ECON 815 - Computational Methods for Economists

Fall 2017

Notes on Simulated Method of Moments Estimators

Simulated Method of Moments

- General idea:
  - Estimate parameters of a structural model by simulating the model and comparing the model "data" to actual data
  - Can do this with a method of moments type methodology- pick some key characteristics of the data (moments) and pick model parameters to create model moments that match these data moments as closely as possible
    * Moments will be functions of endogenous variables (e.g., mean or std deviation of consumption)
  - Application of GMM without closed form solutions to the moments
- Sometimes called Method of Simulated Moments (MSM) or Indirect Inference
- Seminal paper is McFadden (1989, *Econometrica*)
- SMM estimator:

$$\hat{\theta}_{S,T}(W) = argmin_{\theta} \left[\sum_{t=1}^{T} \left(\mu(x_t) - \frac{1}{S} \sum_{s=1}^{S} \mu(x(u_t^s, \theta)) \right) \right]' W_T^{-1}$$
$$\left[\sum_{t=1}^{T} \left(\mu(x_t) - \frac{1}{S} \sum_{s=1}^{S} \mu(x(u_t^s, \theta)) \right) \right]$$

  - Where:
  - x_t are a sequence of observed data
  - $\mu(x_t)$ is a vector of moments from the data (e.g., the sample mean of x, the sample std dev of x)
  - $\mu(x(u_t^s, \theta)$ is a vector of corresponding moments from the S simulations of T observations (note the model moments depend upon the vector of parameters, θ)
  - u_t^s are the $S \times T$ shocks used to simulate the model
  - W_T is the weighting matrix, which may be a function of T, the number of sample observations (we'll talk about weighting below)
- Compared to GMM:
  - Not necessary to find closed form moments condition that will need to hold
  - Instead, matching characteristics of data with model simulations
  - But similar in trying to minimize distance between model and data moments
  - Much more computationally intensive then GMM
    * Will need to solve model for each guess of parameters
    * May need to iterate 1000s of times to find parameters that minimize distance between model and data
- Benefits of SMM:
  - Easy interpretation of moments and whether they are important

- Already simulating model, so ready to run counterfactuals
- Know that model matches key features of data
- Do not need continuous functions (e.g. FOCs) like with GMM
  * Makes ideal for discrete choice models
  * Good for optimal stopping problems
    · MLE good for discrete choice, but if noise in data it may be hard to define "inaction"
    · Anyway, MLE in this context (with no analytical solutions to decision rules thus needing to simulate model) can be thought of as a form of SMM
- Can often do with aggregate data
  * A real benefit when micro data is confidential
- Allows a researcher interested in a structural model to link results explicitly to existing reduced form empirical evidence

• SMM algorithm:

  1. Guess the vector of parameters
  2. Solve the model to generate the decision rules of agents
  3. Use the decision rules to simulate the model
  4. Calculate moments from the simulated model
  5. Compare the moments from the model simulations to those from the data (weighting the distance in some way)
  6. Update the parameter vector based on the above
  7. Repeat until the data and model moments are as close as possible

• Note:

  - SMM often done in 2 steps
  - In first, estimate/calibrate some parameters that don't require structural estimation
  - In second, estimate remaining parameters through SMM
  - This saves time and reduces the number of moment conditions needed

• Identification/Choosing moments:

  - Need at least as many moments as parameters (order condition) to identify model
  - Need to have parameters have differential impact on moments (rank condition) to identify model
  - Thus want to pick moments that respond to parameters and respond to different parameters in different ways
  - Want to pick moments that are important for the economic question/ widely used in the literature
    * It's also nice if they relate to the work of others and/or matter for key questions in the literature

• Std. Errors and weighting matrices

  - As with GMM, weighting matrix is redundant with exactly identified model - should be able to get distance between model and data moments to zero
  - With an over-identified model, weighting matters
  - Identity matrix is consistent, but inefficient
    * Units affect weight on moments
  - Optimal weight matrix

* Two ways to compute the optimal weighting matrix:
  1. Weight moments more heavily if they are better identified
     · Thus, use the inverse of the variance-covariance matrix of the data moments
     · The reference is Gourieroux, Monfort, and Renault (1993, *Journal of Applied Econometrics*)
     · Calculate by bootstrapping the data - calculate the moments N times, then use these N obs of the moments and calculate the covariance between them:
     · $\hat{W}_T^* = \frac{1}{T}\sum_{t=1}^{T}\left(\mu(x_t) - \frac{1}{N}\sum_{n=1}^{N}\mu(x_{n,t})\right)'\left(\mu(x_t) - \frac{1}{N}\sum_{n=1}^{N}\mu(x_{n,t})\right)$
  2. Simulate the moments using independent distributions of shocks.
     · Here,

