \tau_{0j} \ln(\mu_{jt-1}) + \tau_{1j} + \tau_{2j} \boldsymbol{p_{t-1}} + \varepsilon_{jt}^l l_t^j \text{ and } \varepsilon_{jt}^l \sim iidN(0, \sigma_\varepsilon^l)$$

(2) Practical Specification Issues

- Using this modelling device, the effect of children on the utility of leisure is captured by a single state variable, $\boldsymbol{\mu_{jt}}$
- This reduces the computational burden of solving a model with fertility by orders of magnitude
- Thinking of clever ways to reduce the state space like this is more valuable than all the books on numerical analysis you can ever read

(2) Practical Specification Issues

- Some other examples of reducing state space:
- Erdem, Imai, Keane (QME, 2003) dynamic demand paper: Using "quality weighted inventory" to avoid keeping track of inventory of all brands.
- Iskhakov-Keane (2019) paper on age pension:
  – Assume super balance is proportional to human capital to avoid having to track multiple asset classes
  – Model pension and super payouts as functions of state variables already in the model. This is a simplification relative to the actual rules, which involve additional state variables (were we to model them in full detail). Estimate the simple functions from data in first stage.

A Useful "Trick" when the state space is Large

- Value functions can almost always be written as:
- $V(s) = a(i)+b(j)c(k)d(j,k)$
- Where s indexes all state points, and *i*, *j*, *k* index subsets of the whole state space.
- E.g., *i* could index consumption levels, *j* could index work experience levels, *k* could index education levels.
- Typically, construction of a, b, c, d functions involves complex calculations, but these only need be done at a small subset of the state space.
- Always calculate a, b, c, d separately, save them, and put them together at the end.
- Then you only need to do addition/multiplication operations at all s=1,…,S

A Useful "Trick" when the state space is Large

- You should only loop over *i*, *j*, *k* once, at the very end:
- Do *i*=1,..,I
- Do *j* =1,..,J
- Do *k* = 1,…,K
- *s* = S(*i*,*j*,*k*) Mapping function from i,j,k to s
- V(*s*) = a(*i*)+b(*j*)c(*k*)d(*j*,*k*) Table look ups, add/multiply
- Continue
- Continue
- Continue

3) Solving the Model
4) Understanding How the Model Works

- Let’s assume you have settled on a model that you think is fit for purpose
- Next I want to talk about computational and programming issues at a high level
- As far as I am aware these things are hardly ever talked about
- But they are extremely important

3) Solving the Model
4) Understanding How the Model Works

- It is crucial that steps (3) and (4) be done in tandem
- You should be writing programs to:
  - (i) solve the model
  - (ii) simulate the model
  - (iii) calculate descriptive statistics

  AT THE SAME TIME
- This is the only way to make sure the behavior of the model makes sense

3) Solving the Model
4) Understanding How the Model Works

- Always begin by programming a simple special case of the model – where theory makes a clear prediction of how it should behave
- E.g., in a life-cycle model start with consumption as the only choice (and no borrowing constraints).
- Make sure that if $\rho=1/(1+r)$ then consumption is flat over the life-cycle.
- If your simulations don't line up with basic theory you've done something wrong

3) Solving the Model
4) Understanding How the Model Works

- This "start small" approach may seem obvious
- But my experience is that most people try to program up their full-blown model right from the start
- This is a big mistake, both from a programming point of view and with respect to <u>developing economic intuition</u>

3) Solving the Model
4) Understanding How the Model Works

- You should add mechanisms or features to the model one at a time
- Always rig the program so if a parameter θ is set to zero the new mechanism is shut down
- Make sure you get the EXACT same simulation results as before if you set θ = 0
- If not, then you have introduced a bug

3) Solving the Model
4) Understanding How the Model Works

- Once you have introduced a new mechanism, manipulate the new parameters related to that mechanism to see what they do
- Make sure the simulation results make intuitive sense (Spend plenty of time on this!)
- If they don't make sense, you probably have a bug
- *Sometimes*, basic intuition turns out to be wrong and you have <u>learned some economics</u>

3) Solving the Model
4) Understanding How the Model Works

- Continue introducing new mechanisms step-by-step until you have your full model
- Make sure to check carefully for bugs at each step
- Make sure to check that the simulation results continue to make intuitive sense at each step
- That is, does each new feature affect behavior in a way that makes economic sense?

