Agent-based Models: Methodology, Calibration and Estimation
Part 1: 'simple' estimation methods

Sylvain Barde

School of Economics, University of Kent

July 3, 2023

CEF pre-conference workshop, Nice

Outline

The pedagogical approach

Download material from

https://github.com/Sylvain-Barde/abm-validation-cef-2023

- Ideally, I'd have liked a proper workshop - where **you** do the work!
- But:
    - This is a short session with a lot to cover
    - There is uncertainty over the distribution of ability in the audience
    - Murphy's law: anything presentation that is IT based will fail and you embarrass yourself
- So instead, I will be going through a series of pre-prepared Jupyter notebooks
    - Please do ask questions in the workshop if you have them.
    - Please also get in touch if you hit a snag running these notebooks at a later date.

Disclaimers - Not the central bank kind...

- I am focusing on parameter estimation. Validation as a whole is a larger topic.
- Time constraints mean I cannot be exhaustive:
  - Not all methodologies will be covered, not all literature mentioned
  - So apologies in advance to colleagues whom I don't mention.
  - The workshop is more about communicating a 'design philosophy'
- This is especially true for the second part of the workshop
  - The methodology I will present is not necessarily the best or unique
  - ... but it's what I know.
- These PDF slides are there mainly to provide a summary of the main points when you come back to the material. I might not go over everything in them, and I might also duplicate things with the notebooks.
- So don't take this **any of this** as gospel. It's a guide, not a prescription!
  - I will try to flag what is fact (or as close as possible) and what is opinion

Flavours of Indirect inference

- So you have a simulation that runs a non-linear, emergent, hidden model for which you don't have tractable analytical representations, and you want to estimate the parameters from data.
    - Non-parametric is probably you best/only bet.
    - Instead of determining the statistical properties of your model from a functional form, you directly get them from simulated data from your model.
- The general framework this is indirect inference (Gourieroux et al., 1993, 1996), where you estimate an auxiliary model on the simulated data and compare it to the equivalent model on the real data.
- Smith (2008) offers a really good, non-technical explanation on interpretations of II depending on the loss function used
    - Using the Wald metric leads to simulated method of moments
    - Using the likelihood ration leads to simulated likelihood
- This part is well established and well-understood.
    - As a result, this section is relatively 'future proof'

Literature in the ABM community

- Relatively large literature around estimation of ‘small’ ABMs, usually financial ABMs, where
  - The models are computationally light
  - Univariate with plentiful data
- A far from exhaustive list:
  - *SMM* - Gilli and Winker (2003), Grazzini and Richiardi (2015)
  - *SML* - Kukacka and Barunik (2017)
  - *Sequential Monte Carle* - Lux (2018);
  - *Approximate Baysian Computing* - Grazzini et al. (2017)
  - *Bayesian estimation with NN* - Platt (2022).
- See Platt (2020) in particular for a good comparison of some of these methods.
- The methods and models examined follow in this vein.

Simulated Methods of Moments

Let $\mathbf{\Phi}(Y)$ be a function that produces a vector of moments from a dataset, and let

$$\Delta\mathbf{\Phi}(\theta) = \mathbf{\Phi}(Y^{sim}(\theta)) - \mathbf{\Phi}(Y^{emp})$$

Be the difference between the empirical and simulated moments for a given parameter vector θ. SMM aims to find the value of θ that minimises the following distance:

$$D = \Delta\mathbf{\Phi}(\theta)^T W \Delta\mathbf{\Phi}(\theta)$$

Issues:

- For identification, you need at least as many moments as you have parameters. In general, more...
- Estimates are sensitive to the choice of weights matrix W
- You want to put more weight on well-estimated parameters, less weight ones with more uncertainty.

Simulated Methods of Moments

The optimal matrix is $W = \Sigma^{-1}$, where Σ is the variance-covariance of the simulated moments evaluated at the parameter estimates $\hat{\theta}$.

This is a bit circular... How to choose the W matrix?:

- Keep it simple and go for $W = I$. This is not recommended.
- Start with $W = I$ and run an iterated estimation, updating W each time.
- Go for the Newey-West HAC estimator, which is proven to be optimal (although there is a bandwidth to tune...)
- Or use a bootstrap to get the variance-covariance (Tubbenhauer et al., 2021). We'll use this, because it's less hassle to program from scratch!

Simulated Likelihood Estimation, specifically NPSMLE

Use a Kernel method to obtain a approximation of the likelihood, where X_t are conditioning variables that can include lags of Y_t:

$$p(Y_t \mid X_t, \theta)$$

This can be done:

- Directly, over an ensemble of simulated series from the model.
  - See Grazzini et al. (2017)
- Or 'online' using a one-step-ahead implementation with fixed innovations ε_i :

$$\hat{p}(Y_t \mid X_t, \theta) = \frac{1}{N} \sum_i K\left(Y_t^{sim}(\theta, \varepsilon_i) - Y_t^{emp}\right)$$

  - Kukacka and Barunik (2017), based on the NPSMLE of Kristensen and Shin (2012)
  - This is much more efficient*, which is why I am demonstrating it.

* terms and conditions apply

A simple AR(2) example

The purpose is to use a simple example showcase:

- The properties of each method
- The actual implementation proposed here

This example allows to establish:

- The sensitivity of SML to correct shuffling:
    - This is the cost of the methodology's efficiency
    - SMM in an ensemble method, so it is much less sensitive
- The costs of SMM relative to SML
    - Extracting moments requires simulating the T dimension
    - Moments need to be picked
    - The W matrix needs to be correct, although this is minor

Brock & Hommes model

I guess we actually need to estimate/validate an ABM, right?

