Numerical Methods for Simulation of Structural Models

Macro Applications

Fernando Antônio de Barros Júnior

UFPR

fabarrosjr@gmail.com

May 27, 2019

Overview

Value Function Iteration

Calibration of Macro models

Calibration Applied

Simulated Method of Moments

SMM Applied

Indirect Inference

Identification

Dynamics

- The most of the economic phenomenon can be represented in a static model
  - Sometimes it requires TWO periods
- Quantify an economic phenomenon requires a dynamic model the most of the time
- Solving a dynamic model can be tricky, but we have some tools...
- Today: Value Functions!
- Let's go to the board...

Value function iteration

- Well known, basic algorithm of dynamic programming
- Well suited for parallelization
- It will always work (perhaps quite slowly)
- Subject to the curse of dimensionality

An example

RBC model

$$V(k, z) = \underset{c,k'}{\text{Max}}\, u(c) + \beta EV(k', z)$$

$$\text{s.t. } c + k' \leq zf(k) + (1 - \delta)k$$

Iteration

- Start with an initial guess for the value function, V^0
- Maximize the right hand side and compute V^1
- Repeat procedure until $||V^j - V^{j-1}|| < \varepsilon$, where ε is a small number

The simplest way

- Discretize everything!
- Create grids for the states and for the controls
- What are the relevant k and z? $\mathcal{K} \times \mathcal{Z}$
- How can V be represented on a computer? As an array
- How does one compute $E[z]$? Markov chain
- How does one compute the optimal c and k'? Checking all k'

Questions

- What is an appropriate/optimal discretization of $\mathcal{K} \times \mathcal{Z}$?
- How does one discretize the AR1 into a Markov chain? Tauchen, Rouwenhorst
- How does one assess the quality of the approximation?
- tips about $\mathcal{K}$
  - Upper and lower bounds: $[\underline{k}, \bar{k}] = [0.7k_{ss}, 1.3k_{ss}]$
  - Simulate and check if it binds!
  - Where should the points be? Linear spaced; Concentrated, usually around $\underline{k}$; Stochastic
  - How many points should there be? Ideally, adding points wouldn't change your results. Check!

Brute Force Grid Search

Brute force grid search is checking every $k' \in \mathcal{K}$ value and taking the maximum. (Be careful with feasibility!)

Advantages

- Always works - can handle discontinuities, non-convex budget constraints, non-concave objective functions
- Some problems are naturally discrete choice
- Easy for compilers to optimize and easy to make parallel

Disadvantages

- Very slow relatively
- Suffers greatly from the "curse of dimensionality"
- Most problems allow faster algorithms

Exploiting Monotonicity

Very often, one has policy functions that are monotone, e.g.,

$$k'(k_2, z) \geq k'(k_1, z) \; \forall \; k_2 > k_1 \text{ and } z$$

is monotone increasing in capital

This can exploited when finding the optimal $k'(k, z)$:

1. Compute $k'(k_i, z)$ by checking all $k_1, \ldots, k_n$. Set $i = 1$
2. Compute $k'(k_{i+1}, z)$ by checking all $k'(k_i, z), \ldots, k_n$
3. If $i = n$, Stop. O/W increment i and go to 2.

Exploiting Monotonicity

Advantages

- Typically much faster than brute-force
- Applies to many models
- Can still parallelize easily, e.g., over the z values

Disadvantages

- Dangerous if not sure that monotonicity holds
- If it doesn't hold, then VFI may not converge or may converge to something wrong
- This danger can be greatly mitigated by either:
  - explore monotonicity every 9 out of 10 iterations
  - randomly check (say $\approx$ every 1000th loop over states) that the policy found using monotonicity agrees with brute force.
- Note: if monotonicity almost holds, you can "relax" the monotonicity assumption: find $k'(k_{i+1}, z)$ by checking all $k'(k_{i-5}, z), \ldots, k_n$ for example

Another Trick: Multigrid

Basic idea: solve first a problem in a coarser grid and use it as a guess for more refined solution

It works like such:

1. Let there be an initial grid $\mathcal{K}_0$. Let $i = 0$
2. Compute the optimal value V on $\mathcal{K}_i$. If the grid $\mathcal{K}_i$ is "fine enough", stop
3. Define a finer grid $\mathcal{K}_{i+1}$ and interpolate V from $\mathcal{K}_i$ onto $\mathcal{K}_{i+1}$
4. Go to 2

Example: Non-stochastic growth model

```
while (dist>tol && it<itmax)
    it = it + 1;
    for i=1:N;
        C = Grid(i)^alpha+((1-delta)*Grid(i))-Grid;
        ind = find(C<0);
        C(ind) = eps;
        aux = C.^(1-theta)./(1-theta) + beta*V;
        [TV(i),pos] = max(aux);
        Pol(i) = Grid(pos);
    end
    dist = abs(max(TV-V));
    V = TV;
end
```

Comparison

$\underline{k} = 1$, $\bar{k} = 1000$ and $V^0 = \mathbf{0}$

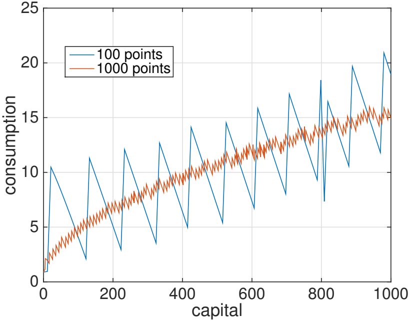
Multigrid

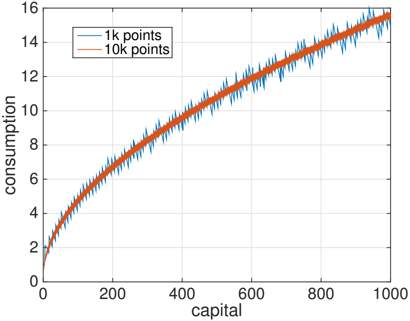
- $\mathcal{K}_1$: 100 points; iterations: 461
- $\mathcal{K}_2$: 1000 points; iterations: 375
- $\mathcal{K}_3$: 10000 points; iterations: 82
- Total time: 143.85 seconds.

Brute Force

- $\mathcal{K}$: 10000 points; iterations: 454
- Total time: 782.60 seconds.

