Nonlinear state-space models

A. Motivation

Things we lose from linearization:

(1) Statistical representation of recessions.

$y_t = \mu_{s_t} + \varepsilon_t$

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

$\Rightarrow y_t \sim ARMA(1,1)$

Could find optimal linear projection using Kalman filter

$$\hat{y}_t = a + \mathbf{h}'\hat{\boldsymbol{\xi}}_{t+1|t}$$

$\hat{\boldsymbol{\xi}}_{t+1|t}$ is linear in $y_t, y_{t-1}, \ldots, y_1$

But optimal forecast is

$$\hat{y}_t = \mu_1 P(s_{t+1} = 1 | y_t, y_{t-1}, \ldots, y_1) + \mu_0 P(s_{t+1} = 0 | y_t, y_{t-1}, \ldots., y_1)$$

which is nonlinear in $y_t, y_{t-1}, \ldots, y_1$

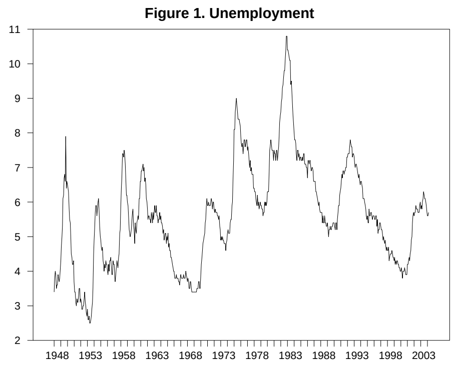


Figure 1. Unemployment
11
10
9
8
7
6
5
4
3
2
1948 1953 1958 1963 1968 1973 1978 1983 1988 1993 1998 2003

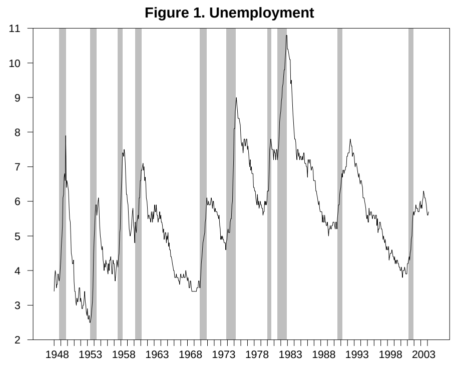
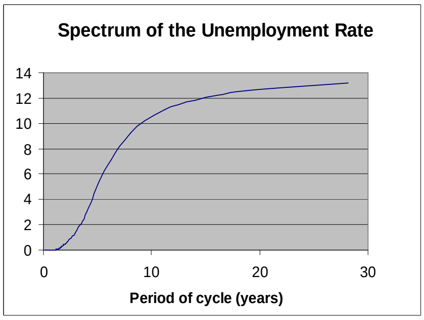


Figure 1. Unemployment
11
10
9
8
7
6
5
4
3
2
1948 1953 1958 1963 1968 1973 1978 1983 1988 1993 1998 2003


Spectrum of the Unemployment Rate
14
12
10
8
6
4
2
0
0 10 20 30
Period of cycle (years)

$$y_t = \underset{(0.028)}{0.060} + \underset{(0.037)}{1.117}\, y_{t-1} - \underset{(0.037)}{0.128}\, y_{t-2} + \underset{(0.007)}{0.158}\, v_t$$

$v_t \sim$ Student t (4.42)

Figure 3. Simulated Unemployment

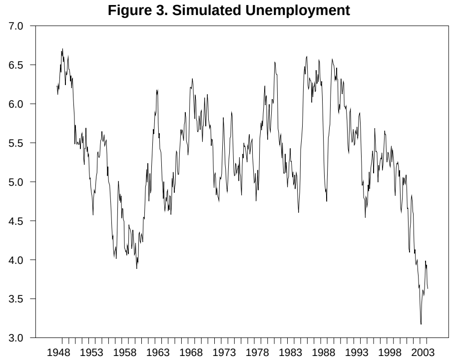
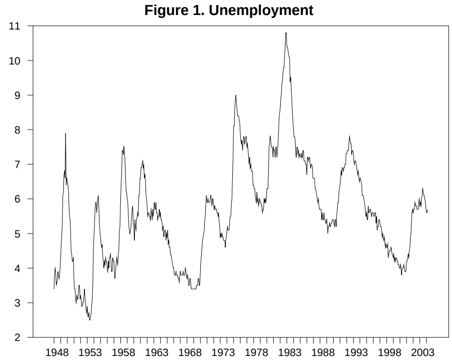


Figure 1. Unemployment



Things we lose from linearization:
(2) Economic characterization of risk aversion.

