ML-4
7/22/11

Filtering and Likelihood Inference

Jesús Fernández-Villaverde

University of Pennsylvania

July 10, 2011

Motivation

- Filtering, smoothing, and forecasting problems are pervasive in economics.

- Examples:

    1. Macroeconomics: evaluating likelihood of DSGE models.

    2. Microeconomics: structural models of individual choice with unobserved heterogeneity.

    3. Finance: time-varying variance of asset returns.

- However, filtering is a complicated endeavor with no simple and exact algorithm.

Environment I

- Discrete time $t \in \{1, 2, ...\}$.
- Why discrete time?
  1. Economic data is discrete.
  2. Easier math.
- Comparison with continuous time:
  1. Discretize observables.
  2. More involved math (stochastic calculus) but often we have extremely powerful results.

Environment II

- States S_t.
- We will focus on continuous state spaces.
- Comparison with discrete states:
  1. Markov-Switching models.
  2. Jumps and continuous changes.
- Initial state S_0 is either known or it comes from $p(S_0; \gamma)$.
- Properties of $p(S_0; \gamma)$? Stationarity?

State Space Representations

- Transition equation:

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

- Measurement equation:

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

- f and g are measurable functions.
- Interpretation. Modelling origin.
- Note Markov structure.

Shocks

- $\{W_t\}$ and $\{V_t\}$ are independent of each other.
- $\{W_t\}$ is known as process noise and $\{V_t\}$ as measurement noise.
- W_t and V_t have zero mean.
- No assumptions on the distribution beyond that.
- Often, we assume that the variance of W_t is given by R_t and the variance of V_t by Q_t.

DSGE Models and State Space Representations

- We have the solution of a DSGE model:

$$\begin{aligned} S_t &= P_1 S_{t-1} + P_2 Z_t \\ Y_t &= R_1 S_{t-1} + R_2 Z_t \end{aligned}$$

- This has nearly the same form that

$$\begin{aligned} S_t &= f\left(S_{t-1}, W_t; \gamma\right) \\ Y_t &= g\left(S_t, V_t; \gamma\right) \end{aligned}$$

- We only need to be careful with:
  1. To rewrite the measurement equation in terms of S_t instead of S_{t-1}.
  2. How we partition Z_t into W_t and V_t.
- Later, we will present an example.

Generalizations I

We can accommodate many generalizations by playing with the state definition:

1. Serial correlation of shocks.
2. Contemporaneous correlation of shocks.
3. Time changing state space equations.

Often, even infinite histories (for example in a dynamic game) can be tracked by a Lagrangian multiplier.

Generalizations II

- However, some generalizations can be tricky to accommodate.

- Take the model:

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

$$Y_t = g\left(S_t, V_t, Y_{t-1}; \gamma\right)$$

Y_t will be an infinite-memory process.

Conditional Densities

- From $S_t = f\left(S_{t-1}, W_t; \gamma\right)$, we can compute $p\left(S_t | S_{t-1}; \gamma\right)$.

- From $Y_t = g\left(S_t, V_t; \gamma\right)$, we can compute $p\left(Y_t | S_t; \gamma\right)$.

- From $S_t = f\left(S_{t-1}, W_t; \gamma\right)$ and $Y_t = g\left(S_t, V_t; \gamma\right)$, we have:

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

and hence we can compute $p\left(Y_t | S_{t-1}; \gamma\right)$.

Filtering, Smoothing, and Forecasting

- Filtering: we are concerned with what we have learned up to current observation.

- Smoothing: we are concerned with what we learn with the full sample.

- Forecasting: we are concerned with future realizations.

Goal of Filtering I

- Compute conditional densities: $p\left(S_t|y^{t-1};\gamma\right)$ and $p\left(S_t|y^t;\gamma\right)$.

- Why?

  1. It allows probability statements regarding the situation of the system.

  2. Compute conditional moments: mean, $s_{t|t}$ and $s_{t|t-1}$, and variances $P_{t|t}$ and $P_{t|t-1}$.

  3. Other functions of the states. Examples of interest.

- Theoretical point: do the conditional densities exist?

Goals of Filtering II

- Evaluate the likelihood function of the observables y^T at parameter values γ:

$$p\left(y^T;\gamma\right)$$

- Given the Markov structure of our state space representation:

$$p\left(y^T;\gamma\right)=\prod_{t=1}^{T}p\left(y_t|y^{t-1};\gamma\right)$$

- Then:

$$\begin{aligned}p\left(y^T;\gamma\right) &= p\left(y_1|\gamma\right)\prod_{t=2}^{T}p\left(y_t|y^{t-1};\gamma\right)\\ &= \int p\left(y_1|s_1;\gamma\right)dS_1\prod_{t=2}^{T}\int p\left(y_t|S_t;\gamma\right)p\left(S_t|y^{t-1};\gamma\right)dS_t\end{aligned}$$

- Hence, knowledge of $\left\{p\left(S_t|y^{t-1};\gamma\right)\right\}_{t=1}^{T}$ and $p\left(S_1;\gamma\right)$ allow the evaluation of the likelihood of the model.

Two Fundamental Tools

1. Chapman-Kolmogorov equation:

$$p\left(S_t|y^{t-1};\gamma\right)=\int p\left(S_t|S_{t-1};\gamma\right)p\left(S_{t-1}|y^{t-1};\gamma\right)dS_{t-1}$$

2. Bayes' theorem:

$$p\left(S_t|y^{t};\gamma\right)=\frac{p\left(y_t|S_t;\gamma\right)p\left(S_t|y^{t-1};\gamma\right)}{p\left(y_t|y^{t-1};\gamma\right)}$$

where:

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

Interpretation

- All filtering problems have two steps: prediction and update.

  1. Chapman-Kolmogorov equation is one-step ahead predictor.
  2. Bayes' theorem updates the conditional density of states given the new observation.

- We can think of those two equations as operators that map measures into measures.

Recursion for Conditional Distribution

- Combining the Chapman-Kolmogorov and the Bayes' theorem:

$$p\left(S_t|y^t;\gamma\right) = \frac{\int p\left(S_t|S_{t-1};\gamma\right)p\left(S_{t-1}|y^{t-1};\gamma\right)dS_{t-1}}{\int\left\{\int p\left(S_t|S_{t-1};\gamma\right)p\left(S_{t-1}|y^{t-1};\gamma\right)dS_{t-1}\right\}p\left(y_t|S_t;\gamma\right)dS_t}p\left(y_t|S_t;\gamma\right)$$

- To initiate that recursion, we only need a value for s_0 or $p\left(S_0;\gamma\right)$.

