Nonlinear state-space models

A. Motivation

Things we lose from linearization:
(1) Statistical representation of recessions.

\[ y_t = \mu_s + \epsilon_t \]

\[ P(s_t = j | s_{t-1} = i) = p_{ij} \quad i,j = 1,2 \]

\[ \Rightarrow y_t \sim ARMA(1,1) \]

Could find optimal linear projection using Kalman filter

\[ \hat{y}_t = a + h \hat{x}_{t+1|t} \]

\[ \hat{x}_{t+1|t} \text{ is linear in } y_t, y_{t-1}, \ldots, y_1 \]

But optimal forecast is

\[ \hat{y}_t = \mu_1 P(s_{t+1} = 1 | y_t, y_{t-1}, \ldots, y_1) \]

\[ + \quad + \mu_0 P(s_{t+1} = 0 | y_t, y_{t-1}, \ldots, y_1) \]

which is nonlinear in \( y_t, y_{t-1}, \ldots, y_1 \)
Figure 1. Unemployment

Spectrum of the Unemployment Rate

Period of cycle (years)
\[ y_t = 0.060 + 1.117 y_{t-1} - 0.128 y_{t-2} + 0.158 v_t \]
\[\text{Student t (4.42)}\]
Things we lose from linearization:

(2) Economic characterization of risk aversion.

\[ l = E_t \left[ \frac{\beta U'(c_{t+1})(1+r_{j,t+1})}{U'(c)} \right] \]

for \( r_{j,t+1} \) the real return on any asset. Finance: different assets have different expected returns due to covariance between \( r_{j,t+1} \) and \( c_{t+1} \).

\[ l = E_t \left[ \frac{\beta U'(c_{t+1})(1+r_{j,t+1})}{U'(c)} \right] \]

steady state:

\[ l = \frac{\beta U'(c)(1+r_j)}{U'(c)} \]

\[ \beta(1 + r_j) = 1 \text{ for all } j \]

linearization around steady state

\[ U'(c) = E_t[\beta U'(c_{t+1})(1 + r_{j,t+1})] \]

\[ \approx (1 + r)\beta U''(c)E_t(c_{t+1} - c) + \beta U'(c)E_t(r_{j,t+1} - r) \]

same for all \( j \).
Things we lose from linearization:
(3) Role of changes in uncertainty, time-varying volatility.
(4) Behavior of economy when interest rate is at zero lower bound
\[ R_t = \min(R^*_t, \bar{R}) \]

Nonlinear state-space models
A. Motivation
B. Extended Kalman filter

Linear state-space model:
State equation:
\[
\begin{align*}
\xi_{t+1} & = F \xi_t + v_{t+1} \\
& \quad \text{with } v_{t+1} \sim N(0, Q)
\end{align*}
\]
Observation equation:
\[
\begin{align*}
y_t & = A' x_t + H' \xi_t + w_t \\
& \quad \text{with } w_t \sim N(0, R)
\end{align*}
\]
Nonlinear state-space model:

State equation:
\[
\xi_{t+1} = \phi(\xi_t) + v_{t+1} \quad v_{t+1} \sim N(0, Q)
\]

Observation equation:
\[
y_t = a(x_t) + h(\xi_t) + w_t \quad w_t \sim N(0, R)
\]

Suppose at date \( t \) we have approximation to distribution of \( \xi_t \) conditional on
\[
\Omega_t = \{ y_t, y_{t-1}, \ldots, y_1, x_t, x_{t-1}, \ldots, x_1 \}
\]
\[
\xi_t | \Omega_t \sim N(\hat{\xi}_{t|t}, P_{t|t})
\]
goal: calculate \( \hat{\xi}_{t+1|t+1}, P_{t+1|t+1} \)

State equation:
\[
\xi_{t+1} = \phi(\xi_t) + v_{t+1}
\]
\[
\phi(\xi_t) \approx \phi_t + \Phi_t (\xi_t - \hat{\xi}_{t|t})
\]
\[
\phi_t = \phi(\hat{\xi}_{t|t})
\]
\[
\Phi_t = \left. \frac{\partial \phi(\xi_t)}{\partial \xi_t'} \right|_{\xi_t = \hat{\xi}_{t|t}}
\]

\[r \times 1\]

\[r \times r\]
Forecast of state vector:
\[
\xi_{t+1} = \phi_t + \Phi_t(\xi_t - \hat{\xi}_{t|t}) + v_{t+1} \\
\hat{\xi}_{t+1|t} = \phi_t = \phi(\hat{\xi}_{1|1}) \\
P_{t+1|t} = \Phi_tP_{t|t}\Phi_t' + Q
\]

Observation equation:
\[
y_t = a(x_t) + h(\xi_t) + w_t \\
h(\xi_t) = h_t + H'_t(\xi_t - \hat{\xi}_{t|t-1}) \\
h_t = h(\hat{\xi}_{t|t-1}) \\
H'_t = \frac{\partial h(\xi_t)}{\partial \xi_t} \bigg|_{\xi_t = \hat{\xi}_{t|t-1}}
\]
Note \(x_t\) is observed so no need to linearize \(a(x_t)\)

Approximating state equation:
\[
\xi_{t+1} = \phi_t + \Phi_t(\xi_t - \hat{\xi}_{t|t}) + v_{t+1}
\]
Approximating observation equation:
\[
y_t = a(x_t) + h_t + H'_t(\xi_t - \hat{\xi}_{t|t-1}) + w_t
\]
A state-space model with time-varying coefficients
Forecast of observation vector:
\[ y_{t+1} = a(x_{t+1}) + h_{t+1} + H'_{t+1}(\xi_{t+1} - \hat{\xi}_{t+1|t}) + w_{t+1} \]
\[ \hat{y}_{t+1|t} = a(x_{t+1}) + h_{t+1} \]
\[ = a(x_{t+1}) + h(\hat{\xi}_{t+1|t}) \]
\[ E(y_{t+1} - \hat{y}_{t+1|t})(y_{t+1} - \hat{y}_{t+1|t})' = H'_{t+1}P_{t+1|t}H_{t+1} + R \]