Grid choice


100 points
1000 points
consumption
capital
0
200
400
600
800
1000
0
5
10
15
20
25


1k points
10k points
consumption
capital
0
200
400
600
800
1000
0
2
4
6
8
10
12
14
16

Other methods

- Exploiting concavity
- Endogenous grid method
- trick: accelerator

Calibration

- What it is
- Why we do it
- Pitfalls
- Why theorists do not like it

Calibration

- Most widespread view: Calibration is the process by which researchers choose the parameters (and functional forms) of their economic models from various sources. Most commonly, this is done by
  - the use of time series averages of the levels or ratios of economic variables
  - the estimation of single equations
  - reference to econometric studies based on either macro- or micro-data
  - setting the parameters so that the model replicates certain empirical facts such as conditional or unconditional moments of the data

Calibration: a Critique

- Lack of statistical formality associated with calibration exercises
- Judgments of "good" or "bad", or as Kydland and Prescott (1996, p. 71) put it, judgments of whether "... the predictions of theory match the observations...", are necessarily subjective
- How do we compare two theories using calibrated models?
- Micro to Macro (Aggregation problem)
  - Houthakker (1956) demonstrated that the aggregation of Leontief micro production technologies yields an aggregate Cobb-Douglas production function

Applying calibration

Barros and Gomes (2019): a decomposition of the welfare cost of economic fluctuations

- Consumers are risk averse
- If they can not smooth consumption (incomplete markets)
  - $\Rightarrow$ welfare loss
- Shocks may hit the consumption
- Stabilization policies are always in place
  - How good they are?
  - Can they be improved?
- Can we fully smooth shocks?

The welfare cost of business cycles

- Lucas (1987): propose a measure to this cost
  - A constant λ that solves

$$\mathbb{E}_0\left(\sum_{t=0}^{\infty}\beta^t u\left([1+\lambda]C_t\right)\right)=\sum_{t=0}^{\infty}\beta^t u\left(\bar{C}_t\right) \tag{1}$$

- This calculation requires a model for consumption (and preferences)
- Lucas (1987):
  - representative consumer
  - CRRA preference
  - log-normal consumption: iid around a linear trend
  - the benefit of improving stabilization policies is 0.04% of *per capita* consumption

Some details of the welfare cost of cycles

- Let $\{C_t\}_{t=0}^{\infty}$ be a risk consumption sequence
- and $\{\bar{C}\}_{t=0}^{\infty}$ represents the previous consumption sequence without volatility
- Following Lucas (1987), we can define the welfare cost of business cycles as a scalar λ that solves

$$\mathbb{E}_0\left(\sum_{t=0}^{\infty}\beta^t u\left([1+\lambda]C_t\right)\right)=\sum_{t=0}^{\infty}\beta^t u\left(\bar{C}_t\right), \tag{2}$$

where

$$u(C)=\begin{cases}\dfrac{C^{1-\gamma}}{1-\gamma}, \text{ if } \gamma>0, \gamma\neq 1 \\ \ln(C), \text{ if } \gamma=1\end{cases}.$$

The welfare cost of macro uncertainty for log-normal consumption

Assumption (1)

The consumption process can be expressed as $C_t = \bar{C}_t X_t$, *where* $X_t = e^{-0.5\sigma_t^2 + x_t}$, $x_t \sim \mathcal{N}(0, \sigma_t^2)$ *and* $\bar{C}_t = \alpha_0 (1+\alpha_1)^t$.

Assumption (2)

Suppose that $\sum_{t=0}^{\infty} \left(\beta (1+\alpha_1)^{1-\gamma}\right)^t e^{-0.5\gamma(1-\gamma)\sigma_t^2} < \infty$.

Proposition

Let $\Gamma \equiv \beta(1+\alpha_1)^{1-\gamma}$. *Suppose that* $\Gamma < 1$. *Under assumption 1 and 2*

$$\lambda = \begin{cases} \exp\left(\frac{1-\beta}{2} \sum_{t=0}^{\infty} \beta^t \sigma_t^2\right) - 1, & \text{if } \gamma = 1 \\ \left[\frac{\sum_{t=0}^{\infty} \Gamma^t}{\sum_{t=0}^{\infty} \Gamma^t e^{-0.5\gamma(1-\gamma)\sigma_t^2}}\right]^{\frac{1}{1-\gamma}} - 1, & \text{if } \gamma > 0, \gamma \neq 1 \end{cases}.$$

A upper bound to stabilization policies

- Suppose that exist a constant $\theta \in [0, 1]$
- given θ, define an consumption series $\{C_t(\theta)\}_{t=0}^{\infty}$, where

$$C_t(\theta) = C_t^{1-\theta} \bar{C}_t^{\theta}, \qquad \forall\, t. \tag{3}$$

- Under assumption 1,

$$C_t(\theta) = \bar{C}_t X_t^{1-\theta}. \tag{4}$$

- $C_t(\theta)$ is a counterfactual consumption for C_t

Compensating consumers for not improving economic policy

- How much are the consumers willing to pay for the best stabilization policy?
- Analogously to Lucas (1987), we can define a constant $\lambda^P(\theta)$, which solves

$$\mathbb{E}_0\left(\sum_{t=0}^{\infty}\beta^t u\left([1+\lambda^P(\theta)]C_t\right)\right)=\mathbb{E}_0\left(\sum_{t=0}^{\infty}\beta^t u\left(C_t(\theta)\right)\right). \quad (5)$$

The partial welfare cost of macro uncertainty

Assumption (3)

Suppose that

$$\sum_{t=0}^{\infty} \left(\beta(1+\alpha_1)^{1-\gamma}\right)^t e^{-0.5(1-\theta)(1-\gamma)(\gamma+\theta-\gamma\theta)\sigma_t^2} < \infty.$$

Proposition

Suppose that $\Gamma < 1$*. Under assumption 1 to 3,*

$$\lambda^P(\theta) = \begin{cases} \exp\left(\theta \frac{1-\beta}{2} \sum_{t=0}^{\infty} \beta^t \sigma_t^2\right) - 1, \ \textit{if } \gamma = 1 \\ \left[\frac{\sum_{t=0}^{\infty} \Gamma^t e^{-0.5(1-\theta)(1-\gamma)(\gamma+\theta-\gamma\theta)\sigma_t^2}}{\sum_{t=0}^{\infty} \Gamma^t e^{-0.5\gamma(1-\gamma)\sigma_t^2}}\right]^{\frac{1}{1-\gamma}} - 1, \ \textit{if } \gamma > 0, \gamma \neq 1 \end{cases}$$