- Applying the Chapman-Kolmogorov equation once more, we get $\left\{p\left(S_t|y^{t-1};\gamma\right)\right\}_{t=1}^{T}$ to evaluate the likelihood function.

Initial Conditions I

- From previous discussion, we know that we need a value for s_1 or $p(S_1;\gamma)$.
- Stationary models: ergodic distribution.
- Non-stationary models: more complicated. Importance of transformations.
- Initialization in the case of Kalman filter.
- Forgetting conditions.
- Non-contraction properties of the Bayes operator.

Smoothing

- We are interested on the distribution of the state conditional on all the observations, on $p\left(S_t|y^T;\gamma\right)$ and $p\left(y_t|y^T;\gamma\right)$.

- We compute:

$$p\left(S_t|y^T;\gamma\right) = p\left(S_t|y^t;\gamma\right)\int \frac{p\left(S_{t+1}|y^T;\gamma\right)p\left(S_{t+1}|S_t;\gamma\right)}{p\left(S_{t+1}|y^t;\gamma\right)}dS_{t+1}$$

a backward recursion that we initialize with $p\left(S_T|y^T;\gamma\right)$, $\{p\left(S_t|y^t;\gamma\right)\}_{t=1}^{T}$ and $\left\{p\left(S_t|y^{t-1};\gamma\right)\right\}_{t=1}^{T}$ we obtained from filtering.

Forecasting

- We apply the Chapman-Kolmogorov equation recursively, we can get $p\left(S_{t+j}|y^{t};\gamma\right),\ j\geq 1$.

- Integrating recursively:

$$p\left(y_{l+1}|y^{l};\gamma\right)=\int p\left(y_{l+1}|S_{l+1};\gamma\right)p\left(S_{l+1}|y^{l};\gamma\right)dS_{l+1}$$

from $t+1$ to $t+j$, we get $p\left(y_{t+j}|y^{T};\gamma\right)$.

- Clearly smoothing and forecasting require to solve the filtering problem first!

Problem of Filtering

- We have the recursion

$$p\left(S_t|y^t;\gamma\right) = \frac{\int p\left(S_t|S_{t-1};\gamma\right) p\left(S_{t-1}|y^{t-1};\gamma\right) dS_{t-1}}{\int \left\{\int p\left(S_t|S_{t-1};\gamma\right) p\left(S_{t-1}|y^{t-1};\gamma\right) dS_{t-1}\right\} p\left(y_t|S_t;\gamma\right) dS_t} p\left(y_t|S_t;\gamma\right)$$

- A lot of complicated and high dimensional integrals (plus the one involved in the likelihood).
- In general, we do not have closed form solution for them.
- *Translate*, *spread*, and *deform* (TSD) the conditional densities in ways that impossibilities to fit them within any known parametric family.

Exception

- There is one exception: linear and Gaussian case.
- Why? Because if the system is linear and Gaussian, all the conditional probabilities are also Gaussian.
- Linear and Gaussian state spaces models *translate* and *spread* the conditional distributions, but they do not *deform* them.
- For Gaussian distributions, we only need to track mean and variance (sufficient statistics).
- Kalman filter accomplishes this goal efficiently.

Linear Gaussian Case

- Let the following system:

  - Transition equation

$$s_t = Fs_{t-1} + G\omega_t, \ \omega_t \sim \mathcal{N}(0, Q)$$

  - Measurement equation

$$y_t = Hs_t + \upsilon_t, \ \upsilon_t \sim \mathcal{N}(0, R)$$

- Assume we want to write the likelihood function of $y^T = \{y_t\}_{t=1}^T$.

The State Space Representation is Not Unique

- Take the previous state space representation.
- Let B be a non-singular squared matrix conforming with F.
- Then, if $s_t^* = Bs_t$, $F^* = BFB^{-1}$, $G^* = BG$, and $H^* = HB^{-1}$, we can write a new, equivalent, representation:
  - Transition equation

$$s_{t+1}^* = F^* s_t^* + G^* \omega_t, \; \omega_t \sim \mathcal{N}(0, Q)$$

  - Measurement equation

$$y_t = H^* s_t^* + \upsilon_t, \; \upsilon_t \sim \mathcal{N}(0, R)$$

Example I

- AR(2) process:

$$y_t = \rho_1 y_{t-1} + \rho_2 z_{t-2} + \sigma_\upsilon \upsilon_t, \;\; \upsilon_t \sim \mathcal{N}(0, 1)$$

- Model is not apparently not Markovian.
- However, it is trivial to write it in a state space form.
- In fact, we have many different state space forms.

Example I

- State Space Representation I:

$$\begin{pmatrix} y_t \\ \rho_2 y_{t-1} \end{pmatrix} = \begin{pmatrix} \rho_1 & 1 \\ \rho_2 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ \rho_2 y_{t-2} \end{pmatrix} + \begin{pmatrix} \sigma_\upsilon \\ 0 \end{pmatrix} \upsilon_t$$

$$y_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} y_t \\ \rho_2 y_{t-1} \end{pmatrix}$$

- State Space Representation II:

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} = \begin{pmatrix} \rho_1 & \rho_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} \sigma_\upsilon \\ 0 \end{pmatrix} \upsilon_t$$

$$y_t = \begin{pmatrix} 1 & \rho_2 \end{pmatrix} \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}$$

- Rotation $B = \begin{pmatrix} 1 & 0 \\ 0 & \rho_2 \end{pmatrix}$ on the second system to get the first one.

Example II

- MA(1) process:

$$y_t = \upsilon_t + \theta \upsilon_{t-1},\ \ \upsilon_t \sim \mathcal{N}\left(0, \sigma_\upsilon^2\right), \text{ and } \mathbb{E}\upsilon_t \upsilon_s = 0 \text{ for } s \neq t.$$

- State Space Representation I:

$$\begin{aligned} \begin{pmatrix} y_t \\ \theta \upsilon_t \end{pmatrix} &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ \theta \upsilon_{t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ \theta \end{pmatrix} \upsilon_t \\ y_t &= \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} y_t \\ \theta \upsilon_t \end{pmatrix} \end{aligned}$$

- State Space Representation II:

$$\begin{aligned} s_t &= \upsilon_{t-1} \\ y_t &= s x_t + \upsilon_t \end{aligned}$$

- Again both representations are equivalent!

