# Simulation Estimators – Part II

## Summer School in Structural Estimation

Toni M. Whited

June 2023

# Outline

1. Identification Again
2. Indirect Inference
3. Tips

# Identification checks

- There are several ways to check whether your model is well identified
- Check #1: Check the standard errors
  - Huge standard errors are usually a symptom of poor identification
  - The parameters' standard errors depend on $\partial g\left(y_{is}\left(\theta\right)\right)/\partial\theta$, the sensitivity of moments to parameters
  - If the sensitivity is low, the derivative will be near zero, which will produce huge standard errors for the structural parameters
- Check #2: Does the Jacobian matrix $\partial g\left(y_{is}\left(\theta\right)\right)/\partial\theta$ have full rank?
  - If not, your model is not locally identified
  - This is almost the same as a standard error check, but not quite.

# Computer output from an estimation gone wrong

```
 Parameter estimates:
&        0.100 &        0.231 &        0.490 &        0.756 &        0.518 &        0.128 &        0.000 &        0.008 &
&(       2.033)&(       0.013)&(       0.000)&(       0.073)&(       0.264)&(       0.000)&(       0.004)&(       0.000)&(

 Gradient matrix:

      0.00000000      0.00000000      1.02758239      0.00000000      0.00000000     38.25073505      0.00000000      0.00000000
      0.00000000      0.00000000      0.28512603      0.00000000      0.00000000      4.82100673      0.00000000      0.00000000
      0.00000000      0.00000000     -9.97948112      0.00000000      0.00000000    370.80545008      0.00000000      0.00000000
      0.00000000      0.00000000      0.40639694      0.00000000      0.00000000      1.02101543      0.00000000      0.00000000
      0.00000000      0.00000000      0.84485716      0.00000000      0.00000000      0.04210265      0.00000000      0.00000000
      0.00000000      0.00000000   -138.58951690      0.00000000      0.00000000      0.00889156     -0.00000004      0.00000000
      0.00000000      0.05056345      0.03244400     -0.16356671      0.00000000     -5.85773601     -2.79010277     -2.79010369
      0.00000000      0.01667884      0.00342592     -0.00333117      0.00000000     -0.09697136     -0.07479289     -0.07479291
      0.00000000     -0.29607166      1.09787363      0.14226033      0.00000000    111.43914416      0.00006803      0.00000173
     -2.06371354      3.34618706     10.68850788    -11.02873288     -0.77796138    -74.85216998    736.96355852   -153.90999884
     -0.01039442      0.02335486     -0.62675135     -0.05792594     -0.00419307     -6.10626047      3.66442260     -0.78400489
      0.00000000      0.04861350     -3.09825605      0.11134717      0.00898020    -25.21239839      2.03722313      1.89152864
      0.00000000      0.00122617     -0.74021211     -0.00036989      0.00009098     -0.66126219     -0.00761401     -0.00819525
      0.00000000      0.08164404     -1.21010451     -0.13848512     -0.00552660     -3.77971077     -2.64720025     -2.53142955
      0.00000000      0.00444937      0.02217510     -0.00057544      0.00000972     -0.26535678     -0.01661087     -0.01716714
      0.00000000      0.00038676      0.00024816     -0.00125113      0.00000000     -0.04480597     -0.02134156      0.11961819
```

# Identification checks

- Check # 3: Can estimation recover the true parameter values?
  - 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?
- Check #4: Start searching from different initial parameter guesses
  - Should reach same final estimate regardless of initial guess
  - If not: Coding errors? Stuck in local min? Model not well identified?
- Check #5: Babysit your code
  - Some parameters converge faster than other. Keep track of this.
  - If one or two parameters are changing drastically during the estimation, you have a problem.

# Local identification diagnostic

- Andrews, Gentzkow and Shapiro (2017) provide a local diagnostic that measures the sensitivity of $\hat{\boldsymbol{\theta}}$ to the estimated moments, $\hat{\boldsymbol{g}}$.
- Let $\boldsymbol{G}$ be the Jacobian of $g$ w.r.t $\theta$.
- The diagnostic is (and it should look familiar):

$$\boldsymbol{S} = -(\boldsymbol{G}'\boldsymbol{\Xi}\boldsymbol{G})^{-1}\boldsymbol{G}'\boldsymbol{\Xi}.$$

- This measure is not scale invariant, so they propose the following normalization:

$$\boldsymbol{S}_{l,p}\sqrt{\frac{\boldsymbol{\Lambda}_{l,l}}{\boldsymbol{V}_{p,p}}},$$

  where $\boldsymbol{V}_{p,p}$ a moment variance and $\boldsymbol{\Lambda}_{l,l}$ is a parameter variance.
- This diagnostic captures both
  - the steepness of the gradients
  - the precision of the estimation of the moments that identify the parameters

# Identification has more than one meaning

- A parameter can be identified (in the formal statistical sense) without being economically interesting
- Two examples:
  - Regression of endogenous $Y$ on endogenous $X$
  - Regression of endogenous $Y$ on exogenous $X$ (without a clear and interesting economic mechanism)
- In both cases, the econometric objective function has a unique minimum.
- Your ultimate goal: Identify **economically interesting** parameters
  - The parameters may be elasticities defining causal effects
  - But they need not be!
  - **Not all economically interesting parameters are causal elasticities**

# “Identification is not causality, and vice versa”

- (Kahn and Whited 2018), Review of Corporate Finance Studies
- Exogenous variation is:
  - Always necessary to identify a causal relation
  - Never sufficient for identifying an economically interesting parameter
    - You also need an economic model (either mathematical or verbal)
  - Only sometimes necessary to identify an economically interesting parameter
- Interesting parameters can sometimes be identified without exogenous variation
  - This is often what’s going on in structural corporate finance.