The nonsmoothable welfare cost of macro uncertainty

Again, we can compensate consumer for the inability of economic policy in smoothing shocks

$$\mathbb{E}_0\left(\sum_{t=0}^{\infty}\beta^t u\left([1+\lambda^{NS}(\theta)]C_t(\theta)\right)\right)=\mathbb{E}_0\left(\sum_{t=0}^{\infty}\beta^t u\left(\bar{C}_t\right)\right). \quad (6)$$

Proposition

Suppose that $\Gamma < 1$. Under assumption 1 and 3,

$$\lambda^{NS}(\theta)=\begin{cases}\exp\left((1-\theta)\frac{1-\beta}{2}\sum_{t=0}^{\infty}\beta^t\sigma_t^2\right)-1, & \text{if } \gamma=1\\ \left[\frac{\sum_{t=0}^{\infty}\Gamma^t}{\sum_{t=0}^{\infty}\Gamma^t e^{-0.5(1-\theta)(1-\gamma)(\gamma+\theta-\gamma\theta)\sigma_t^2}}\right]^{\frac{1}{1-\gamma}}-1, & \text{if } \gamma>0,\gamma\neq 1\end{cases} \quad (7)$$

Welfare Cost of Macro Uncertainty

$$U\left(\{(1+\lambda)C_t\}_{t=0}^{\infty}\right) = U\left(\left\{\bar{C}_t\right\}_{t=0}^{\infty}\right)$$

$$C_t \longrightarrow \mathbb{E}_0[C_t] \; = \; \bar{C}_t$$

C_t: Observed Consumption

$\mathbb{E}_0[C_t] = \bar{C}_t$: Consumption without risk

Partial Welfare Cost of Macro Uncertainty

$$U\left(\left\{\left(1+\lambda^{P}\right) C_t\right\}_{t=0}^{\infty}\right) = U\left(\left\{C_t(\theta)\right\}_{t=0}^{\infty}\right)$$

C_t $\qquad$ $C_t(\theta) \;=\; C_t^{1-\theta}\,\bar{C}_t^{\theta}$ $\qquad$ $\mathbb{E}_0[C_t] \;=\; \bar{C}_t$

Observed Consumption

Consumption without risk

Non-smoothable Cost of Macro Uncertainty

$$U\left(\left\{\left(1+\lambda^N\right) C_t(\theta)\right\}_{t=0}^{\infty}\right) = U\left(\left\{\bar{C}_t\right\}_{t=0}^{\infty}\right)$$

C_t

$C_t(\theta) \;=\; C_t^{1-\theta}\,\bar{C}_t^{\theta}$

$\mathbb{E}_0[C_t] \;=\; \bar{C}_t$

Observed Consumption

Consumption without risk

A decomposition of the welfare cost of business cycles

- From the previous results, we can show that

Proposition

Suppose that $\Gamma < 1$. Under assumption 1 to 3, the welfare cost of business cycles can be decomposed as

$$(1+\lambda) = (1+\lambda^{G}(\theta))(1+\lambda^{NS}(\theta)). \tag{8}$$

Corollary

Follows from the previous proposition that

$$\lambda \approx \lambda^{G}(\theta) + \lambda^{NS}(\theta). \tag{9}$$

Decomposition of Lucas's welfare cost

- $C_t = \alpha_0(1+\alpha_1)^t e^{-0.5\sigma_\mu^2 + x_t^L}$, where $x_t^L \sim \mathcal{N}(0, \sigma_\mu^2)$

$$\lambda = \begin{cases} \exp\left(\frac{\sigma_\mu^2}{2}\right) - 1, \text{ if } \gamma = 1 \\ \exp\left(\frac{\gamma\sigma_\mu^2}{2}\right) - 1, \text{ if } \gamma > 0, \gamma \neq 1 \end{cases} \tag{10}$$

$$\lambda^P(\theta) = \begin{cases} \exp\left(\theta\frac{\sigma_\mu^2}{2}\right) - 1, \text{ if } \gamma = 1 \\ \exp\left(\frac{\gamma\sigma_\mu^2}{2} - \frac{(1-\theta)(\gamma+\theta-\gamma\theta)\sigma_\mu^2}{2}\right) - 1, \text{ if } \gamma > 0, \gamma \neq 1 \end{cases} \tag{11}$$

$$\lambda^N(\theta) = \begin{cases} \exp\left((1-\theta)\frac{\sigma_\mu^2}{2}\right) - 1, \text{ if } \gamma = 1 \\ \exp\left(\frac{(1-\theta)(\gamma+\theta-\gamma\theta)\sigma_\mu^2}{2}\right) - 1, \text{ if } \gamma > 0, \gamma \neq 1 \end{cases} \tag{12}$$

Decomposition of Obstfeld's welfare cost

- $C_t = \alpha_0(1+\alpha_1)^t e^{-0.5\sigma_\mu^2 + x_t^O}$, where
  $x_t^O = \sum_{i=1}^t \varepsilon_i,\ \varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2)$

$$\lambda = \begin{cases} \exp\left(\frac{\beta\sigma_\varepsilon^2}{2(1-\beta)}\right) - 1, \text{ if } \gamma = 1 \\ \left[\frac{1-\Gamma e^{-0.5\gamma(1-\gamma)\sigma_\varepsilon^2}}{1-\Gamma}\right]^{\frac{1}{1-\gamma}} - 1, \text{ if } \gamma > 0, \gamma \neq 1 \end{cases} \tag{13}$$

$$\lambda^P(\theta) = \begin{cases} \exp\left(\theta\frac{\beta\sigma_\varepsilon^2}{2(1-\beta)}\right) - 1, \text{ if } \gamma = 1 \\ \left[\frac{1-\Gamma e^{-0.5\gamma(1-\gamma)\sigma_\varepsilon^2}}{1-\Gamma e^{-0.5(1-\theta)(1-\gamma)(\gamma+\theta-\gamma\theta)\sigma_\varepsilon^2}}\right]^{\frac{1}{1-\gamma}} - 1, \text{ if } \gamma > 0, \gamma \neq 1 \end{cases} \tag{14}$$

$$\lambda^N(\theta) = \begin{cases} \exp\left((1-\theta)\frac{\beta\sigma_\varepsilon^2}{2(1-\beta)}\right) - 1, \text{ if } \gamma = 1 \\ \left[\frac{1-\Gamma e^{-0.5(1-\theta)(1-\gamma)(\gamma+\theta-\gamma\theta)\sigma_\varepsilon^2}}{1-\Gamma}\right]^{\frac{1}{1-\gamma}} - 1, \text{ if } \gamma > 0, \gamma \neq 1 \end{cases} \tag{15}$$