Example III

- Now we explore a different issue.
- Random walk plus drift process:

$$y_t = y_{t-1} + \beta + \sigma_v v_t, \;\; v_t \sim \mathcal{N}(0,1)$$

- This is even more interesting: we have a unit root and a constant parameter (the drift).
- State Space Representation:

$$\begin{aligned} \begin{pmatrix} y_t \\ \beta \end{pmatrix} &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y_{t1} \\ \beta \end{pmatrix} + \begin{pmatrix} \sigma_v \\ 0 \end{pmatrix} v_t \\ y_t &= \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} y_t \\ \beta \end{pmatrix} \end{aligned}$$

Some Conditions on the State Space Representation

- We only consider stable systems.

- A system is stable if for any initial state s_0, the vector of states, s_t, converges to some unique s^*.

- A necessary and sufficient condition for the system to be stable is that:

$$|\lambda_i(F)| < 1$$

  for all i, where $\lambda_i(F)$ stands for eigenvalue of F.

Introducing the Kalman Filter

- Developed by Kalman and Bucy.
- Wide application in science.
- Basic idea.
- Prediction, smoothing, and control.
- Different derivations.

Some Definitions

Definition

Let $s_{t|t-1} = \mathbb{E}\left(s_t|y^{t-1}\right)$ be the best linear predictor of s_t given the history of observables until $t-1$, i.e., y^{t-1}.

Definition

Let $y_{t|t-1} = \mathbb{E}\left(y_t|y^{t-1}\right) = Hs_{t|t-1}$ be the best linear predictor of y_t given the history of observables until $t-1$, i.e., y^{t-1}.

Definition

Let $s_{t|t} = \mathbb{E}\left(s_t|y^{t}\right)$ be the best linear predictor of s_t given the history of observables until t, i.e., s^{t}.

What is the Kalman Filter Trying to Do?

- Let assume we have $s_{t|t-1}$ and $y_{t|t-1}$.
- We observe a new y_t.
- We need to obtain $s_{t|t}$.
- Note that $s_{t+1|t} = Fs_{t|t}$ and $y_{t+1|t} = Hs_{t+1|t}$, so we can go back to the first step and wait for y_{t+1}.
- Therefore, the key question is how to obtain $s_{t|t}$ from $s_{t|t-1}$ and y_t.

A Minimization Approach to the Kalman Filter

- Assume we use the following equation to get $s_{t|t}$ from y_t and $s_{t|t-1}$:

$$s_{t|t} = s_{t|t-1} + K_t \left(y_t - y_{t|t-1} \right) = s_{t|t-1} + K_t \left(y_t - Hs_{t|t-1} \right)$$

- This formula will have some probabilistic justification (to follow).
- K_t is called the Kalman filter gain and it measures how much we update $s_{t|t-1}$ as a function in our error in predicting y_t.
- The question is how to find the optimal K_t.
- The Kalman filter is about how to build K_t such that optimally update $s_{t|t}$ from $s_{t|t-1}$ and y_t.
- How do we find the optimal K_t?

Some Additional Definitions

Definition

Let $\Sigma_{t|t-1} \equiv \mathbb{E}\left(\left(s_t - s_{t|t-1} \right) \left(s_t - s_{t|t-1} \right)' |y^{t-1} \right)$ be the predicting error variance covariance matrix of s_t given the history of observables until $t-1$, i.e. y^{t-1}.

Definition

Let $\Omega_{t|t-1} \equiv \mathbb{E}\left(\left(y_t - y_{t|t-1} \right) \left(y_t - y_{t|t-1} \right)' |y^{t-1} \right)$ be the predicting error variance covariance matrix of y_t given the history of observables until $t-1$, i.e. y^{t-1}.

Definition

Let $\Sigma_{t|t} \equiv \mathbb{E}\left(\left(s_t - s_{t|t} \right) \left(s_t - s_{t|t} \right)' |y^{t} \right)$ be the predicting error variance covariance matrix of s_t given the history of observables until t, i.e. y^t.

The Kalman Filter Algorithm I

- Given $\Sigma_{t|t-1}$, y_t, and $s_{t|t-1}$, we can now set the Kalman filter algorithm.

- Let $\Sigma_{t|t-1}$, then we compute:

$$
\begin{aligned}
\Omega_{t|t-1} &\equiv \mathbb{E}\left(\left(y_t - y_{t|t-1}\right)\left(y_t - y_{t|t-1}\right)' | y^{t-1}\right) \\
&= \mathbb{E}\begin{pmatrix} H\left(s_t - s_{t|t-1}\right)\left(s_t - s_{t|t-1}\right)' H' \\ +\upsilon_t\left(s_t - s_{t|t-1}\right)' H' \\ +H\left(s_t - s_{t|t-1}\right)\upsilon_t' + \upsilon_t \upsilon_t' | y^{t-1} \end{pmatrix} \\
&= H\Sigma_{t|t-1}H' + R
\end{aligned}
$$

The Kalman Filter Algorithm II

- Let $\Sigma_{t|t-1}$, then we compute:

$$\begin{array}{rcl} \mathbb{E}\left(\left(y_t - y_{t|t-1}\right)\left(s_t - s_{t|t-1}\right)' | y^{t-1}\right) & = & \\ \mathbb{E}\left(\begin{array}{c} H\left(s_t - s_{t|t-1}\right)\left(s_t - s_{t|t-1}\right)' \\ +\upsilon_t\left(s_t - s_{t|t-1}\right)' | y^{t-1} \end{array}\right) & = & H\Sigma_{t|t-1} \end{array}$$

- Let $\Sigma_{t|t-1}$, then we compute:

$$K_t = \Sigma_{t|t-1} H' \left(H\Sigma_{t|t-1} H' + R\right)^{-1}$$

- Let $\Sigma_{t|t-1}$, $s_{t|t-1}$, K_t, and y_t, then we compute:

$$s_{t|t} = s_{t|t-1} + K_t\left(y_t - Hs_{t|t-1}\right)$$

Finding the Optimal Gain

- We want K_t such that $\min \Sigma_{t|t}$.
- Thus:

$$K_t = \Sigma_{t|t-1} H' \left(H \Sigma_{t|t-1} H' + R \right)^{-1}$$

with the optimal update of $s_{t|t}$ given y_t and $s_{t|t-1}$ being:

$$s_{t|t} = s_{t|t-1} + K_t \left(y_t - H s_{t|t-1} \right)$$

- Intuition: note that we can rewrite K_t in the following way:

$$K_t = \Sigma_{t|t-1} H' \Omega_{t|t-1}^{-1}$$

1. If we did a big mistake forecasting $s_{t|t-1}$ using past information ($\Sigma_{t|t-1}$ large), we give a lot of weight to the new information (K_t large).
2. If the new information is noise (R large), we give a lot of weight to the old prediction (K_t small).