# Do you need exogenous variation to do SMM?

- Causal elasticities are not always helpful in structural estimation! (Kahn and Whited 2018)
- Compare SMM estimators of a simple investment model
  - Approach #1: Target moments that measure a causal elasticity
  - Approach #2: Target moments that measure an endogenous elasticity
- Result: The performance of the two approaches is nearly identical!
- In both cases,
  - Multiple moments and multiple parameters interact
  - Identification leans on the structure of the model
- Identification **always** leans on assumptions
  - Even in reduced-form papers with exogenous variation
  - Those papers often lean on a “verbal” economic model

# Causal elasticities can be helpful moments to target . . .

- **If** the “exogenous variation” is the same in the model and the data.
  - For example, do not call an endogenous price in the real data “exogenous,” and then model an exogenous price.
- **If** you want to understand what fundamental forces affect the causal elasticity.

# Outline

# Economic Models and Auxiliary Models

- If a model does not provide a closed form likelihood, or the simulated likelihood needs to be calculated for more than two variables (and two is a stretch), then you can use an auxiliary model to estimate the model parameters.
- A likelihood is a description of the true data generating process.
- An auxiliary model is an approximation of the true data generating process.
  - What if you have a DSGE model of the macroeconomy. The likelihood is impossible to solve for, but a VAR might describe the data approximately.
  - What if you have a model of investment spikes or infrequent price adjustments. The auxiliary model could be a duration model.

# Auxiliary Models

- In practice, 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.

# Auxiliary Models

- 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.
- To the extent that the parameters of the auxiliary model are functions of moments of the data, indirect inference can be thought of as a superset of SMM.
- It falls under the category of a simulated minimum distance (SMD) estimator.

# Estimation: Minimum Distance Style

- Notation:
  - $x_N$ is a data matrix of length $N$
  - $x_N^s$ is a simulated data matrix of length $N$ from simulation $s$, $s = 1, \dots, S$.
  - $\theta$ a vector of **auxiliary** model parameters estimated with **real** data.
  - $\theta^s$ a vector of **auxiliary** model parameters estimated with **simulated** data.
  - $b$ is the vector of parameters from the **economic** model.

- Without loss of generality, the parameters of the auxiliary model can be represented as the solution to the maximization of a criterion function

$$\theta_N = \arg\max_{\theta} J_N \left( x_N, \theta \right),$$

- Examples?

# Estimation: Minimum Distance Style

- Construct $S$ simulated data sets based on a given parameter vector, $b$.
- For each of these data sets, estimate $\theta^s$ by maximizing an analogous criterion function

$$\theta_N^s\left(b\right) = \arg\max_{\theta} J_N\left(x_N^s, \theta^s\left(b\right)\right),$$

- Note that the $\theta_N^s\left(b\right)$, as explicit functions of the structural parameters, $b$.
- The inverse of this function is what Gourieroux and Monfort call a “binding” function.

# Estimation: Minimum Distance Style

- The indirect estimator of $b$ is then defined as the solution to the minimization of

$$\begin{aligned}\hat{b} &= \arg\min_{b} \left[\theta_N - \frac{1}{S}\sum_{h=1}^{S}\theta_N^s(b)\right]' \hat{W}_N \left[\theta_N - \frac{1}{S}\sum_{h=1}^{S}\theta_N^s(b)\right] \\ &\equiv \arg\min_{b} \hat{G}_N' \hat{W}_N \hat{G}_N\end{aligned}$$

- $\hat{W}_N$ is a positive definite matrix that converges in probability to a deterministic positive definite matrix $W$.
- As in GMM the optimal weight matrix is the inverse of the covariance matrix of $\theta$.
- The main difference between SMM and this flavor of indirect inference is that the former uses **moments** and the latter uses **functions of moments**.

# Inference: Minimum Distance Style

- The indirect estimator is asymptotically normal for fixed $S$. Define $J \equiv \text{plim}_{N\to\infty}\left(J_N\right)$. Then

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

  where

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

- The technique provides a test of the overidentifying restrictions of the model, with

$$\frac{NS}{1+S}\hat{G}_N'\hat{W}_N\hat{G}_N$$

  converging in distribution to a $\chi^2$, with degrees of freedom equal to the dimension of $\theta$ minus the dimension of $b$.