3) And 4) Solving the Model

- By the end of this process, you will have a good understanding of how each feature of the model (and each parameter) affects behavior
- I often spend more time manipulating the model to make sure I understand how it works than I spend actually estimating it
- Spending a lot of time learning what each parameter does to behavior also gives you a good intuitive understanding of how the model is "identified" (in the sense of what data patterns drive what parameters)

5) Estimation

- Nonlinear estimation is difficult !!
- You can’t just start the parameter vector at some random place and expect it to converge
- You need to calibrate the model to achieve a half way decent fit before iterating
- This also teaches you a lot about how the model works
- But you should expect that this calibration process will take a long time and lots of patience

5) Estimation

- If you find it impossible to get a decent calibration by hand, it means:
  - there is some important mechanism you have omitted from the model
  - You don’t really understand the model
  - You aren’t trying hard enough!

5) Estimation

- Now that you are ready to estimate you need to write the code for the parameter search algorithm
- I like to use BHHH and the Simplex
- More bugs may be introduced here!!
- So I always start by trying to iterate on just a few parameters at a time to make sure the code is working as expected
- Only build up gradually to the full model as you are sure everything is working properly
- Remember: Bugs or ID problems can appear for some parameters and not others

5) Estimation

- Estimation is *not a mechanical process*
- As the parameters iterate, I look at:
  - Simulations of key statistics vs. actual data
  - Values of parameters where I have some reasonable prior
- Two types of things often go wrong:
  - Some parameters go to strange values
  - Key moments don’t improve or even get worse
- These are often symptoms of bugs in the code or a flaw in the model (e.g. identification problem)

5) Estimation

- Problems in estimation are very frustrating because it can be hard to pin down the source:
  - Bug in estimation code
  - Bug in solving the model that you didn't find in steps 3-4
  - A flaw in the model that you didn't find in steps 3-4
- **You almost never find bugs by reading the code!!!**
- I usually proceed by doing two things:
  - shutting down parts of the model to figure which part is causing the problem
  - Printing out lots of stuff and checking if it makes sense (e.g., do value functions go up when they should?)

5) Estimation

- Let's say you are confident the code is working fine but you are not happy with the fit of the model, or some parameters seem very odd
- You need to think more about the economics of the situation and figure out what important mechanism(s) you left out
- Once you have a good idea what may be wrong, add the new mechanism and go back to step (3)
- Unfortunately this may take 3 months!

6) Validation

- Let’s say the parameters of the model have converged to sensible looking values and the in sample fit looks OK
- At this point most people are pretty exhausted so you probably need to take a vacation
- Maybe write a few IV papers for fun!!
- Once you have recovered your stamina, we get to what may be the most important parts, validation and policy experiments

6) Validation

- Unfortunately, most structural papers don't do much in the way of model validation
- One reason is that the data needed to do validation is often not easily available
- A good opportunity for validation is when an experiment has been run, and you can estimate the model on the "control" data and see if you can forecast the "treatment" data
- Another example is when there are major policy changes (e.g., welfare reform)
- You should think about whether you can validate the model <u>before</u> you estimate it

7) Policy Experiments

- One of the major reasons we do structural estimation is because we can use structural models to do policy experiments
- We may be interested in predicting effects of:
  - Proposed policies
  - Hypothetical policies
- Or we may want to use the model to optimize the parameters of a policy to maximize some objective
- (Note: Experiments only allow us to see effects of policies that have already been implemented)

7) Policy Experiments

- Unfortunately, a common flaw of structural papers is they do a lot of work to solve and estimate the model
- But when that is done they don't report any interesting experiments
- That is why I said in the beginning "A good starting point is to ask what ingredients your model must have to credibly address the economic or policy question of interest"
- You should have some interesting experiments in mind before you even start!!

Conclusion: Why Do Structural Estimation?

- Aside from wanting to do policy experiments, another reason to do structural estimation is because you are interested in a model itself:
- If we view the data through the lens of a particular model, what does it imply about behavior in various situations?
- Given the theoretical restrictions of a particular model, is it even possible to get a good fit to the data, or to do well in out of sample validation exercises?

Why Do Structural Estimation?

- In my view, structural econometrics is a long term program to develop models in a range of areas
- Models that we are confident in using for policy evaluation
- For example, in the long run we may hope to reach a consensus on a "canonical" or "standard" life-cycle labor supply model
- This model will have passed a large number of validation exercises over many years
- It will be standard practice to evaluate tax and welfare policy proposals using this model

Why Do Structural Estimation?

- So, in my view, structural econometrics is a long term program to develop “standard” models in a range of areas
- Obviously, we are a long way from this ideal situation
- But this is the goal we must be striving for if we want to do economics like a real science