Estimation of nonlinear models with Dynare

stephane.adjemian@univ-lemans.fr

September, 2023

Introduction

- Full information estimation of models approximated at higher order...
- Or, in principle, full information estimation of nonlinear models.
- Cannot use the (linear) Kalman filter anymore.
- Dynare provides routines to evaluate the likelihood of models approximated at order $k \geq 1$.

The reduced form model

$$
\begin{aligned}
s_t &= f(s_{t-1}, \varepsilon_t; \boldsymbol{\theta}) \\
y_t &= g(s_t; \boldsymbol{\theta}) + e_t
\end{aligned}
$$

- $f(.)$ and $g(.)$ are the state and measurement equations.
- $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^m$ a vector of m parameters.
- Cannot use the (linear) Kalman filter anymore.
- s_t and y_t are the vectors of state variables and observed variables.
- Innovations ε_t and e_t are the structural shocks and measurement errors.
- $\#(y_t) = \#(e_t)$

Reduced form with second order approximation

- The "state" equations:

$$
\begin{aligned}
s_t = \bar{s}(\boldsymbol{\theta}) &+ g_u(\boldsymbol{\theta})\hat{s}_{t-1} + g_u(\boldsymbol{\theta})\varepsilon_t \\
&+ 0.5 g_{\sigma\sigma}(\boldsymbol{\theta}) \\
&+ 0.5 g_{yy}(\boldsymbol{\theta})\left(\hat{s}_{t-1} \otimes \hat{s}_{t-1}\right) \\
&+ 0.5 g_{uu}(\boldsymbol{\theta})\left(\varepsilon_t \otimes \varepsilon_t\right) \\
&+ 0.5 g_{uy}(\boldsymbol{\theta})\left(\hat{s}_{t-1} \otimes \varepsilon_t\right)
\end{aligned}
$$

- The measurement equations:

$$
y_t = Z s_t + e_t
$$

where Z is a selection matrix.

Properties of the state space model

- $s_t \sim$ first order Markov process:

$$p\left(s_t | s_{0:t-1}\right) = p(s_t \, | s_{t-1})$$

- Observations are conditionally independent:

$$p\left(y_t | y_{1:t-1}, s_{0:t}\right) = p\left(y_t | s_t\right)$$

$\Rightarrow$ We cannot evaluate the likelihood, product of $p(y_t | y_{1:t-1})$ densities, without tracking the state variables.

Nonlinear filter (I)

- Suppose $p(y_{t-1}|s_{t-1})$ is known $\longrightarrow$ How to compute $p(y_t|s_t)$?

- First we can predict the states in t given information in $t-1$:

$$p\left(s_t|y_{1:t-1}\right) = \int p(s_t|s_{t-1}) p\left(s_{t-1}|y_{1:t-1}\right) \mathrm{d}s_{t-1} \tag{1}$$

  where $p(s_t|s_{t-1})$ is defined by the state equations:

$$\begin{aligned} p\left(s_t|s_{t-1}\right) &= \int p\left(s_t|s_{t-1},\varepsilon_t\right) p\left(\varepsilon_t|s_{t-1}\right) \mathrm{d}\varepsilon_t \\ &= \int p\left(s_t|s_{t-1},\varepsilon_t\right) p\left(\varepsilon_t\right) \mathrm{d}\varepsilon_t \end{aligned}$$

Nonlinear filter (II)

- $p\left(s_t|y_{1:t-1}\right)$ can be interpreted as our prior belief about the state variables.

- Use Bayes theorem to update our beliefs:

$$p\left(s_t|y_{1:t}\right) = \frac{p\left(y_t|s_t\right)p\left(s_t|y_{1:t-1}\right)}{p\left(y_t|y_{1:t-1}\right)} \tag{2}$$

  where:

$$p\left(y_t|y_{1:t-1}\right) = \int p\left(y_t|s_t\right)p\left(s_t|y_{1:t-1}\right)\mathrm{d}s_t$$

  which is the conditional density required to evaluate the likelihood of the model.