Updated inference:
\[ \hat{\xi}_{t+1|t+1} = \hat{\xi}_{t+1|t} + K_{t+1}(y_{t+1} - \hat{y}_{t+1|t}) \]
\[ K_{t+1} = P_{t+1|t}H_{t+1}(H'_{t+1}P_{t+1|t}H_{t+1} + R)^{-1} \]
Start from \( \hat{\xi}_{0|0} \) and \( P_{0|0} \) reflecting prior information

Approximate log likelihood:
\[ -\frac{Tn}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^{T} \log |\Omega_t| \]
\[ - \frac{1}{2} \sum_{t=1}^{T} \varepsilon_t' \Omega_t^{-1} \varepsilon_t \]
\[ \Omega_t = H'_{t}P_{t|t-1}H_{t} + R \]
\[ \varepsilon_t = y_t - a(x_t) - h(\hat{\xi}_{t|t-1}) \]
Nonlinear state-space models

A. Motivation
B. Extended Kalman filter
C. Importance sampling

Suppose we want to sample from a density that we can only calculate up to a constant:

\[ p(x) = k q(z) \]

where we know \( q(.) \) but not \( k \)

Examples:
(1) Calculate conditional density want to know \( p(x|y) \)
do know \( p(y|x) \) and \( p(x) \)

\[
p(x|y) = \frac{p(y|x)p(x)}{\int_{x} p(y|x)p(x)dx} = kp(y|x)p(x)
\]
can’t calculate \( \int_{x} p(y|x)p(x)dx \)
(2) Generic Bayesian problem:

\[ p(Y|\theta) = \text{likelihood (known)} \]

\[ p(\theta) = \text{prior (known)} \]

goal: calculate

\[ p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{G} \]

for \( G = \int p(Y|\theta)p(\theta)d\theta \)

Analytical approach: choose \( p(\theta) \)
from a family such that \( G \) can be found with clever algebra.

Numerical approach: satisfied to be able to generate draws

\( \theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(D)} \)

from the distribution \( p(\theta|Y) \) without ever knowing the distribution (i.e., without calculating \( G \))

Importance sampling:

Step (1): Generate \( \theta^{(j)} \) from an (essentially arbitrary) “importance density” \( g(\theta) \).

Step (2): Calculate

\[ \omega^{(j)} = \frac{p(Y|\theta^{(j)})p(\theta^{(j)})}{g(\theta^{(j)})} \cdot \]

Step (3): Weight the draw \( \theta^{(j)} \) by \( \omega^{(j)} \) to simulate distribution of \( p(\theta|Y) \).
Examples:

\[ E(\theta|Y) = \int \theta p(\theta|Y) d\theta \]
\[ \approx \frac{\sum_{j=1}^{D} \theta^{(j)} \omega^{(j)}}{\sum_{j=1}^{D} \omega^{(j)}} \]
\[ \equiv \theta^* \]

\[ \text{Var}(\theta|Y) \approx \frac{\sum_{j=1}^{D} (\theta^{(j)} - \theta^*)(\theta^{(j)} - \theta^*)^T \omega^{(j)}}{\sum_{j=1}^{D} \omega^{(j)}} \]

\[ \text{Prob}(\theta_2 < 0) \approx \frac{\sum_{j=1}^{D} \delta_{\theta^{(j)} < \omega} \omega^{(j)}}{\sum_{j=1}^{D} \omega^{(j)}} \]
How does this work?

\[
\frac{\sum_{j=1}^{D} \theta_{(j)} \omega_{(j)}}{\sum_{j=1}^{D} \omega_{(j)}} = \frac{D^{-1} \sum_{j=1}^{D} \theta_{(j)} \omega_{(j)}}{D^{-1} \sum_{j=1}^{D} \omega_{(j)}}
\]

Numerator:

\[
D^{-1} \sum_{j=1}^{D} \theta_{(j)} \omega_{(j)} \overset{P}{\rightarrow} E[\theta_{(j)} \omega_{(j)}]
\]

\[
= \int \theta \omega(\theta) g(\theta) d\theta
\]

\[
= \int \theta p(Y|\theta) p(\theta) g(\theta) d\theta
\]

\[
= \int \theta p(Y|\theta) p(\theta) d\theta
\]

Denominator:

\[
D^{-1} \sum_{j=1}^{D} \omega_{(j)} \overset{P}{\rightarrow} E[\omega_{(j)}]
\]

\[
= \int \omega(\theta) g(\theta) d\theta
\]

\[
= \int p(Y|\theta) p(\theta) g(\theta) d\theta
\]

\[
= \int p(Y|\theta) p(\theta) d\theta
\]

\[
p(Y)
\]
Conclusion:

\[
\frac{\sum_{j=1}^{D} \theta^{(j)} x^{(j)}}{\sum_{j=1}^{D} x^{(j)}} \xrightarrow{p} \int \frac{\theta p(y|\theta)p(\theta)d\theta}{p(y)} = \int \theta p(y|\theta)d\theta
\]

What’s required of \( g(.) \)?

\[
\theta^{(j)} x^{(j)} = \frac{g(y|\theta^{(j)})p(\theta^{(j)})}{g[\theta^{(j)}]} \quad \text{should satisfy Law of Large Numbers.}
\]

Khintchine’s Theorem: If \( \{x_j\}_{j=1}^{D} \) is i.i.d. with finite mean \( \mu \), then \( D^{-1} \sum_{j=1}^{D} x_j \xrightarrow{p} \mu \)

Note:
- does not require \( x_j \) to have finite variance
- \( \theta^{(j)} \) are drawn i.i.d. from \( g(\theta) \) by construction
So we only need
\[
E(\theta|Y) = \int_{\mathcal{N}} \theta p(\theta|Y) d\theta \text{ exists}
\]
\[
p(\theta|Y) = kp(Y|\theta)p(\theta)
\]
support of \( g(\theta) \) includes \( \mathcal{N} \)

However, convergence may be very slow if variance of
\[
\frac{\theta^{(l)} p(Y|\theta^{(l)}) p(\theta^{(l)})}{g(\theta^{(l)})}
\]
is infinite.

Practical observations:
- works best if \( g(\theta) \) has fatter tails than \( p(Y|\theta)p(\theta) \)
- works best when \( g(\theta) \) is a good approximation to \( p(\theta|Y) \)

Always produces an answer, good idea to check it.

