Estimating Macroeconomic Models: A Likelihood Approach

J. Fernandez Villaverde and JF. Rubio Ramirez

Presented by Diego Anzoategui

April 21, 2014

Introduction

- Most DSGE models do not imply a likelihood function (or posterior) that can be easily evaluated analytically or numerically.
- One way of dealing with this problem is to linearize the model and get an approximation to the likelihood.
- However, there are contexts in which a linearization can generate considerable approximation errors.
- This errors accumulate period by period: the longer the sample, the larger the error.

Introduction

- This paper presents one way of evaluating the likelihood (or posterior kernel) of non-linear and/or non-normal macroeconomic models: **Particle Filter (PF)**.

- General idea:
  1. For a given set of parameters, solve the equilibrium of the model using any non-linear solution method;
  2. With the solution in hand, construct a state space representation (define the state and measurement equations);
  3. Use particle filter to evaluate the likelihood function (posterior kernel) of the model.
  4. Find estimates of parameters through ML or simulating the posterior using a MCMC routine.

Model

- Assume that we have a model with the following state-space representation.

$$S_t = f\left(S_{t-1}, W_t; \gamma\right)$$

$$Y_t = g\left(S_t, V_t; \gamma\right)$$

  where S_t are the states, Y_t are observables, and W_t and V_t are shocks that are independent from each other.

- Note that most macroeconomic models do not admit closed-form solutions for functions f and g. The PF only requires a numerical procedure to approximate them.

Likelihood function

- We want $p\left(\mathcal{Y}^T;\gamma\right)$. where $\mathcal{Y}_t$ denotes a realization of Y_t and $\mathcal{Y}^T = \{\mathcal{Y}_t\}_{t=1}^T$. Using the prediction error decomposition,

$$p\left(\mathcal{Y}^T;\gamma\right) = \prod_{t=1}^{T} p\left(\mathcal{Y}_t|\mathcal{Y}^{t-1};\gamma\right)$$

- Note the following,

$$\begin{aligned} p\left(\mathcal{Y}_t|\mathcal{Y}^{t-1};\gamma\right) &= \int p\left(\mathcal{Y}_t, S_t|\mathcal{Y}^{t-1};\gamma\right) dS_t \\ &= \int p\left(\mathcal{Y}_t|S_t;\gamma\right) p\left(S_t|\mathcal{Y}^{t-1};\gamma\right) dS_t \end{aligned}$$

Assumptions

Assumption 1

There exists a partition of $\{W_t\}$ into two sequences $\{W_{1,t}\}$ and $\{W_{2,t}\}$, such that $W_t = (W_{1,t}, W_{2,t})$ and $dim(W_{2,t}) + dim(V_t) \geq dim(Y_t)$ for all t.

Assumptions

Assumption 1

There exists a partition of $\{W_t\}$ into two sequences $\{W_{1,t}\}$ and $\{W_{2,t}\}$, such that $W_t = (W_{1,t}, W_{2,t})$ and $dim(W_{2,t}) + dim(V_t) \geq dim(Y_t)$ for all t.

Assumption 1 (for these slides)

$dim(V_t) \geq dim(Y_t)$ for all t, which implies $W_{1,t} = W_t$

Assumptions

Assumption 2

For all γ, S^t realization $s^t = \{s_0, s_1, .., s_t\}$ and t, the system,

$$\mathcal{Y}_m = g\left(s_m, V_m; \gamma\right) \text{ for } m = 1, 2, .., t$$

has a unique solution $\nu^t(s^t, \mathcal{Y}^t; \gamma)$, where $\nu_t = g^{-1}(s_t, \mathcal{Y}_t; \gamma)$.

Assumptions

Assumption 2

For all γ, S^t realization $s^t = \{s_0, s_1, .., s_t\}$ and t, the system,

$$\mathcal{Y}_m = g\left(s_m, V_m; \gamma\right) \text{ for } m = 1, 2, .., t$$

has a unique solution $\nu^t(s^t, \mathcal{Y}^t; \gamma)$, where $\nu_t = g^{-1}(s_t, \mathcal{Y}_t; \gamma)$.