- $p\left(y_t|s_t\right)$, likelihood of s_t, is defined by the measurement equations.

$\Rightarrow$ Iterate over (1) and (2).

Approximated nonlinear filter (I)

Approximate the distribution of $s_t|y_{1:t}$

- Suppose the distribution of $s_t|y_{1:t}$ can be accurately approximated by a set of nodes and weights $\left\{\left(s_t^i, \omega_t^i\right)\right\}_{i=1}^N$, with non negative weights summing up to one.

$\Rightarrow$ When $N \to \infty$ $\sum_{i=1}^N \omega_i h(s_t^i)$ converges to $\mathbb{E}_{p(s_t|y_{1:t})}[h(s)]$.

- Can be used to approximate the predictive densities:

$$\widehat{p}\left(s_t|y_{1:t-1}\right) = \sum_{i=1}^N p(s_t|s_{t-1}^i)\omega_{t-1}^i$$

- Particles $\leftarrow$ random deviates from $s_t|y_{1:t}$.
- How to sample from $s_t|y_{1:t}$?

Approximated nonlinear filter (II)

Importance sampling (a)

- Suppose we can sample from $q\left(s_t|y_{1:t}\right)$ but not from $q\left(s_t|y_{1:t}\right)$.
- Then:

$$\mathbb{E}_{p(s_t|y_{1:t})}\left[h(s)\right] = \int \underbrace{\frac{p\left(s_t|y_{1:t}\right)}{q\left(s_t|y_{1:t}\right)}}_{\widetilde{w}_t(s_t)} q\left(s_t|y_{1:t}\right) h(s_t) \mathrm{d}s_t$$

$$= \mathbb{E}_{q(s_t|y_{1:t})}\left[\widetilde{\omega}_t(s_t) h(s_t)\right]$$

- By the Bayes theorem:

$$p\left(s_t|y_{1:t}\right) = \frac{p\left(y_{1:t}|s_t\right) p\left(s_t\right)}{p\left(y_{1:t}\right)}$$

- Unormalized weights:

$$\widehat{\omega}_t(s_t) = \frac{p\left(y_{1:t}|s_t\right) p\left(s_t\right)}{q\left(s_t|y_{1:t}\right)} = p\left(y_{1:t}\right) \widetilde{\omega}_t(s_t)$$

Approximated nonlinear filter (II)

Importance sampling (b)

With particles:

$$\hat{\omega}_t^i = \frac{p\left(y_{1:t}|s_t^i\right)p\left(s_t^i\right)}{q\left(s_t^i|y_{1:t}\right)}$$

Then

$$\widehat{\mathbb{E}}_{p(s_t|y_{1:t})}\left[h(s)\right] = \sum_{i=1}^{N} \widetilde{\omega}_i h(s_t^i)$$

where

$$\widetilde{\omega}_i = \frac{\widehat{\omega}_i}{\sum_{i=1}^{N}\widehat{\omega}_i}$$

Approximated nonlinear filter (III)

Sequential importance sampling

Let's pick a proposal q such that:

$$q\left(s_t|y_{1:t}\right) = q\left(s_t|s_{t-1}, y_t\right) q(s_{t-1}|y_{1:t-1})$$

Then, one can show that:

$$\hat{w}_t(s_t) = \hat{w}_{t-1}(s_{t-1}) \frac{p\left(y_t|s_t\right) p\left(s_t|s_{t-1}\right)}{q\left(s_t|s_{t-1}, y_t\right)}$$

A particle's current weight depends on its previous level and an incremental weight. This increment will put more weight on a particle if she is more likely with respect to its history and the current observation.