The Brock and Hommes (1998) is a good place to start:

- It is very popular in the literature, and has been used a lot
- It is simple enough to simulate, yet complex enough to generate a challenge

Main features:

- Multiple populations of traders, each with a price-forming strategy $f_{i,t} = g_i y_{t-1} + b_i$
- Populations evolve according to a logit with intensity of choice β:

$$n_{i,t} = \frac{\exp(\beta U_{i,t})}{\sum_i \exp(\beta U_{i,t})}$$

- $U_{i,t}$ encodes the past profitability of each price-forming strategy.

We are going to be following Kukacka and Barunik (2017) here.

Issues and takeaways - SMM

- SMM Pros:
  - Relies on ensemble methods. Doesn't care about your noise specification
  - Assuming identification, can usually produce good estimates
  - Overall, robust and reliable - not much to go wrong.
- SMM Cons
  - Relies on ensemble methods. Need to simulate $N \times T$ in each iteration.
  - This is compounded by the larger parameter space. 'Only' 2 more, yet time cost is dramatically higher... (due to gradient calculation)
  - Picking good moments can be tricky - as much art as science
  - This allows for 'researcher discretion'
  - Note: that is actually not a bad thing - know thy model!

Issues and takeaways - SMLE

- SMLE Pros:
  - Very desirable theoretically (Likelihoods are great, after all).
  - In particular, no need to worry about moments.
  - Very efficient thanks to the one-step-ahead approach if using NPSMLE. Only need $N \times 1$ vector for each iteration
- SMLE Cons
  - Very reliant on correct specification of the noise and correct handling of pseudo-random numbers.
  - NPSMLE requires a one-step ahead specification of the model. Without that, you have to resort to KDE ensemble estimation (Grazzini et al., 2017), which negates the advantage.
  - Finally, the one-step-ahead assumption means the model can be conditioned on past lags of the observable variable alone - difficult to handle hidden models.

Some final generic advice - MEGA disclaimer here...

- You might gather I quite like parameter recovery exercises...:
  - Separates parameter uncertainty from model uncertainty
  - Gives an idea of whether your method is working OK
- Ergodicity is an issue we've not even covered.
  - ABMs in general are fart from ergodic.
  - Problem is, most methods require this!
  - Test your methods with different lengths to see if you have a problem.
  - Always keep your simulation length in line with the empirical data.
- If possible, write your ABM classes with a ‘step’ method.
  - It really helps, see Vandin et al. (2022) for example.
  - But it might not always be possible.
- Finally, **estimation** should inform the **design** of your ABM.
  - When adding behavioural rules/parameters, keep in mind the constraints around estimating them.

Brock, William A and Cars H Hommes (1998) "Heterogeneous beliefs and routes to chaos in a simple asset pricing model," *Journal of Economic dynamics and Control*, Vol. 22, pp. 1235–1274.

Gilli, Manfred and Peter Winker (2003) "A global optimization heuristic for estimating agent based models," *Computational Statistics & Data Analysis*, Vol. 42, pp. 299–312.

Gourieroux, Christian, Monfort Gourieroux, Alain Monfort, and Director Alain Monfort (1996) *Simulation-based econometric methods*: Oxford university press.

Gourieroux, Christian, Alain Monfort, and Eric Renault (1993) "Indirect inference," *Journal of applied econometrics*, Vol. 8, pp. S85–S118.

Grazzini, Jakob and Matteo Richiardi (2015) "Estimation of ergodic agent-based models by simulated minimum distance," *Journal of Economic Dynamics and Control*, Vol. 51, pp. 148–165.

Grazzini, Jakob, Matteo G Richiardi, and Mike Tsionas (2017) "Bayesian estimation of agent-based models," *Journal of Economic Dynamics and Control*, Vol. 77, pp. 26–47.

Kristensen, Dennis and Yongseok Shin (2012) "Estimation of dynamic models with nonparametric simulated maximum likelihood," *Journal of Econometrics*, Vol. 167, pp. 76–94.

Kukacka, Jiri and Jozef Barunik (2017) "Estimation of financial agent-based models with simulated maximum likelihood," *Journal of Economic Dynamics and Control*, Vol. 85, pp. 21–45.

Lux, Thomas (2018) "Estimation of agent-based models using sequential Monte Carlo methods," *Journal of Economic Dynamics and Control*, Vol. 91, pp. 391–408.

Platt, Donovan (2020) "A comparison of economic agent-based model calibration methods," *Journal of Economic Dynamics and Control*, Vol. 113, p. 103859.

——— (2022) "Bayesian estimation of economic simulation models using neural networks," *Computational Economics*, Vol. 59, pp. 599–650.

Smith, Anthony (2008) "Indirect inference," *The New Palgrave Dictionary of Economics, 2nd Edition (forthcoming)*.

Tubbenhauer, Tobias, Christian Fieberg, and Thorsten Poddig (2021) "Multi-agent-based VaR forecasting," *Journal of Economic Dynamics and Control*, Vol. 131, p. 104231.

Vandin, Andrea, Daniele Giachini, Francesco Lamperti, and Francesca Chiaromonte (2022) "Automated and distributed statistical analysis of economic agent-based models," *Journal of Economic Dynamics and Control*, Vol. 143, p. 104458.