Identification of consumption process parameters

Lucas (1987)

- We can the following model

$$\log(C_t) = \pi_0 + \pi_1 t + u_t,$$

- Then, $\hat{\sigma}_\mu^2 = \mathrm{Var}(\hat{u}_t)$ and $\hat{\alpha}_1 = \exp(\hat{\pi}_1) - 1$

Obstfeld (1994)

- We can the following model

$$\Delta \log(C_t) = \delta_0 + z_t,$$

- Then, $\hat{\sigma}_\varepsilon^2 = \mathrm{Var}(\hat{z}_t)$ and $\hat{\alpha}_1 = \exp(\hat{\delta}_0 - 0.5\hat{\sigma}_\varepsilon^2) - 1$

Estimation and calibration

Estimation

- USA data: 1947Q1 - 2017Q3

	Lucas (1987)	Obstfeld (1994)
$\hat{\alpha}_1$	0.0052	0.0047
$\hat{\sigma}^2_\mu$	0.0026	
$\hat{\sigma}^2_\varepsilon$		0.000025

Calibration

- $\beta \in \{0.97, 0.98, 0.99\}$
- $\gamma \in \{1, 2.5, 5, 10, 20\}$
- $\theta \in \{0.25, 0.50, 0.75\}$

Lucas's consumption process

γ	1	2.5	5	10	20
λ	0.12	0.32	0.64	1.27	1.95
		$\theta = 0.25$			
$\lambda^P(\theta)$	0.03	0.12	0.26	0.54	0.82
$\lambda^N(\theta)$	0.09	0.20	0.38	0.75	1.11
		$\theta = 0.50$			
$\lambda^P(\theta)$	0.06	0.21	0.45	0.94	1.42
$\lambda^N(\theta)$	0.06	0.11	0.19	0.35	0.51
		$\theta = 0.75$			
$\lambda^P(\theta)$	0.09	0.28	0.58	1.19	1.80
$\lambda^N(\theta)$	0.03	0.04	0.06	0.10	0.15

Obstfeld consumption process $\theta = 0.25$

γ		1	2.5	5	10	20
	λ	0.04	0.08	0.12	0.17	0.19
$\beta = 0.97$	$\lambda^P(\theta)$	0.01	0.03	0.05	0.07	0.08
	$\lambda^N(\theta)$	0.03	0.05	0.07	0.10	0.11
	λ	0.06	0.11	0.16	0.20	0.21
$\beta = 0.98$	$\lambda^P(\theta)$	0.02	0.04	0.06	0.08	0.09
	$\lambda^N(\theta)$	0.05	0.07	0.10	0.11	0.12
	λ	0.12	0.18	0.22	0.24	0.24
$\beta = 0.99$	$\lambda^P(\theta)$	0.03	0.07	0.09	0.10	0.10
	$\lambda^N(\theta)$	0.09	0.12	0.13	0.14	0.14

Estimation

- Several approaches:
  - Maximum likelihood estimation - either classical or using Bayesian methods
  - Generalized Method of Moments (GMM)
  - **Simulation-based methods:** SMM, EMM, Indirect Inference

Why shall we use simulation-based methods?

- ML
  - Requires the (closed-form) representation of the likelihood function
  - Often, models are too complex $\Rightarrow$ ML is intractable
  - *Can use Simulated Maximum Likelihood instead*
- GMM
  - Performs poorly in small sample (*post-war quarterly data is considered small sample!*)
  - Can not deal with unobserved variables
  - Often problems with identification and choice of orthogonality conditions

Simulation-Based Estimators

- **Main idea:** We choose the parameters of the model to minimize the distance between functions of actual and simulated data
- Those functions can be:
  - Moments such as variances, covariances, autocorrelations $\Rightarrow$ SMM
  - VAR coefficients, impulse response functions $\Rightarrow$ Indirect inference

Simulated Method of Moments

- Developed by McFadden (1989), Lee and Ingram (1991) ...
- Let θ denote the $K \times 1$ vector of structural parameters associated with our economic model
- In the DSGE case that would be $\theta = (\beta, \alpha, \rho, \delta, \ldots)$
- Let $\{y_t(\theta_0)\}_{t=1}^{T}$ be a sequence of **observed data** and let $\{\tilde{y}_t(\theta)\}_{t=1}^{T}$ be a set of **simulated data**
- θ_0 denotes the ""“true”"" but unobserved set of parameters
- The simulations are done by fixing θ and by using T draws of the stochastic element

Simulated Method of Moments

- Denote by $\mu[y_T]$ a $J \times 1$ vector of functions of the **observed data**
- ... and denote by $\mu[\tilde{y}_T]$ the same $J \times 1$ vector of functions computed using the **simulated time series** $\{\tilde{y}_t(\theta)\}$
- The SMM estimator is defined as

$$\hat{\theta}_T^S(W) = \arg\min_{\theta} \left[\mu[y_T] - \frac{1}{S}\sum_{s=1}^{S}\mu[y\tilde{(\theta)}_T]\right]' W_T \left[\mu[y_T] - \frac{1}{S}\sum_{s=1}^{S}\mu[y\tilde{(\theta)}_T]\right]$$

where W_T is a $J \times J$ symmetric and positive definite weighting matrix and S is the number of simulations

Properties of SMM

- Difference to GMM: Instead of calculating the theoretical moments $\mu[y_T]$ we approximate them with simulations