# Outline

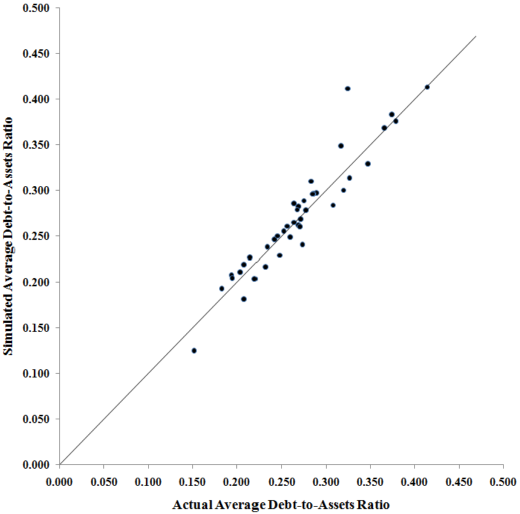
# Miscellaneous Pieces of Advice from the Trenches

- Heterogeneity
- The black box problem
- Choosing the weight matrix
- Computation

# 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 or at best an industry equilibrium with very limited heterogeneity.
- 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 Splits[1]




[1] DeAngelo, DeAngelo and Whited (2011)

# Do Not Construct a Black Box

- More parameters $\neq$ a better model!!!!!!
- Different features of the data should change when underlying parameters change.
- This is less likely to happen in a model with many extraneous parameters.
- 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.

# How Not to Construct a Black Box

- The following is eerily the same as, but not taken from, the material in this video by Michael Keane:

  `https://www.youtube.com/watch?v=0hazaPBAYWE`

- If you don't believe me, maybe you will believe him.

# How Not to Construct a Black Box

- Start simple!!!
- Make sure the simple model is **right**.
- Figure out where the simple model succeeds and fails.
- Add a feature that will help you answer the question you want to answer.
- Make sure the slightly more complicated model is right.
- Figure out where the more complicated model succeeds and fails.
- Converge.

# The question comes first — Not the model

- Before going structural, convince yourself that a structural approach is absolutely necessary.
- Are there data limitations?
- Are you interested in something besides a causal elasticity?

# Use the right weighting matrix

- Many structural estimation papers use odd weighting matrices.
- Identity weight matrix
  - This choice mechanically puts more weight on moments that are larger in absolute value.
  - Not an economically sensible choice.
  - Bazdresch, Kahn and Whited (2018) note poor finite sample performance.

# Use the right weighting matrix

- Inverse of the squared moments on the diagonal
  - This choice minimizes the percentage differences in the moments.
  - This choice mechanically puts more weight on moments that are smaller in absolute value.
  - Not an economically sensible choice.

# Use the right weighting matrix

- Inverse moment variance along the diagonal
  - This choice puts the least weight on moments that are estimated with the least precision.
  - A statistically more or less sensible choice.
  - The old version of Bazdresch et al. (2018) (before the referees got to it) note poor finite sample performance.

# Use the right weighting matrix

- Inverse clustered moment covariance matrix
  - This choice minimizes overall model error.
  - Takes into account moment covariances.
  - Bazdresch et al. (2018) note good finite sample performance.
- If at all possible, do this.

# Do not treat the process of estimating the model mechanically!

- First, try to calibrate the model by hand.
- You will learn about which parts of the model are useful for understanding which parts of the data.
- You will learn a lot about the economics of the model.
- You will also learn a lot about identification.
- You will end up with a good starting value for your estimation.
- Only after you have done this should you start your estimation running.

# 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
- Multi-start algorithms, combined with Nelder Meade tend to work very poorly, except, maybe.
- Tiktak algorithm (Arnoud, Guvenen and Kleineberg 2017), but it uses Dropbox for parallelization . . .
- Use the simulated annealing (SA), particle swarm (PS), or differential evolution algorithm (DE) to avoid local minima
- DE and PS can be parallelized, but I have found that they can converge way too fast or not at all, depending on technical parameter settings. In technical terms, they're downright squirrelly.
- Use the same seed for the random-number generator each time you simulate data off the model. Do not mess up this step.

# Software

- Do not use Matlab, R, Python, 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
- People tell me that Numba $+$ Python is fast but unstable for large problems.
- My initial experimentation with Julia has not been bad. Still not as fast as Fortran.
- Learn how to exploit multiple processors, a graphics card, a supercomputer, . . . .

# Remember that you are actually doing estimation

- 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.
- You know what to do!

Andrews, I., Gentzkow, M.„ Shapiro, J.M., 2017. Measuring the sensitivity of parameter estimates to sample statistics. Quarterly Journal of Economics 132, 1553–1592.

Arnoud, A., Guvenen, F.„ Kleineberg, T., 2017. Benchmarking global optimization algorithms. Manuscript, Yale University.

Bazdresch, S., Kahn, R.J.„ Whited, T.M., 2018. Estimating and testing dynamic corporate finance models. Review of Financial Studies 31, 322–361.

DeAngelo, H., DeAngelo, L.„ Whited, T.M., 2011. Capital structure dynamics and transitory debt. Journal of Financial Economics 99, 235–261.

Kahn, R.J.„ Whited, T.M., 2018. Identification is not causality, and vice-versa. Review of Corporate Finance Studies 7, 1–21.