1. Try special cases where result is known analytically.
2. Try different \( g(\cdot) \) to see if get the same result.
3. Use analytic results for components of \( \theta \) in order to keep dimension that must be importance-sampled small.
Nonlinear state-space models

A. Motivation
B. Extended Kalman filter
C. Importance sampling
D. Particle filter

State equation:

$$\xi_{t+1} = \phi_t(\xi_{t}, v_{t+1})$$

Observation equation:

$$y_t = h_t(\xi_{t}, w_t)$$

$\phi_t(.)$ and $h_t(.)$ known functions

(may depend on unknown $\theta$)

$\{w_t, v_t\}$ have known distribution (e.g., i.i.d., perhaps depend on $\theta$)

$\Omega_t = \{y_t, y_{t-1}, \ldots, y_1\}$
$\Lambda_t = \{\xi_t, \xi_{t-1}, \ldots, \xi_0\}$

output for step $t$:

$$p(\Lambda_t|\Omega_t)$$

represented by a series of particles:

$$\{\xi^{(i)}_t, \xi^{(i)}_{t-1}, \ldots, \xi^{(i)}_0\}_{i=1}^{P}$$
Goal: use observations on history of $y_t$ through date $t$,

$$\Omega_t = \{y_t, y_{t-1}, \ldots, y_1\}$$

to form inference about states

$$\Lambda_t = \{\xi_t, \xi_{t-1}, \ldots, \xi_0\}$$

and also evaluate likelihood

$$p(y_t | \Omega_{t-1})$$

Input for step $t + 1$ of iteration

approximation to $p(\Lambda_t | \Omega_t)$

Output for step $t + 1$:

approximation to $p(\Lambda_{t+1} | \Omega_{t+1})$

Particle $i$ is associated with weights $\hat{\omega}_t^{(i)}$

summing to one such that particles

can be used to simulate draw from

$p(\Lambda_t | \Omega_t)$, e.g.

$$E(\xi_{t-1} | \Omega_t) = \sum_{i=1}^{D} \hat{\xi}_{t-1}^{(i)} \hat{\omega}_t^{(i)}$$
Output of step \( t + 1 \): 
\[
p(\Lambda_{t+1}|\Omega_{t+1})
\]
keep particles \( \{\xi_t^{(i)}, \xi_{t-1}^{(i)}, \ldots, \xi_0^{(i)}\}_{i=1}^D \)
append \( \{\xi_t^{(i)}\}_{i=1}^D \) and recalculate weights \( \hat{\omega}_t^{(i)} \)
byproduct: 
\[
p(y_{t+1}|\Omega_t)
\]

Method: Sequential Importance Sampling
At end of step \( t \) have generated 
\[
\Lambda_t^{(i)} = \{\xi_t^{(i)}, \xi_{t-1}^{(i)}, \ldots, \xi_0^{(i)}\}
\]
from some known importance density 
\[
g_t(\Lambda_t|\Omega_t) = \tilde{g}_t(\xi_t|\Lambda_{t-1}, \Omega_t)g_{t-1}(\Lambda_{t-1}|\Omega_{t-1})
\]
We will also have calculated (up to a constant that does not depend on \( \xi_t \)) the true value of \( p_t(\Lambda_t|\Omega_t) \) so weight for particle \( i \) is proportional to 
\[
\omega_t^{(i)} = \frac{p_t(\Lambda_t^{(i)}|\Omega_t)}{g_t(\Lambda_t^{(i)}|\Omega_t)}
\]
\[ \omega_t^{(i)} = \frac{p_t(\Lambda_t^{(i)} | \Omega_t)}{g_t(\Lambda_t^{(i)} | \Omega_t)} \]

**Step i + 1:**

\[ p_{t+1}(\Lambda_{t+1} | \Omega_{t+1}) = \frac{p(y_{t+1} | \xi_{t+1}) p(\xi_{t+1} | \xi_t) p_t(\Lambda_t | \Omega_t)}{p(y_{t+1} | \Omega_t)} \]

\[ \propto p(y_{t+1} | \xi_{t+1}) p(\xi_{t+1} | \xi_t) p_t(\Lambda_t | \Omega_t) \]

known from obs eq known from state eq known at \( t \)

\[ \omega_{t+1}^{(i)} = \frac{p_{t+1}(\Lambda_{t+1}^{(i)} | \Omega_{t+1})}{g_{t+1}(\Lambda_{t+1}^{(i)} | \Omega_{t+1})} \]

\[ \propto \frac{p(y_{t+1} | \xi_{t+1}^{(i)}) p(\xi_{t+1}^{(i)} | \xi_t^{(i)}) p_t(\Lambda_t^{(i)} | \Omega_t)}{g_{t+1}(\xi_{t+1}^{(i)} | \Lambda_t^{(i)} \Omega_{t+1}) g_t(\Lambda_t^{(i)} | \Omega_t)} \]

\[ = \frac{p(y_{t+1} | \xi_{t+1}^{(i)}) p(\xi_{t+1}^{(i)} | \xi_t^{(i)})}{g_{t+1}(\xi_{t+1}^{(i)} | \Lambda_t^{(i)} \Omega_{t+1})} \frac{p_t(\Lambda_t^{(i)} | \Omega_t)}{g_t(\Lambda_t^{(i)} | \Omega_t)} \]

\[ = \tilde{\omega}_{t+1}^{(i)} \omega_t^{(i)} \]

\[ \hat{\omega}_t^{(i)} = \frac{\omega_t^{(i)}}{\sum_{i=1}^{D} \omega_t^{(i)}} \]

\[ \hat{E}(\xi_{t-1} | \Omega_t) = \sum_{i=1}^{D} \hat{\omega}_t^{(i)} \xi_{t-1}^{(i)} \]

\[ \hat{P} (\xi_{1,t} > 0 | \Omega_t) = \sum_{i=1}^{D} \hat{\omega}_t^{(i)} \delta_{[\xi_{1,t}>0]} \]
\[
\tilde{\omega}_{t+1}^{(i)} = \frac{p(y_{t+1}|\tau^t)^{\frac{1}{2}}p(\xi_{t+1}|y_{t+1})}{\tilde{p}(y_{t+1}|\Omega_t)}
\]
\[
\tilde{p}(y_{t+1}|\Omega_t) = \sum_{i=1}^{D} \tilde{\omega}_{t+1}^{(i)} \tilde{\omega}_{t+1}^{(i)}
\]
\[
\hat{\mathcal{L}}(\theta) = \sum_{t=1}^{T} \log \tilde{p}(y_{t+1}|\Omega_{t-1})
\]

Classical: choose \( \theta \) to max \( \hat{\mathcal{L}}(\theta) \)

Bayesian: draw \( \theta \) from posterior distribution by embedding \( \hat{\mathcal{L}}(\theta) \)
in Metropolis-Hastings algorithm

Flury and Shephard, Econometric Theory, 2011

How start algorithm for \( t = 0 \)?