- Conditions needed:
- y_t e $\tilde{y}_t(\theta)$ are stationary, ergodic and independent
- Consistency: $\mu[y_T] \rightarrow^p \mu_y$ as $T \rightarrow \infty$ and $\mu[\tilde{y}_T^S(\theta)] \rightarrow^p \mu_{\tilde{y}}(\theta)$ as $S \rightarrow \infty$
- Identifiability: Under the null hypothesis that the model is 'true', there exists a θ^* such that $\mu_y = \mu_{\tilde{y}}(\theta^*)$
- Then $\sqrt{T}[\hat{\theta}_T^S(W) - \theta_0] \rightarrow \mathcal{N}(0, Q_S(W))$
- If $W_T = I$, $\hat{\theta}$ is consistent but inefficient

SMM Algorithm

1. Take the relevant data series and compute $\mu[y_T]$
2. Compute S histories of shocks
3. Guess θ^0 and solve the model for its equilibrium
4. Simulate the model S times and produce $\tilde{y}_t(\theta^0)$ and compute $\mu[\tilde{y}_T^S]$
5. Evaluate the objective function and get θ^1 using grid-search or gradient methods
6. Repeat the previous two steps until

$$\left[\mu[y_T] - \frac{1}{S}\sum_{s=1}^{S}\mu[y(\tilde{\theta}^i)_T]\right]' W_T \left[\mu[y_T] - \frac{1}{S}\sum_{s=1}^{S}\mu[y(\tilde{\theta}^i)_T]\right] < \varepsilon$$

$$\text{or } \|\theta^i - \theta^{i-1}\| < \varepsilon$$

Some Issues

- In step 4 we need to choose a sequence of shocks for the model
- IMPORTANT: Must use the same sequence of shocks during the iterations otherwise don't know if objective function changes because parameters change or shocks change
- How to get standard errors for $\hat{\theta}_T^S(W)$ Two alternatives:
  - Hessian of the objective function
  - Monte Carlo approach: Repeat algorithm for different sequences of the shocks, plot the histogram of the resulting θ and compute standard errors from this distribution

Example: Estimation of the New Keynesian Phillips curve

$$\pi_t = \beta \mathbb{E}_t \pi_{t+1} + \frac{(1-\eta)(1-\beta\eta)}{\eta} mc_t$$

- Em que mc_t are the marginal costs, η is tr probability of not changing the prices, β is the discount factor and π_t is the inflation rate
- Let's write the equation as

$$\pi_{t+1} = \frac{1}{\beta}\pi_t - \frac{(1-\eta)(1-\beta\eta)}{\beta\eta} mc_t + \varepsilon_{t+1}$$

- where ε is an expectation error and $\mathbb{E}_t(\varepsilon_{t+1}) = 0$

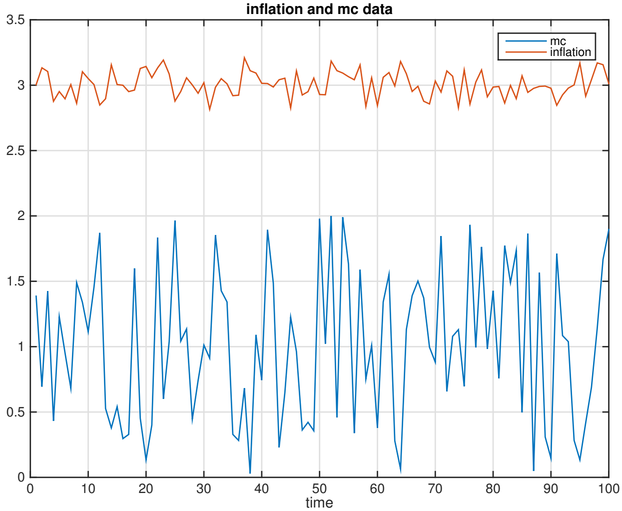
Example: Estimation of the New Keynesian Phillips curve

- Let's consider the following (data) moments in the estimation:

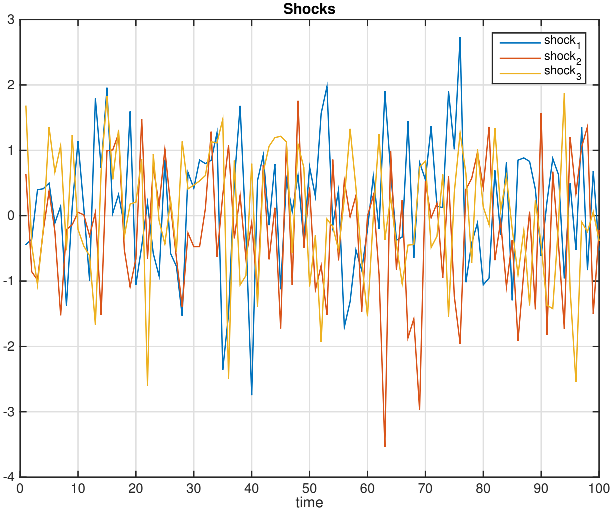
$$\mu[y_T] = \left[\frac{1}{T}\sum_t (\pi_{t+1}\pi_t), \quad \frac{1}{T}\sum_t (\pi_{t+2}\pi_t), \quad \frac{1}{T}\sum_t (\pi_{t+3}\pi_t) \right]'$$

- Compute S series of $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ with T periods
- Simulate $\tilde{y}_S^T(\theta)$ and compute $\mu[\tilde{y}_T^S(\theta)]$
- ...
- Matlab

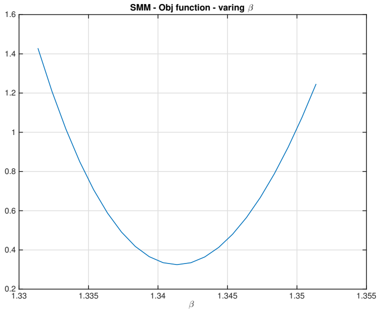
Data

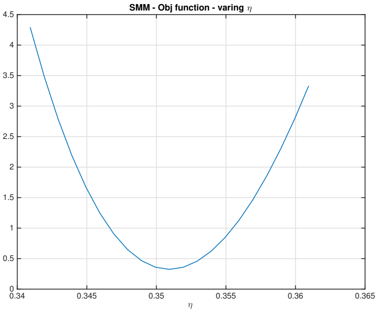

inflation and mc data
3.5
3
2.5
2
1.5
1
0.5
0
0
10
20
30
40
50
60
70
80
90
100
time
mc
inflation