Approximated nonlinear filter (III)

Generic particle filter algorithm

$\left\{s_0^i, w_0^i\right\}_{i=1}^N \leftarrow$ initial sample of particles.

for $t \leftarrow 1, T$ **do**

  for all $i \leftarrow 1, N$ **do**

    $\tilde{s}_t^i \leftarrow q\left(s_t | s_{t-1}^i, y_t\right)$

    $\hat{w}_t^i \leftarrow w_{t-1}^i \frac{p\left(y_t | \tilde{s}_t^i\right) p\left(\tilde{s}_t^i | s_{t-1}^i\right)}{q\left(\tilde{s}_t^i | s_{t-1}^i, y_t\right)}$

  end for

  $\tilde{w}_t^i \leftarrow \frac{\hat{w}_t^i}{\sum_{i=1}^N \hat{w}_t^i}$

  $\left\{s_t^i, w_t^i\right\}_{i=1}^N \leftarrow \left\{\tilde{s}_t^i, \tilde{w}_t^i\right\}_{i=1}^N$

end for

Issue 1

Degeneracy

- Weights degenerate as t goes to infinity...
- Up to the point where all but one particles have zero weights.
- We don't want to sum-up a distribution with a single point.

$\Rightarrow$ Resampling (kill particles with low weights and replicate good particles).

Approximated nonlinear filter (IV)

Particle filter algorithm with resampling

$\left\{s_0^i, w_0^i\right\}_{i=1}^N \leftarrow$ initial sample of particles.
for $t \leftarrow 1, T$ **do**
    for all $i \leftarrow 1, N$ **do**
        $\tilde{s}_t^i \leftarrow q\left(s_t | s_{t-1}^i, y_t\right)$
        $\hat{w}_t^i \leftarrow w_{t-1}^i \frac{p\left(y_t | \tilde{s}_t^i\right) p\left(\tilde{s}_t^i | s_{t-1}^i\right)}{q\left(\tilde{s}_t^i | s_{t-1}^i, y_t\right)}$
    end for
    $\tilde{w}_t^i \leftarrow \frac{\hat{w}_t^i}{\sum_{i=1}^N \hat{w}_t^i}$
    $N_t^\star \leftarrow \left(\sum_{i=1}^N \left(\tilde{w}_t^i\right)^2\right)^{-1}$
    if $N_t^\star < \alpha N$ **then**
        $\left\{s_t^i, w_t^i\right\}_{i=1}^N \leftarrow$ Resampling step, with $w_t^i = {}^1/_N$
    else
        $\left\{s_t^i, w_t^i\right\}_{i=1}^N \leftarrow \left\{\tilde{s}_t^i, \tilde{w}_t^i\right\}_{i=1}^N$
    end if
end for

Set the proposal distribution

The blind proposal

$$q\left(s_t|s_{t-1}, y_t\right) = p\left(s_t|s_{t-1}\right)$$

- Does not use current observation.
- Weights recursive expression simplifies into:

$$\widehat{\omega}_t\left(s_t\right) \propto \widetilde{\omega}_{t-1}\left(s_{t-1}\right) p\left(y_t|s_t\right)$$

- Conditional density of observation is:

$$p\left(y_t|\tilde{s}_t^i\right) = (2\pi)^{-\frac{n}{2}}\ \ |R|^{-\frac{1}{2}} e^{-\frac{1}{2}\left(y_t - g(\tilde{s}_t^i;\boldsymbol{\theta})\right)' R^{-1}\left(y_t - g(\tilde{s}_t^i;\boldsymbol{\theta})\right)}$$

$\Rightarrow$ Measurement errors are mandatory!

Issues 2 and 3

Likelihood

$$p\left(y_{1:T}|\boldsymbol{\theta}\right) = p\left(y_1|s_0;\boldsymbol{\theta}\right)p\left(s_0|\boldsymbol{\theta}\right)\prod_{t=2}^{T}p\left(y_t|y_{1:t-1};\boldsymbol{\theta}\right)$$
$$\approx \sum_{i=1}^{N}\widetilde{\omega}_{t-1}^{i}p\left(y_t|s_t^i;\boldsymbol{\theta}\right)$$

1. How to choose the initial distribution for the states (s_0)?
2. Non differentiability of the likelihhod because of the resampling step.

Estimation with particle filter

- Do not use gradient based methods to estimate the mode...
- Simplex "works", but still issues with the hessian at the estimated mode.
- Better to run directly the MCMC.
- Slice? SMC...
- Do we really need particle (or nonlinear) filters?