Draw \( \xi_{0}^{(i)} \) from \( p(\xi_{0}) \)
(prior distribution or hypothesized unconditional distribution)

How choose importance density \( \tilde{\xi}_{t+1}(\xi_{t+1}|\Lambda_{t}, \Omega_{t+1}) \)?

(1) Bootstrap filter
\[
\tilde{\xi}_{t+1}(\xi_{t+1}|\Lambda_{t}, \Omega_{t+1}) = p(\xi_{t+1}|\xi_{t})
\]
known from state equation
\[
\xi_{t+1} = \phi_{t}(\xi_{t}, v_{t+1})
\]
But better performance from adaptive filters that also use \( y_{t+1} \)
Note that for bootstrap filter
\[ \tilde{g}_{t+1}(\xi_{t+1}|\Lambda_t, \Omega_{t+1}) = p(\xi_{t+1}|\xi_t) \]
\[ \tilde{\omega}_t^{(i)} = \frac{p(y_t|\xi_t^{(i)})p(\xi_t^{(i)}|\xi_t)}{\tilde{g}_{t+1}(\xi_t^{(i)}|\Lambda_t^{(i)}, \Omega_{t+1})} = p(y_t|\xi_t^{(i)}) \]

Separate problem for particle filter:
one history \( \Lambda_t^{(i)} \) comes to dominate
the others (\( \tilde{\omega}_t^{(i)} \to 1 \) for some \( i \))

Partial solution to degeneracy problem:
Sequential Importance Sampling
with Resampling
Before finishing step \( t \), now resample
\( \{\Lambda_t^{(j)}\}_{j=1}^\mathcal{D} \) with replacement
by drawing from the distribution
\[ \Lambda_t^{(i)} = \begin{cases} 
\Lambda_t^{(1)} & \text{with probability } \tilde{\omega}_t^{(1)} \\
\vdots & \\
\Lambda_t^{(\mathcal{D})} & \text{with probability } \tilde{\omega}_t^{(\mathcal{D})}
\end{cases} \]
Result: repopulate $\{\Lambda_t^{(j)}\}$ by replicating most likely elements
(weights for $\Lambda_t^{(j)}$ are now $\hat{\omega}_t^{*^{(j)}} = 1/D$)

(1) Resampling does not completely solve degeneracy because early-sample elements of
$\Lambda_t^{(j)} = \{\xi_t^{(j)}, \xi_{t-1}^{(j)}, \ldots, \xi_0^{(j)}\}$ will tend to be the same for all $j$ as $t$ gets large
(2) Does help in the sense that have full set of particles to grow from $t$ forward

(3) Have good inference about $p(\xi_{t-k}|\Omega_t)$ for small $k$
(4) Have poor inference about $p(\xi_{t-k}|\Omega_t)$ for large $k$
   (separate smoothing algorithm can be used if goal is $p(\xi_t|\Omega_T)$)
(5) Better to form inference about $p(\xi_{t-k} | \Omega_t)$ or $p(\mathbf{y}_t | \Omega_{t-1})$ using 
$\{\Lambda_i^{(j)}\}$ and weights $\{\hat{w}_i^{(j)}\}$ rather than
$\{\Lambda_i^{(j)}\}$ and weights $\{1/D\}$
(6) Better not to resample every $t$

How choose importance density $\tilde{g}_{t+1}(\xi_{t+1} | \Lambda_t, \Omega_{t+1})$?

(1) Bootstrap filter:
$\tilde{g}_{t+1}(\xi_{t+1} | \Lambda_t, \Omega_{t+1}) = p(\xi_{t+1} | \xi_{t})$

(2) Auxiliary particle filter:
use $\mathbf{y}_{t+1}$ to get better proposal density for $\xi_{t+1}$

Example: from state equation
$\xi_{t+1} = \phi_i(\xi_{t}, \mathbf{v}_{t+1})$
we have guess for likely value for $\xi_{t+1}$ associated with particle $i$.
e.g. $\xi^{(i)}_{t+1} = \phi_i(\xi^{(i)}_{t}, 0)$
And from observation equation we can calculate how likely it would be to observe $y_{t+1}$ if $\xi_{t+1}$ took on this value:

$$p(y_{t+1}|\tilde{\xi}^{(i)}_{t+1}) = p[y_{t+1}|\phi_{t}(\xi^{(i)}_{t}, 0)] = \tilde{r}(\xi^{(i)}_{t}, y_{t+1})$$

Calculate

$$\tau_{t}^{(i)} = \tilde{r}_{t}^{(i)} \omega_{t}^{(i)}$$
$$\tilde{r}_{t}^{(i)} = \tilde{r}(\xi^{(i)}_{t}, y_{t+1})$$
$$\hat{r}_{t}^{(i)} = \frac{r_{t}^{(i)}}{\sum_{i=1}^{D} r_{t}^{(i)}}$$

Resample historical particles with prob $\hat{r}_{t}^{(i)}$

$$\Lambda_{t}^{(i)} = \begin{cases} 
\Lambda_{t}^{(1)} & \text{with probability } \hat{r}_{t}^{(1)} \\
\vdots \\
\Lambda_{t}^{(P)} & \text{with probability } \hat{r}_{t}^{(P)}
\end{cases}$$
Draw $\xi_t^{(j)}$ from proposal density
$$\tilde{g}_{t+1}(\xi_t^{(j)} | \Lambda_t^{(j)}, \Omega_{t+1}) = p(\xi_t^{(j)} | \xi_t^{(j)})$$
From what importance density did we generate proposed $\Lambda_{t+1}^{(j)}$?