Sample Shocks for Simulations

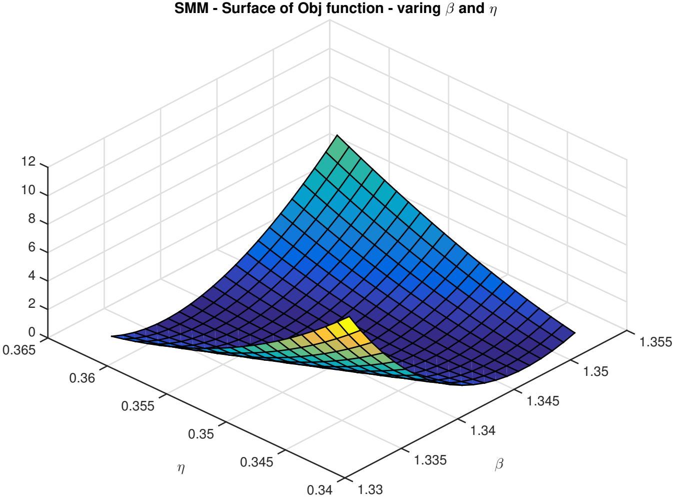


Objective Function - β and η

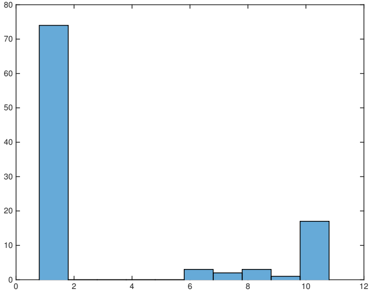




Objective Function - Surface



Cheating - Monte Carlo Simulation



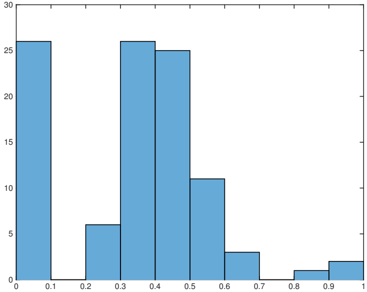


Figure: distribution of β and η

Monte Carlo Simulation - Cheating a lot

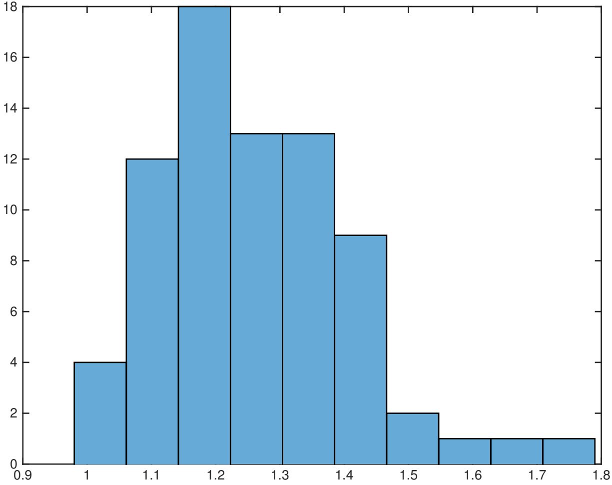


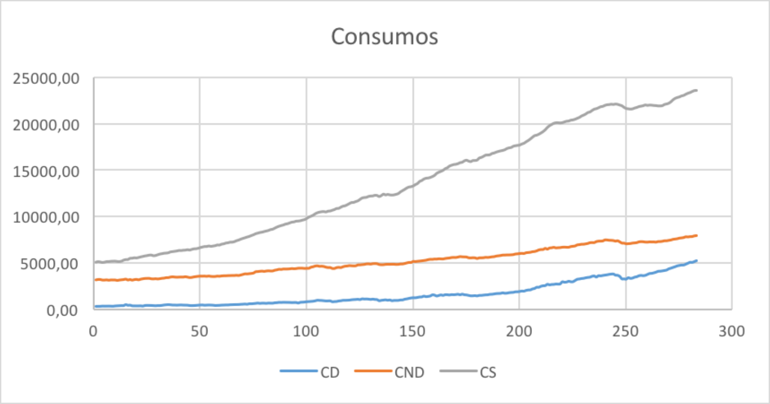
Figure: distribution of β

SMM applied to my research

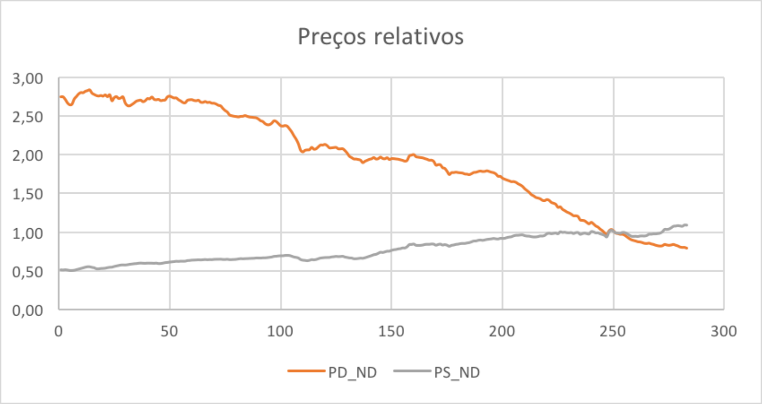
Barros and Gomes (2019): A study on the components of aggregate consumption

- Hall (1988): Inter and intratemporal elasticities influence the quantitative implications of various economic policy decisions
- Intratemporal substitution *indirectly* affects the measure of the intertemporal substitution
  - only for nonhomothetic preferences
- Ogaki and Reinhart (1998): nonhomotheticity is essential to obtain unbiased estimates of the intra and intertemporal substitutions.
- Implications for asset pricing
- The literature focuses on nondurables (+ services) and durables

Aggregate Consumption



Relative Prices



An Euler Equation

Hall's problem

$$E_t \left[\sum_{j=0}^{\infty} \beta^j u(c_{t+j}) \right],$$

s.t. a I.B.C. where

$$u(c) = \frac{c^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}}$$

We have the following Euler Equation

$$E_t \left[\beta \left(\frac{c_{t+1}}{c_t} \right)^{-\frac{1}{\sigma}} (1 + r_{t+1}) \right] = 1.$$