- Note that this implies that we can evaluate $p\left(\mathcal{Y}_t | S_t; \gamma\right)$ by a change of variables.

$$p\left(\mathcal{Y}_t | S_t; \gamma\right) = p(\nu_t; \gamma)|J_g|^{-1}$$

Assumptions

Assumption 3

For all $\gamma \in \Gamma$, s^t, and t, the model gives some positive probability to the data $\mathcal{Y}^T$,that is,

$$p\left(\mathcal{Y}_t | s_t; \gamma\right) > 0$$

for all $\gamma \in \Gamma$, s^t, and t.

Assumptions

Assumption 3

For all $\gamma \in \Gamma$, s^t, and t, the model gives some positive probability to the data $\mathcal{Y}^T$,that is,

$$p\left(\mathcal{Y}_t | s_t; \gamma\right) > 0$$

for all $\gamma \in \Gamma$, s^t, and t.

- Assumption 1 and 3 are jointly a sufficient condition for the model not to be stochastically singular.

A particle filter

$$p\left(\mathcal{Y}_t|\mathcal{Y}^{t-1};\gamma\right) = \int p\left(\mathcal{Y}_t|S_t;\gamma\right) p\left(S_t|\mathcal{Y}^{t-1};\gamma\right) dS_t$$

- By Assumption 2, $p\left(\mathcal{Y}_t|S_t;\gamma\right)$ is easy to evaluate.
- Hence, conditional on having N evenly weighted draws $\{\hat{s}^i_{t|t-1}\}_{i=1}^N$ from $p\left(S_t|\mathcal{Y}^{t-1};\gamma\right)$,

$$p(\mathcal{Y}_t|\mathcal{Y}^{t-1}) \simeq \frac{1}{N}\sum_{i=1}^{N} p(\mathcal{Y}_t|\hat{s}^i_{t|t-1};\gamma)$$

- A particle filter specifies a way of getting these draws (or particles) $\hat{s}^i_{t|t-1}$.

A particle filter

$$p\left(\mathcal{Y}_t|\mathcal{Y}^{t-1};\gamma\right) = \int p\left(\mathcal{Y}_t|S_t;\gamma\right) p\left(S_t|\mathcal{Y}^{t-1};\gamma\right) dS_t$$

- By Assumption 2, $p\left(\mathcal{Y}_t|S_t;\gamma\right)$ is easy to evaluate.
- Hence, conditional on having N evenly weighted draws $\{\hat{s}^i_{t|t-1}\}_{i=1}^N$ from $p\left(S_t|\mathcal{Y}^{t-1};\gamma\right)$,

$$p(\mathcal{Y}_t|\mathcal{Y}^{t-1}) \simeq \frac{1}{N}\sum_{i=1}^{N} p(\mathcal{Y}_t|\hat{s}^i_{t|t-1};\gamma)$$

- A particle filter specifies a way of getting these draws (or particles) $\hat{s}^i_{t|t-1}$.

A particle filter

$$p\left(\mathcal{Y}_t|\mathcal{Y}^{t-1};\gamma\right) = \int p\left(\mathcal{Y}_t|S_t;\gamma\right) p\left(S_t|\mathcal{Y}^{t-1};\gamma\right) dS_t$$

- By Assumption 2, $p\left(\mathcal{Y}_t|S_t;\gamma\right)$ is easy to evaluate.
- Hence, conditional on having N evenly weighted draws $\{\hat{s}^i_{t|t-1}\}_{i=1}^N$ from $p\left(S_t|\mathcal{Y}^{t-1};\gamma\right)$,

$$p(\mathcal{Y}_t|\mathcal{Y}^{t-1}) \simeq \frac{1}{N}\sum_{i=1}^{N} p(\mathcal{Y}_t|\hat{s}^i_{t|t-1};\gamma)$$

- A particle filter specifies a way of getting these draws (or particles) $\hat{s}^i_{t|t-1}$.

A particle filter

- We want draws from $p\left(S_t|\mathcal{Y}^{t-1};\gamma\right)$ for every t.
- Notation.
  Let $\{\hat{s}^i_{t-1}\}^N_{i=1}$ be a sequence of N iid draws from $p\left(S_{t-1}|\mathcal{Y}^{t-1};\gamma\right)$.
  Let $\{\hat{s}^i_{t|t-1}\}^N_{i=1}$ be a sequence of draws from $p\left(S_t|\mathcal{Y}^{t-1};\gamma\right)$.
- The filter in this paper is a sampling/importance resampling (SIR) algorithm.