If we had resampled using original weights proportional to $\omega_t^{(j)}$, then
$\Lambda_t^{(j)}$ would represent an i.i.d. sample of size $D$ drawn from $p_t(\Lambda_t | \Omega_t)$.
When we resampled using weights proportional to $\tilde{\tau}(\xi_t^{(j)}, y_{t+1})\omega_t^{(j)}$,
$\Lambda_t^{(j)}$ represents an i.i.d. sample with density proportional to
$\tilde{\tau}(\Lambda_t, y_{t+1})p_t(\Lambda_t | \Omega_t)$.

$\xi_t^{(j)}$ was then drawn from $p(\xi_t^{(j)} | \xi_t^{(j)})$.
Proposal density evaluated at $\Lambda_{t+1}^{(j)}$ is thus
$$g_{t+1}(\Lambda_{t+1}^{(j)} | \Omega_{t+1}) = p(\xi_{t+1}^{(j)} | \xi_{t+1}^{(j)})\tilde{\tau}(\Lambda_t^{(j)}, y_{t+1})p(\Lambda_t^{(j)} | \Omega_t)$$
Target density is

\[ p_{t+1}(\Lambda_{t+1}|\Omega_{t+1}) \]
\[ \propto p(y_{t+1}|\xi_{t+1})p(\xi_{t+1}|\xi_{t})p_{t}(\Lambda_{t}|\Omega_{t}) \]

Desired weights are thus proportional to

\[ \omega^{(i)}_{t+1} = \frac{p(y_{t+1}|\xi_{t+1}^{(i)})p(\xi_{t+1}^{(i)}|\xi_{t}^{(i)})p_{t}(\Lambda_{t}^{(i)}|\Omega_{t})}{p(\xi_{t+1}^{(i)}|\xi_{t}^{(i)})\tilde{\tau}(\Lambda_{t}^{(i)}, y_{t+1})p_{t}(\Lambda_{t}^{(i)}|\Omega_{t})} \]
\[ = \frac{p(y_{t+1}|\xi_{t+1}^{(i)})}{\tilde{\tau}(\Lambda_{t}^{(i)}, y_{t+1})} \]

Summary of auxiliary particle filter:
1. Calculate measure of how useful \( \Lambda_{t}^{(i)} \) is for predicting \( y_{t+1} \), e.g.

\[ \tilde{\tau}^{(i)}_{t} = p[y_{t+1}|\xi_{t+1}^{(i)} = \phi_{t}(\xi_{t}^{(i)}, 0)] \]

where \( p(y_{t+1}|\xi_{t+1}^{(i)}) \) comes from obs eq and \( \xi_{t+1}^{(i)} = \phi_{t}(\xi_{t}, \nu_{t+1}) \) is state eq
(2) Resample $\Lambda_t^{(i)}$ from $\Lambda_t^{(i)}$ with probabilities proportional to $z_t^{(i)} \omega_t^{(i)}$

(3) Generate $\xi_{t+1}^{(j)}$ from $\phi_t(\xi_t^{(j)}, v_{t+1})$

(4) Calculate weights

$$\omega_{t+1}^{(j)} = \frac{p(y_{t+1}|z_{t+1}^{(j)})}{z_t^{(j)}}$$

$$\hat{p}(y_{t+1}|\Omega_t) = \sum_{i=1}^{D} \omega_t^{(j)} \hat{\omega}_t^{(j)}$$

$$\hat{\omega}_t^{(j)} = \frac{\omega_t^{(j)}}{\sum_{j=1}^{D} \omega_t^{(j)}}$$

$$\hat{E}(\xi_{t+1}|\Omega_{t+1}) = \sum_{j=1}^{D} \hat{\omega}_{t+1}^{(j)} \xi_{t+1}^{(j)}$$

$$\hat{L}(\theta) = \sum_{t=0}^{T-1} \log \hat{p}(y_{t+1}|\Omega_t)$$

---

**Nonlinear state-space models**

A. Motivation
B. Extended Kalman filter
C. Importance sampling
D. Particle filter
E. Example: estimating a DSGE using higher-order approximation (Fernandez-Villaverde and Rubio-Ramirez, REStud, 2007)
Background on perturbation methods

Example:

\[
\begin{align*}
\max_{\{c_t, k_{t+1}\}} & \quad E_0 \sum_{t=0}^{\infty} \beta^t \log c_t \\
\text{s.t.} & \quad c_t + k_{t+1} = e^{\varepsilon_t} k_t^a + (1 - \delta) k_t \quad t = 1, 2, \ldots \\
& \quad z_t = \rho z_{t-1} + \sigma \varepsilon_t \quad t = 1, 2, \ldots \\
& \quad k_0, z_0 \text{ given} \\
& \quad \varepsilon_t \sim N(0, 1)
\end{align*}
\]

Approach: we will consider a continuum of economies indexed by \( \sigma \) and study solutions as \( \sigma \to 0 \) (that is, as economy becomes deterministic).

We seek decision rules of the form

\[
\begin{align*}
c_t &= c(k_t, z_t; \sigma) \\
k_{t+1} &= k(k_t, z_t; \sigma)
\end{align*}
\]

Write F.O.C. as \( E_t a(k_t, z_t; \sigma, \varepsilon_{t+1}) = 0 \)

\[
\begin{align*}
a_1(k_t, z_t; \sigma, \varepsilon_{t+1}) &= \frac{1}{c(k_t, z_t; \sigma)} - \\
& \quad \beta \frac{ak(k_t, z_t; \sigma)^{\varepsilon_t} \exp(\rho z_{t+1})}{c(k_t, z_t; \sigma) \exp(\rho z_{t+1})} \\
a_2(k_t, z_t; \sigma, \varepsilon_{t+1}) &= c(k_t, z_t; \sigma) + k(k_t, z_t; \sigma) \\
& \quad - e^{\varepsilon_t} k_t^a - (1 - \delta) k_t
\end{align*}
\]
Zero-order approximation
(deterministic steady state)
\[
\sigma = 0 \\
z_t = z = 0 \\
k_t = k \\
a(k, 0; 0) = 0
\]