Nonhomothetic preferences

Our approach

$$u\left(C_t, S_t, K_t\right) = \frac{1}{1-1/\sigma} v\left(C_t, S_t, K_t\right)^{1-1/\sigma}$$

$$v\left(C_t, S_t, K_t\right) = \left[C_t^{1-\frac{1}{\theta}} + \alpha_s S_t^{1-\frac{\eta}{\theta}} + \alpha_k K_t^{1-\frac{\zeta}{\theta}}\right]^{\frac{\theta}{\theta-1}},$$

Thus, the Euler Equation can be represented as

$$\beta E_t \left[\left(\frac{\hat{f}\left(C_{t+1}, S_{t+1}, K_{t+1}\right)}{\hat{f}\left(C_t, S_t, K_t\right)} \right)^{\frac{\theta}{\theta-1}\frac{\sigma-1}{\sigma}-1} \left(\frac{C_{t+1}}{C_t}\right)^{-\frac{1}{\theta}} \left(1 + r_{t+1}\right) \right] = 1,$$

where

$$\hat{f}\left(C_t, S_t, K_t\right) = C_t^{1-\frac{1}{\theta}} + \alpha_s S_t^{1-\frac{\eta}{\theta}} + \alpha_k K_t^{1-\frac{\zeta}{\theta}}.$$

Estimation

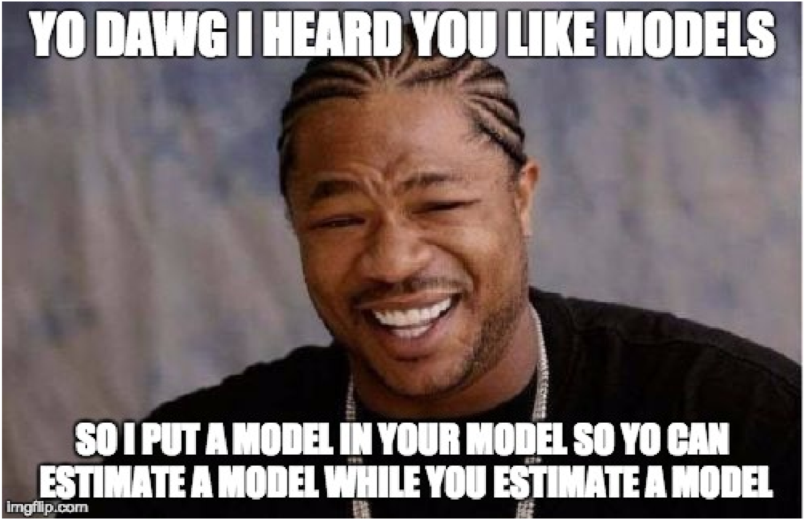
- GMM? Maybe not.
- SMM
  - 7 parameters
  - 7 or more target statistics (growth rates, ratios, etc...)
  - Partial equilibrium (use actual data for prices)
  - Income process: AR(1) [additional parameters]
- In progress...

Estimation

- GMM? Maybe not.
- SMM
  - 7 parameters
  - 7 or more target statistics (growth rates, ratios, etc...)
  - Partial equilibrium (use actual data for prices)
  - Income process: AR(1) [additional parameters]
- In progress...
- Prometo que o paper vai sair

Indirect Inference

- Developed by Gourioux et al. (1993) and Smith (1993)
- Indirect inference is a generalization of SMM where μ consists of generic continuous functions of the data
- It is particularly convenient when the model is complex $\Rightarrow$ likelihood is intractable
- The key feature of I.Inf is that it uses a simple auxiliary model to “measure the distance between the model and the data”


YO DAWG I HEARD YOU LIKE MODELS
SO I PUT A MODEL IN YOUR MODEL SO YO CAN
ESTIMATE A MODEL WHILE YOU ESTIMATE A MODEL
imgflip.com

- Main idea:
  - Auxiliary model is estimated both on the actual and on simulated data
  - I.Inf chooses the structural parameters so that the auxiliary parameters from the simulated data are as close as possible to those obtained on actual data.
- Choice of the auxiliary model is key! It has to be good!! More specifically:
  1. Has to provide a good statistical representation of the data
  2. Has to represent well the economic processes captured by the structural model
  3. Easy and quickly to compute

- Points (1)-(2) relate to the proper identification of structural parameters
- If (1)-(2) are not satisfied $\Rightarrow$ Objective function will not be responsive to variation in structural parameters
- Examples for auxiliary models:
  - Structural impulse response functions (DSGE literature)
  - Wage regressions and hazard rate models (Applied labor)
  - Discrete choice models (IO)

- Suppose $\Theta(y_t, \beta)$ is the auxiliary model at hand, where β is a vector of auxiliary parameters and y_t is the actual data
- Let $\hat{\beta}_T$ be estimator of β computed from the observed data
- The way $\hat{\beta}_T$ is computed, clearly depends on the structure of the auxiliary model: Impulse responses: SVAR estimation / Hazard model: ML estimation / etc.

- Notice that the null hypotheses underlying the estimation is that the observed data are generated by the model at the 'true' value of the parameter θ_0
- Given a set of structural parameters q we solve and simulate the model to get artificial data $\tilde{y}_T(\theta)$
- The auxiliary model is then estimated out of the simulated data. The associated estimator of the auxiliary parameters is $\hat{\beta}(\tilde{y}_T(\theta))$

- For the same θ we repeat the last step S times and compute

$$\hat{\beta}_S(\theta) = \frac{1}{S} \sum_s \hat{\beta}(\tilde{y}_T(\theta))$$

- The averaging is done to minimize the error due to changing draws
- The indirect inference estimator of the model's structural parameters is

$$\hat{\theta}_T^S(W) = \arg\min_\theta \left[\hat{\beta}_T - \hat{\beta}_S(\theta)\right]' \times W \times \left[\hat{\beta}_T - \hat{\beta}_S(\theta)\right]$$

Some issues

- How can we check whether our auxiliary model is a good one?
- Gourieroux et al. (1993) propose a global specification test based on the value of the objective function
- The test statistic is given by

$$\zeta_T = \frac{TS}{1+S} \min_{\theta} \left[\hat{\beta}_T - \hat{\beta}_S(\theta)\right]' \times W \times \left[\hat{\beta}_T - \hat{\beta}_S(\theta)\right]$$