Example

- Assume the following model in state space form:
  - Transition equation:
$$s_t = \mu + \omega_t,\ \omega_t \sim N\left(0, \sigma_\omega^2\right)$$
  - Measurement equation:
$$y_t = s_t + \upsilon_t,\ \upsilon_t \sim N\left(0, \sigma_\upsilon^2\right)$$

- Let $\sigma_\upsilon^2 = q\sigma_\omega^2$.
- Then, if $\Sigma_{1|0} = \sigma_\omega^2$, ($s_1$ is drawn from the ergodic distribution of s_t):
$$K_1 = \sigma_\omega^2 \frac{1}{1+q} \propto \frac{1}{1+q}.$$

- Therefore, the bigger σ_υ^2 relative to σ_ω^2 (the bigger q), the lower K_1 and the less we trust y_1.

The Kalman Filter Algorithm III

- Let $\Sigma_{t|t-1}$, $s_{t|t-1}$, K_t, and y_t.
- Then, we compute:

$$\Sigma_{t|t} \equiv \mathbb{E}\left(\left(s_t - s_{t|t}\right)\left(s_t - s_{t|t}\right)'|y^t\right) =$$

$$\mathbb{E}\left(\begin{array}{c} \left(s_t - s_{t|t-1}\right)\left(s_t - s_{t|t-1}\right)' - \\ \left(s_t - s_{t|t-1}\right)\left(y_t - Hs_{t|t-1}\right)' K_t' - \\ K_t\left(y_t - Hs_{t|t-1}\right)\left(s_t - s_{t|t-1}\right)' + \\ K_t\left(y_t - Hs_{t|t-1}\right)\left(y_t - Hs_{t|t-1}\right)' K_t'|y^t \end{array}\right) = \Sigma_{t|t-1} - K_t H \Sigma_{t|t-1}$$

where

$$s_t - s_{t|t} = s_t - s_{t|t-1} - K_t\left(y_t - Hs_{t|t-1}\right).$$

The Kalman Filter Algorithm IV

- Let $\Sigma_{t|t-1}$, $s_{t|t-1}$, K_t, and y_t, then we compute:

$$\Sigma_{t+1|t} = F\Sigma_{t|t}F' + GQG'$$

- Let $s_{t|t}$, then we compute:

  1. $s_{t+1|t} = Fs_{t|t}$
  2. $y_{t+1|t} = Hs_{t+1|t}$

- Therefore, from $s_{t|t-1}$, $\Sigma_{t|t-1}$, and y_t we compute $s_{t|t}$ and $\Sigma_{t|t}$.

- We also compute $y_{t|t-1}$ and $\Omega_{t|t-1}$ to help (later) to calculate the likelihood function of $y^T = \{y_t\}_{t=1}^T$.

The Kalman Filter Algorithm: A Review

We start with $s_{t|t-1}$ and $\Sigma_{t|t-1}$. Then, we observe y_t and:

- $\Omega_{t|t-1} = H\Sigma_{t|t-1}H' + R$
- $y_{t|t-1} = Hs_{t|t-1}$
- $K_t = \Sigma_{t|t-1}H'\left(H\Sigma_{t|t-1}H' + R\right)^{-1}$
- $\Sigma_{t|t} = \Sigma_{t|t-1} - K_t H\Sigma_{t|t-1}$
- $s_{t|t} = s_{t|t-1} + K_t\left(y_t - Hs_{t|t-1}\right)$
- $\Sigma_{t+1|t} = F\Sigma_{t|t}F' + GQG'$
- $s_{t+1|t} = Fs_{t|t}$

We finish with $s_{t+1|t}$ and $\Sigma_{t+1|t}$.

Writing the Likelihood Function

Likelihood function of $y^T = \{y_t\}_{t=1}^T$:

$$\log p\left(y^T | F, G, H, Q, R\right) =$$

$$\sum_{t=1}^{T} \log p\left(y_t | y^{t-1} F, G, H, Q, R\right) =$$

$$-\sum_{t=1}^{T} \left[\frac{N}{2}\log 2\pi + \frac{1}{2}\log\left|\Omega_{t|t-1}\right| + \frac{1}{2}\varsigma_t' \Omega_{t|t-1}^{-1} \varsigma_t\right]$$

where:

$$\varsigma_t = y_t - y_{t|t-1} = y_t - Hs_{t|t-1}$$

is white noise and:

$$\Omega_{t|t-1} = H_t \Sigma_{t|t-1} H_t' + R$$

Initial conditions for the Kalman Filter

- An important step in the Kalman Filter is to set the initial conditions.
- Initial conditions $s_{1|0}$ and $\Sigma_{1|0}$.
- Where do they come from?

Since we only consider stable system, the standard approach is to set:

- $s_{1|0} = s^*$
- $\Sigma_{1|0} = \Sigma^*$

where s solves:

$$
\begin{aligned}
s^* &= F s^* \\
\Sigma^* &= F \Sigma^* F' + GQG'
\end{aligned}
$$

- How do we find Σ^*?

$$\Sigma^* = \left[I - F \otimes F \right]^{-1} vec(GQG')$$

Initial conditions for the Kalman Filter II

Under the following conditions:

1. The system is stable, i.e. all eigenvalues of F are strictly less than one in absolute value.
2. GQG' and R are p.s.d. symmetric.
3. $\Sigma_{1|0}$ is p.s.d. symmetric.

Then $\Sigma_{t+1|t} \rightarrow \Sigma^*$.