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

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}

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

steady state:

$$1 = \frac{\beta U'(c)(1+r_j)}{U'(c)}$$

$\beta(1+r_j) = 1$ for all j

linearization around steady state

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

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:

$$\underset{r\times 1}{\boldsymbol{\xi}_{t+1}} = \underset{r\times r}{\mathbf{F}}\,\underset{r\times 1}{\boldsymbol{\xi}_t} + \underset{r\times 1}{\mathbf{v}_{t+1}} \qquad \mathbf{v}_{t+1} \sim N(\mathbf{0}, \mathbf{Q})$$

Observation equation:

$$\underset{n\times 1}{\mathbf{y}_t} = \underset{n\times k}{\mathbf{A}'}\,\underset{k\times 1}{\mathbf{x}_t} + \underset{n\times r}{\mathbf{H}'}\,\underset{r\times 1}{\boldsymbol{\xi}_t} + \underset{n\times 1}{\mathbf{w}_t} \qquad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{R})$$

Nonlinear state-space model:

State equation:

$$\underset{r\times 1}{\xi_{t+1}} = \underset{r\times 1}{\phi(\xi_t)} + \underset{r\times 1}{\mathbf{v}_{t+1}} \qquad \mathbf{v}_{t+1} \sim N(\mathbf{0}, \mathbf{Q})$$

Observation equation:

$$\underset{n\times 1}{\mathbf{y}_t} = \underset{n\times 1}{\mathbf{a}(\mathbf{x}_t)} + \underset{n\times 1}{\mathbf{h}(\xi_t)} + \underset{n\times 1}{\mathbf{w}_t} \qquad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{R})$$

Suppose at date t we have approximation to distribution of ξ_t conditional on

$\Omega_t = \{\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1, \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_1\}$

$\xi_t|\Omega_t \sim N(\hat{\xi}_{t|t}, \mathbf{P}_{t|t})$

goal: calculate $\hat{\xi}_{t+1|t+1}, \mathbf{P}_{t+1|t+1}$

State equation:

$$\xi_{t+1} = \phi(\xi_t) + \mathbf{v}_{t+1}$$

$$\phi(\xi_t) \simeq \phi_t + \mathbf{\Phi}_t(\xi_t - \hat{\xi}_{t|t})$$

$$\underset{r\times 1}{\phi_t} = \phi(\hat{\xi}_{t|t})$$

$$\underset{r\times r}{\mathbf{\Phi}_t} = \left.\frac{\partial \phi(\xi_t)}{\partial \xi_t'}\right|_{\xi_t = \hat{\xi}_{t|t}}$$

Forecast of state vector:

$$\boldsymbol{\xi}_{t+1} = \boldsymbol{\phi}_t + \boldsymbol{\Phi}_t(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t}) + \mathbf{v}_{t+1}$$

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \boldsymbol{\phi}_t = \boldsymbol{\phi}(\hat{\boldsymbol{\xi}}_{t|t})$$

$$\mathbf{P}_{t+1|t} = \boldsymbol{\Phi}_t \mathbf{P}_{t|t} \boldsymbol{\Phi}_t' + \mathbf{Q}$$

Observation equation:

$$\mathbf{y}_t = \mathbf{a}(\mathbf{x}_t) + \mathbf{h}(\boldsymbol{\xi}_t) + \mathbf{w}_t$$

$$\mathbf{h}(\boldsymbol{\xi}_t) \simeq \mathbf{h}_t + \mathbf{H}_t'(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1})$$

$$\underset{n\times 1}{\mathbf{h}_t} = \mathbf{h}(\hat{\boldsymbol{\xi}}_{t|t-1})$$

$$\underset{n\times r}{\mathbf{H}_t'} = \left.\frac{\partial \mathbf{h}(\boldsymbol{\xi}_t)}{\partial \boldsymbol{\xi}_t'}\right|_{\boldsymbol{\xi}_t = \hat{\boldsymbol{\xi}}_{t|t-1}}$$