$$\begin{aligned}\hat{W}_T^* =& \frac{1}{T}\sum_{t=1}^{T}\left[\mu(x_t) - \frac{1}{S}\sum_{s=1}^{S}\mu(x(u_t^s,\hat{\theta}_{S,T}))\right]\left[\mu(x_t) - \frac{1}{S}\sum_{s=1}^{S}\mu(x(u_t^s,\hat{\theta}_{S,T}))\right]' + \\ & \frac{1}{S}\frac{1}{T}\sum_{s=1}^{S}\sum_{t=1}^{T}\left[\mu(x(u_t^s,\hat{\theta}_{S,T})) - \frac{1}{L}\sum_{l=1}^{L}\mu(x(u_t^l,\hat{\theta}_{S,T}))\right]\left[\mu(x(u_t^s,\hat{\theta}_{S,T})) - \frac{1}{L}\sum_{l=1}^{L}\mu(x(u_t^l,\hat{\theta}_{S,T}))\right]'\end{aligned}$$

     · where u_t^s and u_t^l independent draws of the underlying shocks used to simulate the model

– Standard Errors:
  * Variance covariance matrix for parameter estimates is given by:

$$Q_S(W^*) = \left(1 + \frac{1}{S}\right)\left[\frac{\partial\mu(\hat{\theta}_{S,T})}{\partial\theta}' W^{*-1}\frac{\partial\mu(\hat{\theta}_{S,T})}{\partial\theta}\right]^{-1}$$

  * Where $\frac{\partial\mu(\hat{\theta}_{S,T})}{\partial\theta}$ is the derivative of the vector of model moments with respect to the parameter vector (so this will be a $q \times p$ matrix for q moments and p parameters).
  * Calculate the derivatives numerically - moving θ just a bit and calculating the new vector of moments
  * The std errors will be the square roots of the diagonal elements of $Q_S(W)$

• Overidentification tests
  – Hansen's J-test
  – Same as with GMM:

$$\xi_T = \frac{TS}{1+S}\left[\sum_{t=1}^{T}\left(\mu(x_t) - \frac{1}{S}\sum_{s=1}^{S}\mu(x(u_t^s,\hat{\theta}_{S,T}))\right)\right]' W^{*-1}\left[\sum_{t=1}^{T}\left(\mu(x_t) - \frac{1}{S}\sum_{s=1}^{S}\mu(x(u_t^s,\hat{\theta}_{S,T}))\right)\right]$$

  – $\xi_T \sim \chi^2(q-p)$
  – where q is the number of moments, and p is the number of parameters

• How many simulations to calculate model moments?
  – More can make model moments more precise, but no big diff usually between 2000 and 10000
  – Need to be careful of:
    1. Initial values. Be sure to not use first 500 or so simulations when calculating moments so that initial values don't affect results
    2. Random #s. Use same random number draws when doing your simulations, else simulated moments will be affected by draws of numbers.

- Alternatively, you can find the stationary distribution
    * This involves solving the model then using another fixed point algorithm to solve for the fixed point of the stationary distribution
    * But the stationary distribution will be the same for a given set of parameter values - no uncertainty like simulations
    * What does this mean for std errors? Are we really just calibrating at this point since parameter values imply the model moments with no uncertainty?

- Implementation:
  - Find data moments (likely calculate these yourself using microdata, but may also get from others)
  - Make a guess at model parameters
  - Set up model solution (e.g. VFI for a dynamic programming problem)
  - Simulate model/find stationary distribution
  - Calculate model moments
  - Find distance between data and model moments (using optimal weight matrix)
  - Update guess at parameter vector using `fminsearch` or simulated annealing algorithm (or other optimization routine)
  - Continue until minimize distance between data and model moments

- Tips:
  - Use a global optimizer like Simulated Annealing

SMM: Example 1, Cooper and Ejarque, "Financial frictions and investment: requiem in Q" (*Review of Economic Dynamics*, 2003):

- Question: To what extent does market power explain firm investment behavior?
- Theory: Q-theory predicts that firm investment will only be a function of the marginal value of the firm (marginal Q)
  - $\frac{I}{K}_{it} = a_{i0} + a_1 E\bar{q}_{it+1} + a_2 \left(\frac{\pi_{it}}{K_{it}}\right)$
- Observation: Firm investment rates seem to be explained to a significant degree by firm profitability/cash flow
- A natural thought is that this is a sign of costly external finance
- Cooper and Ejarque ask to what extent this is driven by not costly finance, but market power by firms
  - Marginal Q is unobserved and is proxies for by Avg Q (and in theory, Marginal Q = Avg Q if adjustment costs are quadratic and perfect competition, no financial frictions)
  - Market power means Marg Q $\neq$ Avg Q - so measurement error in most Q-regressions
  - This error is correlated with firm profitability - further from avg Q as market power and thus profits increase
- The answer has important implications for how fiscal and monetary policies affect firm investment behavior
  - Monetary policy operates by affecting the cost of capital through fluctuations in the interest rate.
  - Bernanke and Gertler (1995) argue that firm investment not responsive to cost of capital fluctuations.