Remarks

1. There are more general theorems than the one just described.
2. Those theorems are based on non-stable systems.
3. Since we are going to work with stable system the former theorem is enough.
4. Last theorem gives us a way to find Σ as $\Sigma_{t+1|t} \to \Sigma$ for any $\Sigma_{1|0}$ we start with.

The Kalman Filter and DSGE models

- Basic real business cycle model:

$$\max \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left\{ \log c_t + \psi \log \left(1 - l_t\right) \right\}$$

$$c_t + k_{t+1} = k_t^{\alpha} \left(e^{z_t} l_t\right)^{1-\alpha} + \left(1 - \delta\right) k_t$$

$$z_t = \rho z_{t-1} + \sigma \varepsilon_t, \ \varepsilon_t \sim \mathcal{N}(0, 1)$$

- Equilibrium conditions:

$$\frac{1}{c_t} = \beta \mathbb{E}_t \left\{ \frac{1}{c_{t+1}} \left(\alpha k_{t+1}^{\alpha - 1} \left(e^{z_{t+1}} l_{t+1}\right)^{1-\alpha} + 1 - \delta \right) \right\}$$

$$\psi \frac{l_t}{1 - l_t} c_t = \left(1 - \alpha\right) k_t^{\alpha} \left(e^{z_t} l_t\right)^{1-\alpha}$$

$$c_t + k_{t+1} = k_t^{\alpha} \left(e^{z_t} l_t\right)^{1-\alpha} + \left(1 - \delta\right) k_t$$

$$z_t = \rho z_{t-1} + \sigma \varepsilon_t$$

The Kalman Filter and Linearized DSGE Models

- We loglinearize (or linearize) the equilibrium conditions around the steady state.
- Alternatives: particle filter.
- We assume that we have data on:
  1. $\log output_t$
  2. $\log l_t$
  3. $\log c_t$

  s.t. a linearly additive measurement error $V_t = \begin{pmatrix} v_{1,t} & v_{2,t} & v_{3,t} \end{pmatrix}'$.
- Why measurement error? Stochastic singularity.
- Degrees of freedom in the measurement equation.

Policy Functions

- We need to write the model in state space form.
- Remember that a loglinear solution has the form:

$$\widehat{k}_{t+1} = p_1 \widehat{k}_t + p_2 z_t$$

and

$$\begin{aligned}
\widehat{output}_t &= q_1 \widehat{k}_t + q_2 z_t \\
\widehat{l}_t &= r_1 \widehat{k}_t + r_2 z_t \\
\widehat{c}_t &= u_1 \widehat{k}_t + u_2 z_t
\end{aligned}$$

Writing the Likelihood Function

- Transition equation:

$$\underbrace{\begin{pmatrix} 1 \\ \widehat{k}_t \\ z_t \end{pmatrix}}_{s_t} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & p_1 & p_2 \\ 0 & 0 & \rho \end{pmatrix}}_{F} \underbrace{\begin{pmatrix} 1 \\ \widehat{k}_{t-1} \\ z_{t-1} \end{pmatrix}}_{s_{t-1}} + \underbrace{\begin{pmatrix} 0 \\ 0 \\ \sigma \end{pmatrix}}_{G} \underbrace{\epsilon_t}_{\omega_t}.$$

- Measurement equation:

$$\underbrace{\begin{pmatrix} \log output_t \\ \log l_t \\ \log c_t \end{pmatrix}}_{y_t} = \underbrace{\begin{pmatrix} \log y & q_1 & q_2 \\ \log l & r_1 & r_2 \\ \log c & u_1 & u_2 \end{pmatrix}}_{H} \underbrace{\begin{pmatrix} 1 \\ \widehat{k}_t \\ z_t \end{pmatrix}}_{s_t} + \underbrace{\begin{pmatrix} v_{1,t} \\ v_{2,t} \\ v_{3,t} \end{pmatrix}}_{\upsilon}$$

The Solution to the Model in State Space Form

- Now, with y^T, F, G, H, Q, and R as defined before...

- ...we can use the Ricatti equations to evaluate the likelihood function:

$$\log p\left(y^T|\gamma\right) = \log p\left(y^T|F, G, H, Q, R\right)$$

where $\gamma = \{\alpha, \beta, \rho, \psi, \delta, \sigma\}$.

- Cross-equations restrictions implied by equilibrium solution.

- With the likelihood, we can do inference!

Nonlinear Filtering

- Different approaches.
- Deterministic filtering:
  1. Kalman family.
  2. Grid-based filtering.
- Simulation filtering:
  1. McMc.
  2. Sequential Monte Carlo.

Kalman Family of Filters

- Use ideas of Kalman filtering to NLGF problems.
- Non-optimal filters.
- Different implementations:
  1. Extended Kalman filter.
  2. Iterated Extended Kalman filter.
  3. Second-order Extended Kalman filter.
  4. Unscented Kalman filter.

The Extended Kalman Filter

- EKF is historically the first descendant of the Kalman filter.
- EKF deals with nonlinearities with a first order approximation to the system and applying the Kalman filter to this approximation.
- Non-Gaussianities are ignored.

Algorithm

- Given $s_{t-1|t-1}$, $s_{t|t-1} = f\left(s_{t-1|t-1}, 0; \gamma\right)$.
- Then:

$$P_{t|t-1} = Q_{t-1} + F_t P_{t-1|t-1} F_t'$$

where

$$F_t = \left.\frac{df\left(S_{t-1}, W_t; \gamma\right)}{dS_{t-1}}\right|_{S_{t-1}=s_{t-1|t-1}, W_t=0}$$

- Kalman gain, K_t, is:

$$K_t = P_{t|t-1} G_t' \left(G_t P_{t|t-1} G_t' + R_t\right)^{-1}$$

where

$$G_t = \left.\frac{dg\left(S_{t-1}, v_t; \gamma\right)}{dS_{t-1}}\right|_{S_{t-1}=s_{t|t-1}, v_t=0}$$

- Then

$$\begin{aligned} s_{t|t} &= s_{t|t-1} + K_t \left(y_t - g\left(s_{t|t-1}, 0; \gamma\right)\right) \\ P_{t|t} &= P_{t|t-1} - K_t G_t P_{t|t-1} \end{aligned}$$

Problems of EKF

1. It ignores the non-Gaussianities of W_t and V_t.
2. It ignores the non-Gaussianities of states distribution.
3. Approximation error incurred by the linearization.
4. Biased estimate of the mean and variance.
5. We need to compute Jacobian and Hessians.