A particle filter

- It consists of the following steps,

Step 0 **Initialization**. Set $t = 1$, sample $\{\hat{s}_0^i\}_{i=1}^N$ from $p(S_{t-1}|\mathcal{Y}^{t-1};\gamma) = p(S_0;\gamma)$.

Step 1 **Prediction**. Sample $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ using the state equation and the sample $\{\hat{s}_{t-1}^i\}_{i=1}^N$.

Step 2 **Filtering**. Assign to each draw $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ a weight q_i^t (to be defined).

Step 3 **Sampling**. Sample N times from $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ with replacement and probabilities $\{q_i^t\}_{i=1}^N$. Call the new sample $\{\hat{s}_t^i\}_{i=1}^N$: they are draws from $p(S_t|\mathcal{Y}^t;\gamma)$. Go to step 1 if $t < T$ and set t=t+1.

A particle filter

- It consists of the following steps,

Step 0 **Initialization**. Set $t = 1$, sample $\{\hat{s}_0^i\}_{i=1}^N$ from $p(S_{t-1}|\mathcal{Y}^{t-1};\gamma) = p(S_0;\gamma)$.

Step 1 **Prediction**. Sample $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ using the state equation and the sample $\{\hat{s}_{t-1}^i\}_{i=1}^N$.

Step 2 **Filtering**. Assign to each draw $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ a weight q_i^t (to be defined).

Step 3 **Sampling**. Sample N times from $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ with replacement and probabilities $\{q_i^t\}_{i=1}^N$. Call the new sample $\{\hat{s}_t^i\}_{i=1}^N$: they are draws from $p(S_t|\mathcal{Y}^t;\gamma)$. Go to step 1 if $t < T$ and set t=t+1.

A particle filter

- It consists of the following steps,

Step 0 **Initialization**. Set $t = 1$, sample $\{\hat{s}_0^i\}_{i=1}^N$ from $p(S_{t-1}|\mathcal{Y}^{t-1};\gamma) = p(S_0;\gamma)$.

Step 1 **Prediction**. Sample $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ using the state equation and the sample $\{\hat{s}_{t-1}^i\}_{i=1}^N$.

Step 2 **Filtering**. Assign to each draw $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ a weight q_i^t (to be defined).

Step 3 **Sampling**. Sample N times from $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ with replacement and probabilities $\{q_i^t\}_{i=1}^N$. Call the new sample $\{\hat{s}_t^i\}_{i=1}^N$: they are draws from $p(S_t|\mathcal{Y}^t;\gamma)$. Go to step 1 if $t < T$ and set t=t+1.

A particle filter

- It consists of the following steps,

Step 0 **Initialization**. Set $t = 1$, sample $\{\hat{s}_0^i\}_{i=1}^N$ from $p(S_{t-1}|\mathcal{Y}^{t-1};\gamma) = p(S_0;\gamma)$.

Step 1 **Prediction**. Sample $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ using the state equation and the sample $\{\hat{s}_{t-1}^i\}_{i=1}^N$.

Step 2 **Filtering**. Assign to each draw $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ a weight q_i^t (to be defined).

Step 3 **Sampling**. Sample N times from $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ with replacement and probabilities $\{q_i^t\}_{i=1}^N$. Call the new sample $\{\hat{s}_t^i\}_{i=1}^N$: they are draws from $p(S_t|\mathcal{Y}^t;\gamma)$. Go to step 1 if $t < T$ and set t=t+1.

A particle filter

- It consists of the following steps,

Step 0 **Initialization**. Set $t = 1$, sample $\{\hat{s}_0^i\}_{i=1}^N$ from $p(S_{t-1}|\mathcal{Y}^{t-1};\gamma) = p(S_0;\gamma)$.

Step 1 **Prediction**. Sample $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ using the state equation and the sample $\{\hat{s}_{t-1}^i\}_{i=1}^N$.

Step 2 **Filtering**. Assign to each draw $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ a weight q_i^t (to be defined).

Step 3 **Sampling**. Sample N times from $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ with replacement and probabilities $\{q_i^t\}_{i=1}^N$. Call the new sample $\{\hat{s}_t^i\}_{i=1}^N$: they are draws from $p(S_t|\mathcal{Y}^t;\gamma)$. Go to step 1 if $t < T$ and set t=t+1.