First-order approximation:
Since \( E_{\sigma}a(k_t, z_t; \sigma, \varepsilon_{t+1}) = 0 \) for all \( k_t, z_t; \sigma \), it follows that
\[
E_{\sigma}a_k(k_t, z_t; \sigma, \varepsilon_{t+1}) = 0
\]
for \( a_k(k_t, z_t; \sigma, \varepsilon_{t+1}) = \frac{\partial a(k_t, z_t; \sigma, \varepsilon_{t+1})}{\partial k_t} \)
likewise
\[
E_{\sigma}a_z(k_t, z_t; \sigma, \varepsilon_{t+1}) = E_{\sigma}a_{\sigma}(k_t, z_t; \sigma, \varepsilon_{t+1}) = 0
\]

\[
a_1(k, 0; 0) = 0 \\
\quad \Rightarrow \frac{1}{c} - \beta \frac{a_k^{a-1}}{c} = 0 \\
\quad \Rightarrow 1 = \beta a k^{a-1} \\
a_2(k, 0; 0) = 0 \\
\quad \Rightarrow c + k - k^a - (1 - \delta)k \\
\quad \Rightarrow c = k^a - \delta k
\]
\( E_t \left\{ \frac{\partial a_1(k,z;\sigma,s_{n+1})}{\partial k_t} \bigg|_{k_t=k,z_t=0,\sigma=0} \right\} = \)

\[-\frac{1}{c^2} c_k - \frac{\beta a(a-1)k^{a-2}}{c} k_z + \frac{\beta a^{a-1}}{c^2} c_k k_k \]

Since \( c \) and \( k \) are known from previous step, setting this to zero gives us an equation in the unknowns \( c_k \) and \( k_k \) where for example
\[
c_k = \frac{\partial c(k,z;\sigma)}{\partial k_t} \bigg|_{k_t=k,z_t=0,\sigma=0} \]

\[ \frac{\partial a_2(k,z;\sigma)}{\partial k_t} \bigg|_{k_t=k,z_t=0,\sigma=0} = \]

\[ c_k + k_k - a k^{a-1} - (1 - \delta) \]

This is a second equation in \( c_k,k_k \), which together with the first can now be solved for \( c_k,k_k \) as a function of \( c \) and \( k \)

\[ E_t \left\{ \frac{\partial a_1(k,z;\sigma,s_{n+1})}{\partial c_t} \bigg|_{k_t=k,z_t=0,\sigma=0} \right\} = \]

\[-\frac{1}{c^2} c_z - \frac{\beta a(a-1)k^{a-2}}{c} k_z - \frac{\beta a^{a-1} \rho}{c} + \frac{\beta a^{a-1}}{c^2} (c_k k_z + \rho c_z) \]

\[ \frac{\partial a_2(k,z;\sigma)}{\partial c_t} \bigg|_{k_t=k,z_t=0,\sigma=0} = \]

\[ c_z + k_z - k^a \]

setting these to zero allows us to solve for \( c_z,k_z \)
\[
\frac{\partial a_1(k_t, z_t; \sigma, \varepsilon_{t+1})}{\partial \sigma}
\bigg|_{k_t = k, z_t = 0, \sigma = 0} = \frac{-1}{c^2} c\sigma - \frac{\beta \alpha (\alpha - 1) k^{\alpha - 2}}{c} k\sigma - \frac{\beta ak^{\alpha-1}}{c} \varepsilon_{t+1} \\
+ \frac{\beta ak^{\alpha-1}}{c^2} (c_k k\sigma + \varepsilon_{t+1} c_z + c\sigma) \\
\frac{\partial a_2(k_t, z_t; \sigma)}{\partial \sigma}
\bigg|_{k_t = k, z_t = 0, \sigma = 0} = c\sigma + k\sigma 
\]

Taking expectations and setting to zero yields
\[
\frac{-1}{c^2} c\sigma - \frac{\beta \alpha (\alpha - 1) k^{\alpha - 2}}{c} k\sigma \\
+ \frac{\beta ak^{\alpha-1}}{c^2} (c_k k\sigma + c\sigma) = 0 \\
c\sigma + k\sigma = 0 
\]
which has solution \(c\sigma = k\sigma = 0\)
\(\Rightarrow\) volatility, risk aversion play no role in first-order approximation

Now that we’ve calculated derivatives, we have the approximate solutions
\[c(k_t, z_t; \sigma) \approx c + c_k (k_t - k) + c_z z_t + c\sigma \sigma\]
\[k(k_t, z_t; \sigma) \approx k + k_k (k_t - k) + k_z z_t + k\sigma \sigma\]
where we showed that \(c\sigma = k\sigma = 0\)
Thus, first-order perturbation is a way to find linearization or log-linearization.
But we don’t have to stop here. Since $E_t a(k_t, z_t; \sigma, \varepsilon_{t+1}) = 0$ for all $k_t, z_t, \sigma$, second derivatives with respect to $(k_t, z_t; \sigma)$ also have to be zero.

Differentiate each of the 6 equations

$E_t a_k(k_t, z_t; \sigma, \varepsilon_{t+1}) = 0$
$E_t a_z(k_t, z_t; \sigma, \varepsilon_{t+1}) = 0$
$E_t a_\sigma(k_t, z_t; \sigma, \varepsilon_{t+1}) = 0$

with respect to $k_t, z_t$, and $\sigma$.