As the sample size grows, those errors accumulate and the filter diverges.

Iterated Extended Kalman Filter I

- Compute $s_{t|t-1}$ and $P_{t|t-1}$ as in EKF.
- Iterate N times on:

$$K_t^i = P_{t|t-1} G_t^{i\prime} \left(G_t^i P_{t|t-1} G_t^{i\prime} + R_t \right)^{-1}$$

where

$$G_t^i = \left. \frac{dg\left(S_{t-1}, v_t; \gamma\right)}{dS_{t-1}} \right|_{S_{t-1}=s_{t|t-1}^i, v_t=0}$$

and

$$s_{t|t}^i = s_{t|t-1} + K_t^i \left(y_t - g\left(s_{t|t-1}, 0; \gamma\right)\right)$$

- Why are we iterating? How many times?
- Then:

$$\begin{aligned} s_{t|t} &= s_{t|t-1} + K_t \left(y_t - g\left(s_{t|t-1}^N, 0; \gamma \right)\right) \\ P_{t|t} &= P_{t|t-1} - K_t^N G_t^N P_{t|t-1} \end{aligned}$$

Second-order Extended Kalman Filter

- We keep second-order terms of the Taylor expansion of transition and measurement.
- Theoretically, less biased than EKF.
- Messy algebra.
- In practice, not much improvement.

Unscented Kalman Filter I

- Recent proposal by Julier and Uhlmann (1996).
- Based around the unscented transform.
- A set of sigma points is selected to preserve some properties of the conditional distribution (for example, the first two moments).
- Then, those points are transformed and the properties of the new conditional distribution are computed.
- The UKF computes the conditional mean and variance accurately up to a third order approximation if the shocks W_t and V_t are Gaussian and up to a second order if they are not.
- The sigma points are chosen deterministically and not by simulation as in a Monte Carlo method.
- The UKF has the advantage with respect to the EKF that no Jacobian or Hessians is required, objects that may be difficult to compute.

New State Variable

- We modify the state space by creating a new augmented state variable:

$$\mathbb{S}_t = [S_t, W_t, V_t]$$

that includes the pure state space and the two random variables W_t and V_t.

- We initialize the filter with

$$\mathbf{s}_{0|0} = E(\mathbb{S}_t) = E(S_0, 0, 0)$$

$$\mathbb{P}_{0|0} = \begin{bmatrix} P_{0|0} & 0 & 0 \\ 0 & R_0 & 0 \\ 0 & 0 & Q_0 \end{bmatrix}$$

Sigma Points

- Let L be the dimension of the state variable $\mathbb{S}_t$.
- For $t = 1$, we calculate the $2L + 1$ sigma points:

$$
\begin{aligned}
\mathcal{S}_{0,t-1|t-1} &= \mathbf{s}_{t-1|t-1} \\
\mathcal{S}_{i,t-1|t-1} &= \mathbf{s}_{t-1|t-1} - \left((L + \lambda) \mathbb{P}_{t-1|t-1} \right)^{0.5} \text{ for } i = 1, \ldots, L \\
\mathcal{S}_{i,t-1|t-1} &= \mathbf{s}_{t-1|t-1} + \left((L + \lambda) \mathbb{P}_{t-1|t-1} \right)^{0.5} \text{ for } i = L + 1, \ldots, 2L
\end{aligned}
$$

Parameters

- $\lambda = \alpha^2 (L + \kappa) - L$ is a scaling parameter.
- α determines the spread of the sigma point and it must belong to the unit interval.
- κ is a secondary parameter usually set equal to zero.
- Notation for each of the elements of $\mathcal{S}$:

$$\mathcal{S}_i = \left[\mathcal{S}_i^s, \mathcal{S}_i^w, \mathcal{S}_i^v\right] \text{ for } i = 0, ..., 2L$$

Weights

- Weights for each point:

$$\begin{aligned}
\mathcal{W}_0^m &= \frac{\lambda}{L+\lambda} \\
\mathcal{W}_0^c &= \frac{\lambda}{L+\lambda} + \left(1-\alpha^2+\beta\right) \\
\mathcal{W}_0^m &= X_0^c = \frac{1}{2\left(L+\lambda\right)} \text{ for } i = 1, ..., 2L
\end{aligned}$$

- β incorporates knowledge regarding the conditional distributions.
- For Gaussian distributions, $\beta = 2$ is optimal.

Algorithm I: Prediction of States

- We compute the transition of the pure states:

$$\mathcal{S}^{s}_{i,t|t-1} = f\left(\mathcal{S}^{s}_{i,t|t-1}, \mathcal{S}^{w}_{i,t-1|t-1}; \gamma\right)$$

- Weighted state

$$s_{t|t-1} = \sum_{i=0}^{2L} \mathcal{W}^{m}_{i} \mathcal{S}^{s}_{i,t|t-1}$$

- Weighted variance:

$$P_{t|t-1} = \sum_{i=0}^{2L} \mathcal{W}^{c}_{i} \left(\mathcal{S}^{s}_{i,t|t-1} - s_{t|t-1}\right)\left(\mathcal{S}^{s}_{i,t|t-1} - s_{t|t-1}\right)'$$

Algorithm II: Prediction of Observables

- Predicted sigma observables:

$$\mathcal{Y}_{i,t|t-1} = g\left(\mathcal{S}^{s}_{i,t|t-1}, \mathcal{S}^{v}_{i,t|t-1}; \gamma\right)$$

- Predicted observable:

$$y_{t|t-1} = \sum_{i=0}^{2L} \mathcal{W}_i^m \mathcal{Y}_{i,t|t-1}$$

Algorithm III: Update

- Variance-covariance matrices:

$$
\begin{aligned}
P_{yy,t} &= \sum_{i=0}^{2L} \mathcal{W}_i^c \left(\mathcal{Y}_{i,t|t-1} - y_{t|t-1}\right)\left(\mathcal{Y}_{i,t|t-1} - y_{t|t-1}\right)' \\
P_{xy,t} &= \sum_{i=0}^{2L} \mathcal{W}_i^c \left(\mathcal{S}_{i,t|t-1}^s - s_{t|t-1}\right)\left(\mathcal{Y}_{i,t|t-1} - y_{t|t-1}\right)'
\end{aligned}
$$