- Fiscal policy affects both the cost of capital and firm cash flow.
- If financial frictions are large, the cash flow channel may be especially important.

- Problem: Need a structural model to estimate these "deep" parameters of the model (parameters measuring market power, financing costs, etc)
- GMM on Euler equations not viable because external financing costs may be fixed costs - so there is a discontinuity if go from no access to capital markets to some access
- Solution: SMM
- Basic Model:
  - Firm capital accumulation model:
  - Bellman is: $V(K, A) = \max_{K'} \pi(K, A) - p(K' - K(1-\delta)) - C(K', K) + \beta E_{A'|A} V(K', A')$
    - Where K is the capital stock, A is the profitability shock
    - π is the firm's per period profit function
    - C is the cost of adjusting capital function
    - β is the discount factor, p is the price of capital
  - Impose costs to external finance
  - Extensions: Fixed costs in changing capital stock
- Identification:
  - β is set to match the the real, risk free interest rate (in equilibrium, $1 + r = \frac{1}{\beta}$) - this is calibration
  - In baseline case, estimate 5 parameters: curvature of profit function (returns to scale), quadratic adjustment cost, persistence of profit shocks, std dev of shock to profits, external finance cost
  - Moments:
    - a_1 and a_2 from Q-regression above (tells you about adjust costs, costs to external finance)
    - Serial correlation of investment rate (tells you about quadratic cost)
    - std dev of profit to capital ratio (tells you about std dev of profit shock)
    - Avg Q (tells you about returns to scale)
    - Fraction of firms accessing external finance (tells you about fixed cost to external finance)
  - Model is overidentified
  - J-stat rejects over identifying restrictions, but use to compare models
- Data:
  - Firm level data on investment, capital stock, profits - from Compustat
- Results:
  - Once include market power, model has no better fit if allow for costly external finance
  - Therefore disconnect between Q-theory and evidence not due to financial frictions
  - No link between financial frictions and lumpy investment patterns
  - Implications for policy - not really sensitive to cash flow, still marginal Q, we just can't proxy for marginal Q by avg Q (so monetary policy should still work well)

SMM: Example 2, Adda, Dustmann, Meghir, and Robin, "Career Progression and Formal versus On-the-Job Training", Unpublished, 2012):

- Question: What is the impact of vocational training on labor market outcomes (wages, unemployment)?
- Problem with regression analysis:
  - Entering the vocational training track is endogenous
  - Choices have dynamic effects - e.g. choice to undertake vocational training affects entire lifecycle wage profile - and those expected wages later affect educ decisions today
  - Need to find the deep parameters (policy invariant parameters) in order to conduct counterfactual simulations (as they briefly do in Section 6 - showing that low-wage subsidies can reduce incentives to obtain education)
- Basic Model:
  - $V_a^A(G_t, \kappa_{if}^0, \mu_{if}, R_i, \varepsilon_i, \omega_{it}) = \mu_{if} + W_a^A(G_t, X_{it} = 0, T_{fit} = 0, \kappa_{if}^0, \varepsilon_i) - [\lambda_R(R_i, G_t) + \lambda_0(\varepsilon_i)] - \omega_{it}$
  - Where:
  - V_a^A is the value of an apprenticeship to an individual of age a
  - W_a^A is the present value of employment
  - $[\lambda_R(R_i, G_t) + \lambda_0(\varepsilon_i)]$ gives the cost to the apprenticeship (direct and indirect, respectively)
  - It includes the value of the apprenticeship net of direct monetary costs and indirect utility costs of the apprenticeship
  - Note that the value of an apprenticeship itself is a recursive function - which includes the option value of switching apprenticeships and the value of subsequent employment (there's a lot going on here)
  - The state variables include the region of the apprenticeship, R_i, GDP, G_t, experience, X_{it}, tenure, T_{it}, match quality, κ_{if}, transition costs, μ_{if}, unobserved cost shocks, ω_{it} and unobserved heterogeneity, ε_i (ε_i is actually a vector of two, possibly jointly, distributed random variables to account for selection on unobserved returns to education and on ability).
- Identification:
  - Likelihood function is difficult to calculate
  - Additional difficulty due to data - not observed as quarterly frequency as model is set up, so have to deal with time aggregation issues
  - Thus they choose an SMM estimation procedure
  - 118 parameters to estimate
  - Choose 390 moments
  - These moments come from several linear regressions of wages and employment (and functions of these) on a host of covariates included in the model which characterize the career profile of earnings and employment for those who did and did not apprentice.
  - Also use moments on proportion of apprentices by year
- Data:
  - German administrative data - 2% sample of SS data
  - All work spells, start and end dates, 1975-1996
  - Includes spells of apprenticeship training and whether hold qualification
- Results:
  - Returns to apprenticeship driven by changes in the patter of earnings - more growth up front, less later on
  - Apprenticeships do result in less mobility - which affects workers after recessions
  - Unskilled also suffer from recessions, but mostly because of productivity declines