Step 1: Prediction

- We start this step with a sample $\{\hat{s}_{t-1}^i\}_{i=1}^N$ from $p(S_{t-1}|\mathcal{Y}^{t-1};\gamma)$
- It is easy to get a sample $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$.
- Let ω_t^i be a draw from W_t's distribution. Then, a draw $\hat{s}_{t|t-1}^i$ is defined by,

$$\hat{s}_{t|t-1}^i = f(\hat{s}_{t-1}^i, \omega_t^i)$$

- $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ is a sample from $p(S_t|\mathcal{Y}^{t-1};\gamma)$. Remember,

$$p(S_t|\mathcal{Y}^{t-1};\gamma) = \int p(S_t|S_{t-1};\gamma)p(S_{t-1}|\mathcal{Y}^{t-1};\gamma)dS_{t-1}$$

Step 2: Filtering

- We have a sample $\{\hat{s}^i_{t|t-1}\}_{i=1}^N$ from $p(S_t|\mathcal{Y}^{t-1};\gamma)$ from the previous step.
- The aim of step 2 and 3 is to draw a sample $\{\hat{s}^i_t\}_{i=1}^N$ from $p(S_t|\mathcal{Y}^t;\gamma)$.
- By Bayes Law,

$$p(S_t|\mathcal{Y}^t;\gamma) \propto p(S_t|\mathcal{Y}^{t-1};\gamma)p(\mathcal{Y}_t|S_t,\mathcal{Y}^{t-1};\gamma)$$

- Hence, drawing a sample from $p(S_t|\mathcal{Y}^t;\gamma)$ is similar to drawing from $p(S_t|\mathcal{Y}^{t-1};\gamma)$ but using importance weights proportional to $p(\mathcal{Y}_t|S_t,\mathcal{Y}^{t-1};\gamma)$.

Step 2: Filtering

- Note the following,

$$p(S_t|\mathcal{Y}^t;\gamma) = \frac{p(S_t|\mathcal{Y}^t;\gamma)}{p(S_t|\mathcal{Y}^{t-1};\gamma)} p(S_t|\mathcal{Y}^{t-1};\gamma)$$

  And we also know that,

$$\frac{p(S_t|\mathcal{Y}^t;\gamma)}{p(S_t|\mathcal{Y}^{t-1};\gamma)} \propto p(\mathcal{Y}_t|S_t,\mathcal{Y}^{t-1};\gamma)$$

Step 2: Filtering

- But we already have a sample from $p(S_t|\mathcal{Y}^{t-1};\gamma)$ which came from step 1: $\{\hat{s}^i_{t|t-1}\}_{i=1}^N$.

- We can easily compute the weights proportional to $p(\mathcal{Y}_t|S_t,\mathcal{Y}^{t-1};\gamma)$.

$$q_t^i = \frac{p(\mathcal{Y}_t|\hat{s}^i_{t|t-1},\mathcal{Y}^{t-1};\gamma)}{\sum_{i=1}^N p(\mathcal{Y}_t|\hat{s}^i_{t|t-1},\mathcal{Y}^{t-1};\gamma)}$$

- Next step applies the computed weights to get a sample from $p(S_t|\mathcal{Y}^t;\gamma)$.

Step 3: Sampling

- Rubin (1988) proposed the method applied in this step to draw from $p(S_t|\mathcal{Y}^t;\gamma)$ using weights q_t^i.

Corollary 1

Given a draw $\{\hat{s}^i_{t|t-1}\}_{i=1}^N$, let the sequence $\{\tilde{s}^i\}_{i=1}^N$ be a draw with replacement from $\{\hat{s}^i_{t|t-1}\}_{i=1}^N$ where q_t^i is the probability of $\hat{s}^i_{t|t-1}$ being drawn for all i. Then $\{\tilde{s}^i\}_{i=1}^N$ is a draw from $p(S_t|\mathcal{Y}^t;\gamma)$.

- Can use multinomial resampler to get $\{\tilde{s}^i\}_{i=1}^N = \{\hat{s}^i_t\}_{i=1}^N$.

- With $\{\hat{s}^i_t\}_{i=1}^N$ move to Step 1 again, get $\{\hat{s}^i_{t+1|t}\}_{i=1}^N$, and repeat until $t = T$.

A particle filter

- It consists of the following steps,

Step 0 **Initialization**. Set $t = 1$, sample $\{\hat{s}_0^i\}_{i=1}^N$ from $p(S_{t-1}|\mathcal{Y}^{t-1};\gamma) = p(S_0;\gamma)$.