- Kalman gain:

$$
K_t = P_{xy,t} P_{yy,t}^{-1}
$$

Algorithm IV: Update

- Update of the state:

$$s_{t|t} = s_{t|t} + K_t \left(y_t - y_{t|t-1} \right)$$

- the update of the variance:

$$P_{t|t} = P_{t|t-1} + K_t P_{yy,t} K_t'$$

Finally:

$$\mathbb{P}_{t|t} = \begin{bmatrix} P_{t|t} & 0 & 0 \\ 0 & R_t & 0 \\ 0 & 0 & Q_t \end{bmatrix}$$

Grid-Based Filtering

- Remember that we have the recursion

$$p\left(s_t|y^t;\gamma\right) = \frac{\int p\left(s_t|s_{t-1};\gamma\right)p\left(s_{t-1}|y^{t-1};\gamma\right)ds_{t-1}}{\int\left\{\int p\left(s_t|s_{t-1};\gamma\right)p\left(s_{t-1}|y^{t-1};\gamma\right)ds_{t-1}\right\}p\left(y_t|s_t;\gamma\right)ds_t}p\left(y_t|s_t;\gamma\right)$$

- This recursion requires the evaluation of three integrals.
- This suggests the possibility of addressing the problem by computing those integrals by a deterministic procedure as a grid method.
- Kitagawa (1987)and Kramer and Sorenson (1988).

Grid-Based Filtering I

- We divide the state space into N cells, with center point s_t^i, $\left\{ s_t^i : i = 1, ..., N \right\}$.
- We substitute the exact conditional densities by discrete densities that put all the mass at the points $\left\{ s_t^i \right\}_{i=1}^{N}$.
- We denote $\delta \left(x \right)$ is a Dirac function with mass at 0.

Grid-Based Filtering II

- Then, approximated distributions and weights:

$$
\begin{aligned}
p\left(s_t|y^{t-1};\gamma\right) &\simeq \sum_{i=1}^{N}\omega_{t|t-1}^{i}\delta\left(s_t-s_t^i\right)\\
p\left(s_t|y^{t};\gamma\right) &\simeq \sum_{i=1}^{N}\omega_{t|t-1}^{i}\delta\left(s_t-s_t^i\right)\\
\omega_{t|t-1}^{i} &= \sum_{j=1}^{N}\omega_{t-1|t-1}^{j}p\left(s_t^i|s_{t-1}^j;\gamma\right)\\
\omega_{t|t}^{i} &= \frac{\omega_{t|t-1}^{i}p\left(y_t|s_t^i;\gamma\right)}{\sum_{j=1}^{N}\omega_{t|t-1}^{j}p\left(y_t|s_t^j;\gamma\right)}
\end{aligned}
$$

Approximated Recursion

$$p\left(s_t|y^t;\gamma\right) =$$

$$\sum_{i=1}^{N} \frac{\left[\sum_{j=1}^{N} \omega_{t-1|t-1}^{j} p\left(s_t^i|s_{t-1}^j;\gamma\right)\right] p\left(y_t|s_t^i;\gamma\right)}{\sum_{j=1}^{N}\left[\sum_{j=1}^{N} \omega_{t-1|t-1}^{j} p\left(s_t^i|s_{t-1}^j;\gamma\right)\right] p\left(y_t|s_t^j;\gamma\right)} \delta\left(s_t - s_t^i\right)$$

Compare with

$$p\left(s_t|y^t;\gamma\right) =$$

$$\frac{\int p\left(s_t|s_{t-1};\gamma\right) p\left(s_{t-1}|y^{t-1};\gamma\right) ds_{t-1}}{\int \left\{\int p\left(s_t|s_{t-1};\gamma\right) p\left(s_{t-1}|y^{t-1};\gamma\right) ds_{t-1}\right\} p\left(y_t|s_t;\gamma\right) ds_t} p\left(y_t|s_t;\gamma\right)$$

given that

$$p\left(s_{t-1}|y^{t-1};\gamma\right) \simeq \sum_{i=1}^{N} \omega_{t-1|t-1}^{i} \delta\left(s_t^i\right)$$

Problems

- Grid filters require a constant readjustment to small changes in the model or its parameter values.
- Too computationally expensive to be of any practical benefit beyond very low dimensions.
- Grid points are fixed ex ante and the results are very dependent on that choice.

Can we overcome those difficulties and preserve the idea of integration? Yes, through Monte Carlo Integration.

Particle Filtering

- Remember,
  1. Transition equation:
  $$S_t = f\left(S_{t-1}, W_t; \gamma\right)$$
  2. Measurement equation:
  $$Y_t = g\left(S_t, V_t; \gamma\right)$$
- Some Assumptions:
  1. We can partition $\{W_t\}$ into two independent sequences $\{W_{1,t}\}$ and $\{W_{2,t}\}$, s.t. $W_t = (W_{1,t}, W_{2,t})$ and $\dim(W_{2,t}) + \dim(V_t) \geq \dim(Y_t)$.
  2. We can always evaluate the conditional densities $p\left(y_t | W_1^t, y^{t-1}, S_0; \gamma\right)$.
  3. The model assigns positive probability to the data.

Rewriting the Likelihood Function

- Evaluate the likelihood function of the a sequence of realizations of the observable y^T at a particular parameter value γ:

$$p\left(y^T;\gamma\right)$$

- We factorize it as:

$$p\left(y^T;\gamma\right)=\prod_{t=1}^{T}p\left(y_t|y^{t-1};\gamma\right)$$

$$=\prod_{t=1}^{T}\int\int p\left(y_t|W_1^t,y^{t-1},S_0;\gamma\right)p\left(W_1^t,S_0|y^{t-1};\gamma\right)dW_1^t dS_0$$

A Law of Large Numbers

If $\left\{ \left\{ s_0^{t|t-1,i}, w_1^{t|t-1,i} \right\}_{i=1}^{N} \right\}_{t=1}^{T}$ N i.i.d. draws from $\left\{ p\left(W_1^t, S_0 | y^{t-1}; \gamma \right) \right\}_{t=1}^{T}$, then:

$$p\left(y^T; \gamma \right) \simeq \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} p\left(y_t | w_1^{t|t-1,i}, y^{t-1}, s_0^{t|t-1,i}; \gamma \right)$$