Note $\mathbf{x}_t$ is observed so no need to linearize $\mathbf{a}(\mathbf{x}_t)$

Approximating state equation:

$$\boldsymbol{\xi}_{t+1} = \boldsymbol{\phi}_t + \boldsymbol{\Phi}_t(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t}) + \mathbf{v}_{t+1}$$

Approximating observation equation:

$$\mathbf{y}_t = \mathbf{a}(\mathbf{x}_t) + \mathbf{h}_t + \mathbf{H}_t'(\boldsymbol{\xi}_t - \hat{\boldsymbol{\xi}}_{t|t-1}) + \mathbf{w}_t$$

A state-space model with time-varying coefficients

Forecast of observation vector:

$$\mathbf{y}_{t+1} = \mathbf{a}(\mathbf{x}_{t+1}) + \mathbf{h}_{t+1} + \mathbf{H}'_{t+1}(\boldsymbol{\xi}_{t+1} - \hat{\boldsymbol{\xi}}_{t+1|t}) + \mathbf{w}_{t+1}$$

$$\hat{\mathbf{y}}_{t+1|t} = \mathbf{a}(\mathbf{x}_{t+1}) + \mathbf{h}_{t+1} = \mathbf{a}(\mathbf{x}_{t+1}) + \mathbf{h}(\hat{\boldsymbol{\xi}}_{t+1|t})$$

$$E(\mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1|t})(\mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1|t})' = \mathbf{H}'_{t+1}\mathbf{P}_{t+1|t}\mathbf{H}_{t+1} + \mathbf{R}$$

Updated inference:

$$\hat{\boldsymbol{\xi}}_{t+1|t+1} = \hat{\boldsymbol{\xi}}_{t+1|t} + \mathbf{K}_{t+1}(\mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1|t})$$

$$\mathbf{K}_{t+1} = \mathbf{P}_{t+1|t}\mathbf{H}_{t+1}(\mathbf{H}'_{t+1}\mathbf{P}_{t+1|t}\mathbf{H}_{t+1} + \mathbf{R})^{-1}$$

Start from $\hat{\boldsymbol{\xi}}_{0|0}$ and $\mathbf{P}_{0|0}$ reflecting prior information

Approximate log likelihood:

$$-\frac{Tn}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log|\boldsymbol{\Omega}_t| - \frac{1}{2}\sum_{t=1}^{T}\boldsymbol{\varepsilon}'_t\boldsymbol{\Omega}_t^{-1}\boldsymbol{\varepsilon}_t$$

$$\boldsymbol{\Omega}_t = \mathbf{H}'_t\mathbf{P}_{t|t-1}\mathbf{H}_t + \mathbf{R}$$

$$\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \mathbf{a}(\mathbf{x}_t) - \mathbf{h}(\hat{\boldsymbol{\xi}}_{t|t-1})$$

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:

$$p(\mathbf{z}) = kq(\mathbf{z})$$

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

Examples:

(1) Calculate conditional density

want to know $p(\mathbf{x}|\mathbf{y})$

do know $p(\mathbf{y}|\mathbf{x})$ and $p(\mathbf{x})$

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{\int_{\chi} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})d\mathbf{x}}$$

$$= kp(\mathbf{y}|\mathbf{x})p(\mathbf{x})$$

can't calculate $\int_{\chi} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})d\mathbf{x}$

(2) Generic Bayesian problem:

$p(\mathbf{Y}|\boldsymbol{\theta})$ = likelihood (known)

$p(\boldsymbol{\theta})$ = prior (known)

goal: calculate

$$p(\boldsymbol{\theta}|\mathbf{Y}) = \frac{p(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{G}$$

for $G = \int p(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$

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

Numerical approach: satisfied to be able to generate draws

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

from the distribution $p(\boldsymbol{\theta}|\mathbf{Y})$ without ever knowing the distribution (i.e., without calculating G)

Importance sampling:

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

Step (2): Calculate

$$\omega^{(j)} = \frac{p[\mathbf{Y}|\boldsymbol{\theta}^{(j)}]p[\boldsymbol{\theta}^{(j)}]}{g[\boldsymbol{\theta}^{(j)}]}.$$