A particle filter

- It consists of the following steps,

Step 0 **Initialization**. Set $t = 1$, sample $\{\hat{s}_0^i\}_{i=1}^N$ from $p(S_{t-1}|\mathcal{Y}^{t-1};\gamma) = p(S_0;\gamma)$.

Step 1 **Prediction**. Sample $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ using the state equation and the sample $\{\hat{s}_{t-1}^i\}_{i=1}^N$.

Step 2 **Filtering**. Assign to each draw $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ a weight q_i^t.

Step 3 **Sampling**. Sample N times from $\{\hat{s}_{t|t-1}^i\}_{i=1}^N$ with replacement and probabilities $\{q_i^t\}_{i=1}^N$. Call the new sample $\{\hat{s}_t^i\}_{i=1}^N$: they are draws from $p(S_t|\mathcal{Y}^t;\gamma)$. Go to step 1 if $t < T$ and set t=t+1.

Computing the Likelihood

- After this process we end up with $\{\{\hat{s}^i_{t|t-1}\}_{i=1}^N\}_{t=0}^T$.
- It is easy to estimate the likelihood,

$$p(\mathcal{Y}^T|\gamma) \simeq \prod_{t=0}^{T} \frac{1}{N} \sum_{i=1}^{N} p(\mathcal{Y}_t|\hat{s}^i_{t|t-1};\gamma)$$

Pitfalls

- The process is easy and fast.
- However, it has some problems.
  1. Sample impoverishment or weight degeneracy: the sampling step reduces the effective sample size.
  2. Blind proposal. Step 1 ignores information on S_t contained in $\dagger_t$. This could be inefficient: we might potentially need many blind proposals to be in the right part of the likelihood/posterior.
- Good idea to start with a high N.
- Might want to improve blind proposals through an important sampler (Pitt and Shephard 2001).

Estimation: MLE

- Follow the next steps to compute the MLE estimator.

Step 0 Initialization: Set $i = 0$ and an initial γ_i. Set i=i+1

Step 1 Solve the model: for γ_i and compute $f(.,.;\gamma_i)$ and $g(.,.;\gamma_i)$.

Step 2 Evaluating the Likelihood: using the PF compute $p(\mathcal{Y}^T;\gamma_i)$ and get γ_{i+1} from a maximization routine.

Step 3 Stopping rule: If $\|p(\mathcal{Y}^T;\gamma_i) - p(\mathcal{Y}^T;\gamma_{i+1})\| > \xi$ set $i = i + 1$ and go to step 1.

Estimation: MLE

- We need twice differentiability to: (i) be able to get the Hessian that is actually related to $var(\hat{\gamma}_{MLE})$, and (ii) use a gradient based method to do the maximization.
- Lack of differentiability is likely.
- They avoid the second problem by using simulated annealing.

Estimation: MCMC

- Follow the next steps to compute the posterior distribution: $\pi(\gamma|\mathcal{Y}^T) \propto p(\mathcal{Y}^T|\gamma)\pi(\gamma)$.

Step 0 Initialization: Set $i = 0$ and an initial γ_i.Solve the model for γ_i and compute $f(.,.;\gamma_i)$ and $g(.,.;\gamma_i)$. Evaluate the prior $\pi(\gamma_i)$ and approximate $p(\mathcal{Y}^T;\gamma_i)$. Set i=i+1.

Step 1 Proposal draw: Get a draw γ_i^* from a proposal density $q(\gamma_{i-1}, \gamma_i^*)$.

Step 2 Solve the model for γ_i^* and compute $f(.,.;\gamma_i^*)$ and $g(.,.;\gamma_i^*)$.

Step 3 Evaluating the proposal. Get $\pi(\gamma_i^*)$ and $p(\mathcal{Y}^T;\gamma_i^*)$.

Step 4 Accept/Reject. Draw χ from $U[0,1]$. If $\chi \leq \frac{p(\mathcal{Y}^T;\gamma_i^*)\pi(\gamma_i^*)}{p(\mathcal{Y}^T;\gamma_{i-1})\pi(\gamma_{i-1})}$ set $\gamma_i = \gamma_i^*$, otherwise $\gamma_i = \gamma_{i-1}$. Set i=i+1 and go to step 1.

Computation

- The authors estimate a model with 8 states.
- Using PF with $80,000$ particles, each likelihood evaluation takes 12 sec.
- Kalman filter takes a fraction of a second.