The problem of evaluating the likelihood is equivalent to the problem of drawing from

$$\left\{ p\left(W_1^t, S_0 | y^{t-1}; \gamma \right) \right\}_{t=1}^{T}$$

Introducing Particles

- $\left\{s_0^{t-1,i}, w_1^{t-1,i}\right\}_{i=1}^N$ N i.i.d. draws from $p\left(W_1^{t-1}, S_0 | y^{t-1}; \gamma\right)$.
- Each $s_0^{t-1,i}, w_1^{t-1,i}$ is a *particle* and $\left\{s_0^{t-1,i}, w_1^{t-1,i}\right\}_{i=1}^N$ a *swarm of particles*.
- $\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^N$ N i.i.d. draws from $p\left(W_1^{t}, S_0 | y^{t-1}; \gamma\right)$.
- Each $s_0^{t|t-1,i}, w_1^{t|t-1,i}$ is a *proposed particle* and $\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^N$ a *swarm of proposed particles*.
- Weights:

$$q_t^i = \frac{p\left(y_t | w_1^{t|t-1,i}, y^{t-1}, s_0^{t|t-1,i}; \gamma\right)}{\sum_{i=1}^N p\left(y_t | w_1^{t|t-1,i}, y^{t-1}, s_0^{t|t-1,i}; \gamma\right)}$$

A Proposition

Let $\left\{\widetilde{s}_0^i, \widetilde{w}_1^i\right\}_{i=1}^N$ be a draw with replacement from $\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^N$ and probabilities q_t^i. Then $\left\{\widetilde{s}_0^i, \widetilde{w}_1^i\right\}_{i=1}^N$ is a draw from $p\left(W_1^t, S_0|y^t; \gamma\right)$.

Importance of the Proposition:

1. It shows how a draw $\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^N$ from $p\left(W_1^t, S_0|y^{t-1}; \gamma\right)$ can be used to draw $\left\{s_0^{t,i}, w_1^{t,i}\right\}_{i=1}^N$ from $p\left(W_1^t, S_0|y^t; \gamma\right)$.
2. With a draw $\left\{s_0^{t,i}, w_1^{t,i}\right\}_{i=1}^N$ from $p\left(W_1^t, S_0|y^t; \gamma\right)$ we can use $p\left(W_{1,t+1}; \gamma\right)$ to get a draw $\left\{s_0^{t+1|t,i}, w_1^{t+1|t,i}\right\}_{i=1}^N$ and iterate the procedure.

Sequential Monte Carlo

Step 0, Initialization: Set $t \rightsquigarrow 1$ and set $p\left(W_1^{t-1}, S_0 | y^{t-1}; \gamma\right) = p\left(S_0; \gamma\right)$.

Step 1, Prediction: Sample N values $\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^{N}$ from the density $p\left(W_1^t, S_0 | y^{t-1}; \gamma\right) = p\left(W_{1,t}; \gamma\right) p\left(W_1^{t-1}, S_0 | y^{t-1}; \gamma\right)$.

Step 2, Weighting: Assign to each draw $s_0^{t|t-1,i}, w_1^{t|t-1,i}$ the weight q_t^i.

Step 3, Sampling: Draw $\left\{s_0^{t,i}, w_1^{t,i}\right\}_{i=1}^{N}$ with rep. from $\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^{N}$ with probabilities $\left\{q_t^i\right\}_{i=1}^{N}$. If $t < T$ set $t \rightsquigarrow t+1$ and go to step 1. Otherwise go to step 4.

Step 4, Likelihood: Use $\left\{\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^{N}\right\}_{t=1}^{T}$ to compute:

$$p\left(y^T; \gamma\right) \simeq \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} p\left(y_t | w_1^{t|t-1,i}, y^{t-1}, s_0^{t|t-1,i}; \gamma\right)$$

A "Trivial" Application

How do we evaluate the likelihood function $p\left(y^T|\alpha,\beta,\sigma\right)$ of the nonlinear, non-Gaussian process:

$$\begin{aligned} s_t &= \alpha + \beta\frac{s_{t-1}}{1+s_{t-1}} + w_t \\ y_t &= s_t + v_t \end{aligned}$$

where $w_t \sim \mathcal{N}\left(0,\sigma\right)$ and $v_t \sim t\left(2\right)$ given some observables $y^T = \{y_t\}_{t=1}^T$ and s_0.

1. Let $s_0^{0,i} = s_0$ for all i.
2. Generate N i.i.d. draws $\left\{ s_0^{1|0,i}, w^{1|0,i} \right\}_{i=1}^{N}$ from $\mathcal{N}(0, \sigma)$.
3. Evaluate
$$p\left(y_1 | w_1^{1|0,i}, y^0, s_0^{1|0,i}\right) = p_{t(2)}\left(y_1 - \left(\alpha + \beta \frac{s_0^{1|0,i}}{1+s_0^{1|0,i}} + w^{1|0,i}\right)\right).$$
4. Evaluate the relative weights $q_1^i = \frac{p_{t(2)}\left(y_1 - \left(\alpha + \beta \frac{s_0^{1|0,i}}{1+s_0^{1|0,i}} + w^{1|0,i}\right)\right)}{\sum_{i=1}^{N} p_{t(2)}\left(y_1 - \left(\alpha + \beta \frac{s_0^{1|0,i}}{1+s_0^{1|0,i}} + w^{1|0,i}\right)\right)}$.
5. Resample with replacement N values of $\left\{ s_0^{1|0,i}, w^{1|0,i} \right\}_{i=1}^{N}$ with relative weights q_1^i. Call those sampled values $\left\{ s_0^{1,i}, w^{1,i} \right\}_{i=1}^{N}$.
6. Go to step 1, and iterate 1-4 until the end of the sample.

A Law of Large Numbers

A law of the large numbers delivers:

$$p\left(y_1 \middle| y^0, \alpha, \beta, \sigma\right) \simeq \frac{1}{N}\sum_{i=1}^{N} p\left(y_1 \middle| w_1^{1|0,i}, y^0, s_0^{1|0,i}\right)$$

and consequently:

$$p\left(y^T \middle| \alpha, \beta, \sigma\right) \simeq \prod_{t=1}^{T} \frac{1}{N}\sum_{i=1}^{N} p\left(y_t \middle| w_1^{t|t-1,i}, y^{t-1}, s_0^{t|t-1,i}\right)$$