Gives 18 linear equations in the 12 unknowns

$\{c_{ij}, k_{ij}\}_{i,j=\{k,z,\sigma\}}$ with 6 equations redundant by symmetry of second derivatives (e.g., $c_{kz} = c_{zk}$) and where coefficients on $c_{ij}, k_{ij}$ are known from previous step.
We then have second-order approximation to decision functions,
\[ c(k_t, z_t; \sigma) \approx c + c'_1 s_t + (1/2) s_t' C_2 s_t \]
\[ k(k_t, z_t; \sigma) \approx k + k'_1 s_t + (1/2) s_t' K_2 s_t \]
\[ c'_1 = \begin{bmatrix} c_k & c_z & 0 \end{bmatrix} \]
\[ s_t = \begin{bmatrix} (k_t - k) & z_t & \sigma \end{bmatrix} \]
\[ k'_1 = \begin{bmatrix} k_k & k_z & 0 \end{bmatrix} \]

\[ C_2 = \begin{bmatrix} c_{kk} & c_{kz} & c_{k\sigma} \\ c_{zk} & c_{zz} & c_{z\sigma} \\ c_{ak} & c_{az} & c_{a\sigma} \end{bmatrix} \]

\[ K_2 = \begin{bmatrix} k_{kk} & k_{kz} & k_{k\sigma} \\ k_{zk} & k_{zz} & k_{z\sigma} \\ k_{ak} & k_{az} & k_{a\sigma} \end{bmatrix} \]

\[ c(k_t, z_t; \sigma) \approx c + c'_1 s_t + (1/2) s_t' C_2 s_t \]
\[ s_t = \begin{bmatrix} (k_t - k) & z_t & \sigma \end{bmatrix} \]

Note: term on \( \sigma^2 \) in \( s_t' C_2 s_t \) acts like another constant reflecting precautionary behavior left out of certainty-equivalence steady-state \( c \)
We could in principle continue to as high an order approximation as we wanted.

\[ C_t + I_t = A_t K_t^\alpha L_t^{1-\alpha} \]

\[ K_{t+1} = (1-\delta)K_t + U_t I_t \]

\[ \log A_t = \zeta + \log A_{t-1} + \sigma_a \epsilon_{at} \]

\[ \log U_t = \theta + \log U_{t-1} + \sigma_v \epsilon_{vt} \]

\[ \log \sigma_{at} = (1-\lambda_a) \log \bar{\sigma}_a + \lambda_a \log \sigma_{a,t-1} + \tau_a \eta_{at} \]

\[ \log \sigma_{vt} = (1-\lambda_v) \log \bar{\sigma}_v + \lambda_v \log \sigma_{v,t-1} + \tau_v \eta_{vt} \]

\[ E_0 \sum_{t=0}^{\infty} \beta^t \{ \epsilon^d_t, \log C_t + \psi \log (1-L_t) \} \]

\[ d_t = \rho d_{t-1} + \sigma_d \epsilon_{dt} \]

\[ \log \sigma_{dt} = (1-\lambda_d) \log \bar{\sigma}_d + \lambda_d \log \sigma_{d,t-1} + \tau_d \eta_{dt} \]
\[ v_t = (\epsilon_{at}, \epsilon_{vt}, \epsilon_{dt}, \eta_{at}, \eta_{vt}, \eta_{dt})' \]
\[ v_t \sim N(0, I_6) \]
\[ \Omega = \text{diag}\{\sigma^2_{\epsilon_{at}}, \sigma^2_{\epsilon_{vt}}, \sigma^2_{\epsilon_{dt}}, \tau^2_{\eta_{at}}, \tau^2_{\eta_{vt}}, \tau^2_{\eta_{dt}}\} \]

perturbation method: Continuum of economies with variance \( \chi \Omega \), take expansion around \( \chi = 0 \)

Transformations to find steady-state representation:
\[ Z_t = A_{t-1}^{1/(1-a)} U_{t-1}^{a/(1-a)} \]
\[ \tilde{Y}_t = Y_t/Z_t, \quad \tilde{C}_t = C_t/Z_t, \quad \tilde{T}_t = I_t/Z_t \]
\[ \tilde{U}_t = U_t/U_{t-1}, \quad \tilde{A}_t = A_t/A_{t-1}, \quad \tilde{K}_t = K_t/Z_t U_{t-1} \]
\[ \tilde{k} = \log \text{of steady-state value for } \tilde{K} \]
\[ \tilde{\tilde{k}}_t = \log \tilde{K}_t - \tilde{k} \]

state vector for economic model:
\[ \tilde{s}_t = (\tilde{k}_t, \epsilon_{at}, \epsilon_{vt}, \epsilon_{dt}, d_{t-1}, \sigma_{\epsilon_{at}} - \sigma_{\epsilon_{at}}, \sigma_{\epsilon_{vt}} - \sigma_{\epsilon_{vt}}, \sigma_{\epsilon_{dt}} - \sigma_{\epsilon_{dt}})' \]
second-order perturbation:
\[ \tilde{\tilde{k}}_{t+1} = \psi_{k1}^t \tilde{s}_t + (1/2) \tilde{s}_t^t \psi_{k2} \tilde{s}_t + \psi_{k0} \]
\[ \tilde{\tilde{\eta}}_t = \psi_{\eta1}^t \tilde{s}_t + (1/2) \tilde{s}_t^t \psi_{\eta2} \tilde{s}_t + \psi_{\eta0} \]
\[ \tilde{\hat{\eta}}_t = \psi_{\epsilon1}^t \tilde{s}_t + (1/2) \tilde{s}_t^t \psi_{\epsilon2} \tilde{s}_t + \psi_{\epsilon0} \]
\[ \psi_{\eta0} \text{ reflects precautionary effects} \]
However, we will observe actual GDP growth per capita
\[ \Delta \log Y_t = \Delta \log \tilde{Y}_t \]
\[ + \frac{1}{1-\lambda} (\Delta \log A_{t-1} + \alpha \Delta \log U_{t-1}) + \sigma_{yt} \varepsilon_{yt} \]
\[ = h_t (\tilde{s}_t, \tilde{s}_{t-1}) + \sigma_{yt} \varepsilon_{yt} \]
\[ \varepsilon_{yt} = \text{measurement error} \]