- which is asymptotically χ^2 with dim(β) - dim(θ) degrees of freedom
- Hypotheses tests can be carried out in the standard fashion using Wald or Lagrange multiplier tests (see Gourieroux et al. (1993))
- Auxiliary parameters are conditional moments of the data are (often) of interest as well. Especially in wage regressions and duration models

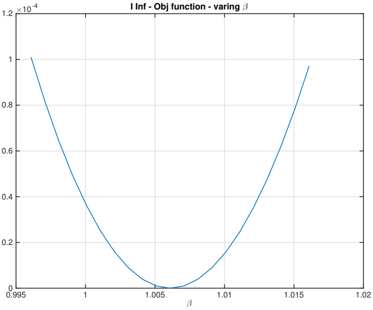
Example: NK Phillips curve continued

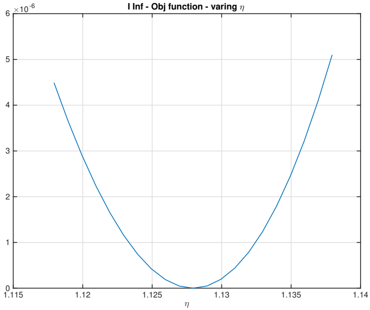
- Let's consider our previous example. How to estimate (β, η) by I.Inf?
- Step 1: Choose an auxiliary model. We take a reduced form equation

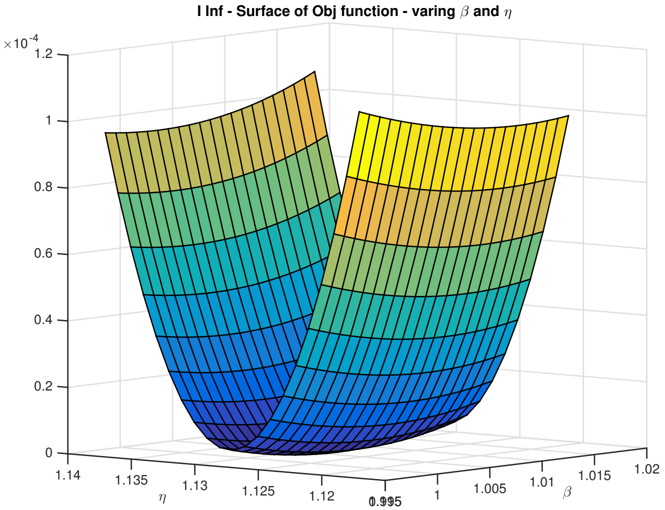
$$\pi_{t+1} = a_1 \pi_t + a_2 mc_t + \varepsilon_{t+1}$$

- and estimate a1 and a2 using actual data
- Step 2: Simulate structural model to obtain artificial data $\tilde{\pi}_t$ and $\tilde{mc}_t$
- Step 3: Estimate the auxiliary model using $\tilde{\pi}_t$ and $\tilde{mc}_t$ and get $\tilde{a}_1(\beta, \eta)$ and $\tilde{a}_2(\beta, \eta)$
- Estimate (β, η) by minimizing

$$[a_1 - \tilde{a}_1(\beta, \eta)]\, W\, [a_1 - \tilde{a}_1(\beta, \eta)]'$$


I Inf - Obj function - varing β
$\times 10^{-4}$
1.2
1
0.8
0.6
0.4
0.2
0
0.995
1
1.005
1.01
1.015
1.02
β


I Inf - Obj function - varing η
$\times 10^{-6}$
6
5
4
3
2
1
0
1.115
1.12
1.125
1.13
1.135
1.14
η


I Inf - Surface of Obj function - varing β and η
$\times 10^{-4}$
1.2
1
0.8
0.6
0.4
0.2
0
1.14
1.135
1.13
1.125
1.12
1.115
η
0.995
1
1.005
1.01
1.015
1.02
β

Summary

```
****************************************
                Results

         OLS      SMM          IF
****************************************
beta   1.0061    1.3414     1.0061
 eta   0.8812    0.3509     1.1279
****************************************
```

Identification

- Let $g(\theta)$ be the distance function that we seek to minimize, i.e.

$$\hat{\theta} = \arg\min_{\theta} g(\theta)$$

- **Identification**: Under what conditions can we recover the structural parameters θ?
- Only if the mapping from the distance function to the parameters is well behaved
- Need:
  - $g(\theta)$ has a unique minimum at $\theta = \theta_0$
  - Hessian is positive definite and has full rank
  - Curvature of $g(\theta)$ is "sufficient'

Identification

- Different objective functions may have different "identification power"
- With complex models (the ones we typically work with), the distance function is a non-linear function of the parameters $\Rightarrow$ too complicated to be worked out analytically
- $\Rightarrow$ Identifiability of θ is problematic Important cases:
  - Observational equivalence: $g(y, T, m_1, \theta_1^*) = g(y, T, m_2, \theta_2^*)$
  - Under-identification: $g(y, T, m, \theta) = \text{constant } \forall \theta$
  - Weak identification: $g(y, T, m, \theta^*) = 0$ and $g(y, T, m, \theta) < \varepsilon$, $\forall \theta$
- m_1 and θ_1 respectively denote model specification i and the associated optimal estimator

Identification

- Under-identification: Certain parameter(s) disappear(s) from the solution) distance function is not responsive to changes in that parameter(s)
- Note (DSGE): Different shocks identify different parameters
- Weak and partial under-identification. Consequences:
  - Algorithm remains stuck at initial conditions
  - Estimates could be random
  - Parameter estimates are inconsistent, asymptotic distribution is non-normal and the standard tests are incorrect!

Identification

- What causes the problems?
- Example DGSE: Law of motion of capital stock is almost invariant to variations of α (capital share) and ρ (AR coefficient)
- Can we reduce problems by
  - Change T - long horizon may have little information
  - DSGE: Match VAR coefficients
  - Alter the objective function
  - Choose different estimation approaches

How to detect identification problems ?

- Ex-post diagnostics
  - Erratic parameter estimates as T increases
  - Large or non-computable standard errors
  - Crazy t-statistics
- General diagnostics:
  - Plot objective function (around estimated values)
  - Condition number of the Hessian (ratio of the largest to the smallest eigenvalue)
- Testing rank of Hessian. Check Cragg and Donald (1997)