Also observe real gross investment per capita \((I_t)\), hours worked per capita \((l_t)\), and relative price of investment goods \(P_t\)
\[ \Delta \log I_t = h_t (\tilde{s}_t, \tilde{s}_{t-1}) + \sigma_{it} \varepsilon_{it} \]
\[ \log l_t = h_t (\tilde{s}_t, \tilde{s}_{t-1}) + \sigma_{it} \varepsilon_{it} \]
\[ \Delta \log P_t = -\Delta \log U_t \]

\[ \tilde{s}_t = (\tilde{k}_t, \tilde{e}_{at}, \tilde{e}_{vt}, \tilde{e}_{dt}, d_{t-1}, \]
\[ \sigma_{at} - \tilde{\sigma}_a, \sigma_{vt} - \tilde{\sigma}_v, \sigma_{dt} - \tilde{\sigma}_d)' \]
\[ \tilde{v}_t = (\varepsilon_{at}, \varepsilon_{vt}, \varepsilon_{dt}, \eta_{at}, \eta_{vt}, \eta_{dt})' \]
\[ \tilde{s}_t = (\tilde{s}_t, \tilde{s}_{t-1}) \]
\[ \text{state equation} \]
\[ \tilde{s}_t = f (\tilde{s}_{t-1}, v_t) \]
\[ f_1 (\tilde{s}_{t-1}, v_t) = \psi_{k1} \tilde{s}_t + (1/2) \tilde{s}_t^\prime \Psi_{k2} \tilde{s}_t + \psi_{k0} \]
\[ f_2(S_{t-1}, v_t) = \varepsilon_{at} \]
\[ \vdots \]
\[ f_5(S_{t-1}, v_t) = \rho d_{t-2} + \sigma_{d,t-1} \varepsilon_{d,t-1} \]
\[ f_6(S_{t-1}, v_t) = \exp[(1 - \lambda_a) \log \bar{\sigma}_a + \lambda_a \log \sigma_{a,t-1} + \tau_a \eta_{at}] - \bar{\sigma}_a \]
\[ \vdots \]
\[ f_{9-16}(S_{t-1}, v_t) = \tilde{s}_{t-1} \]

\[ y_t = (\Delta \log Y_t, \Delta \log I_t, \log \ell_t, \Delta \log P_t)' \]
observation equation:
\[ y_t = h(S_t) + w_t \]

According to the set-up, \( \varepsilon_{vt} \) is observed directly from the change in investment price each period
\[ \log U_t = \theta + \log U_{t-1} + \sigma_{vt} \varepsilon_{vt} \]
\[ \Delta \log P_t = -\Delta \log U_t \]
We only need to generate a draw for
\[ \mathbf{v}_{1t} = (\mathbf{e}_{at}, \mathbf{e}_{dt}, \mathbf{\eta}_{at}, \mathbf{\eta}_{vt}, \mathbf{\eta}_{dt})' \]
in order to have a value for \( \mathbf{vt} \) and value for \( \mathbf{\varepsilon}_{vt} \)
\[ \mathbf{\varepsilon}_{vt} = \frac{-\Delta \log \mathbf{P}_{t+\mathbf{t}}}{\mathbf{\sigma}_{vt}} \]

Initialization:
\( \mathbf{S}_t = \mathbf{f}(\mathbf{S}_{t-1}, \mathbf{v}_t) \)
One approach is to set \( \mathbf{S}_{-N} = \mathbf{0} \), draw \( \mathbf{v}_{-N+1}, \mathbf{v}_{-N+2}, \ldots, \mathbf{v}_0 \) from \( \mathcal{N}(\mathbf{0}, \mathbf{I}_6) \) to obtain \( D \) draws (particles) for \( \{\mathbf{S}_0^{(i)}\}_{i=1}^D \)

Estimation using bootstrap particle filter
As of date \( t \) we have calculated a set
\[ \Lambda_t^{(i)} = \{\mathbf{S}_t^{(i)}, \mathbf{S}_{t-1}^{(i)}, \ldots, \mathbf{S}_0^{(i)}\} \]
for \( i = 1, \ldots, D \)
To update for \( t + 1 \) we do the following:
Step 1: generate $v_{t+1}^{(i)} \sim N(0, I_3)$ for $i = 1, \ldots, D$

Step 2: generate $S_{t+1}^{(i)} = f(\mathbf{S}_t^{(i)}, v_{t+1}^{(i)})$

except for the third element $\mathbf{e}_{t+1}^{(i)}$

Step 3: calculate $w_{t+1}^{(i)} = y_{t+1} - h(S_{t+1}^{(i)})$

and set third element of $S_{t+1}^{(i)}$ equal to fourth element of $w_{t+1}^{(i)}$.

$$\mathbf{e}_{t+1}^{(i)} = -\frac{\Delta \log P_{t+1} + \theta}{\sigma_{t+1}^{(i)}}$$

Step 4: calculate

$$\tilde{\omega}_{t+1}^{(i)} = (2\pi)^{-D/2} |\mathbf{D}_t^{(i)}|$$

$$\times \exp \left( -\frac{1}{2} \left[ w_{t+1}^{(i)} \mathbf{D}_t^{(i)}^{-1} w_{t+1}^{(i)} \right] \right)$$

$$\mathbf{D}_t^{(i)} = \left[ \begin{array}{cccc}
\sigma_{\hat{y}}^2 & 0 & 0 & 0 \\
0 & \sigma_{\hat{e}}^2 & 0 & 0 \\
0 & 0 & \sigma_{\hat{e}}^2 & 0 \\
0 & 0 & 0 & \left( \sigma_{t+1}^{(i)} \right)^2
\end{array} \right]$$

Step 5: Contribution to likelihood is

$$\hat{p}(y_{t+1} | \Omega_t) = D^{-1} \sum_{i=1}^{D} \tilde{\omega}_{t+1}^{(i)} = \overline{\omega}_{t+1}$$

Step 6: Calculate $\hat{\omega}_{t+1}^{(i)} = \tilde{\omega}_{t+1}^{(i)} / \overline{\omega}_{t+1}$

and resample

$$\Lambda_{t+1}^{(i)} = \begin{cases}
\Lambda_{t+1}^{(1)} \text{ with probability } \hat{\omega}_{t+1}^{(1)} \\
\vdots \\
\Lambda_{t+1}^{(D)} \text{ with probability } \hat{\omega}_{t+1}^{(D)}
\end{cases}$$
Structural parameters:
\[ \theta = (a, \delta, \rho, \beta, \psi, \theta, \zeta, \tau_a, \tau_v, \tau_d, \bar{\sigma}_a, \bar{\sigma}_v, \bar{\sigma}_d, \lambda_a, \lambda_v, \lambda_d, \sigma_{\Omega}, \sigma_{\Omega}, \sigma_{\Omega})' \]

Fernandez-Villaverde and Rubio-Ramirez estimate \( \theta \) by maximizing
\[ \hat{L}(\theta) = \sum_{t=1}^{T} \hat{p}(y_t|\Omega_{t-1}; \theta) \]

---

---

---

---