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

Examples:

$$E(\boldsymbol{\theta}|\mathbf{Y}) = \int \boldsymbol{\theta}\, p(\boldsymbol{\theta}|\mathbf{Y}) d\boldsymbol{\theta}$$

$$\simeq \frac{\sum_{j=1}^{D} \boldsymbol{\theta}^{(j)} \omega^{(j)}}{\sum_{j=1}^{D} \omega^{(j)}}$$

$$\equiv \boldsymbol{\theta}^{*}$$

$$\mathrm{Var}(\boldsymbol{\theta}|\mathbf{Y}) \simeq \frac{\sum_{j=1}^{D} (\boldsymbol{\theta}^{(j)} - \boldsymbol{\theta}^{*})(\boldsymbol{\theta}^{(j)} - \boldsymbol{\theta}^{*})' \omega^{(j)}}{\sum_{j=1}^{D} \omega^{(j)}}$$

$$\mathrm{Prob}(\theta_2 < 0) \simeq \frac{\sum_{j=1}^{D} \delta_{[\theta_2^{(j)} < 0]} \omega^{(j)}}{\sum_{j=1}^{D} \omega^{(j)}}$$

How does this work?

$$\frac{\sum_{j=1}^{D} \boldsymbol{\theta}^{(j)} \omega^{(j)}}{\sum_{j=1}^{D} \omega^{(j)}} = \frac{D^{-1} \sum_{j=1}^{D} \boldsymbol{\theta}^{(j)} \omega^{(j)}}{D^{-1} \sum_{j=1}^{D} \omega^{(j)}}$$

Numerator:

$$D^{-1} \sum_{j=1}^{D} \boldsymbol{\theta}^{(j)} \omega^{(j)} \xrightarrow{p} E[\boldsymbol{\theta}^{(j)} \omega^{(j)}]$$

$$= \int \boldsymbol{\theta} \omega(\boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

$$= \int \boldsymbol{\theta} \frac{p(\mathbf{Y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})}{g(\boldsymbol{\theta})} g(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

$$= \int \boldsymbol{\theta} p(\mathbf{Y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

Denominator:

$$D^{-1} \sum_{j=1}^{D} \omega^{(j)} \xrightarrow{p} E[\omega^{(j)}]$$

$$= \int \omega(\boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

$$= \int \frac{p(\mathbf{Y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})}{g(\boldsymbol{\theta})} g(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

$$= \int p(\mathbf{Y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

$$= p(\mathbf{Y})$$

Conclusion:

$$\frac{\sum_{j=1}^{D} \boldsymbol{\theta}^{(j)} \omega^{(j)}}{\sum_{j=1}^{D} \omega^{(j)}} \xrightarrow{p} \frac{\int \boldsymbol{\theta} p(\mathbf{Y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}}{p(\mathbf{Y})}$$

$$= \int \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{Y}) d\boldsymbol{\theta}$$

What's required of $g(.)$?

$\boldsymbol{\theta}^{(j)} \omega^{(j)} = \frac{\boldsymbol{\theta}^{(j)} p[\mathbf{Y}|\boldsymbol{\theta}^{(j)}] p[\boldsymbol{\theta}^{(j)}]}{g[\boldsymbol{\theta}^{(j)}]}$ should satisfy Law of Large Numbers.

Khintchine's Theorem: If $\{\mathbf{x}_j\}_{j=1}^{D}$ is i.i.d. with finite mean $\boldsymbol{\mu}$, then $D^{-1} \sum_{j=1}^{D} \mathbf{x}_j \xrightarrow{p} \boldsymbol{\mu}$

Note:

- does not require $\mathbf{x}_j$ to have finite variance
- $\boldsymbol{\theta}^{(j)}$ are drawn i.i.d. from $g(\boldsymbol{\theta})$ by construction

So we only need

$E(\boldsymbol{\theta}|\mathbf{Y}) = \int_{\aleph} \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{Y}) d\boldsymbol{\theta}$ exists

$p(\boldsymbol{\theta}|\mathbf{Y}) = kp(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$

support of $g(\boldsymbol{\theta})$ includes $\aleph$

However, convergence may be very slow if variance of

$$\frac{\boldsymbol{\theta}^{(j)} p[\mathbf{Y}|\boldsymbol{\theta}^{(j)}] p[\boldsymbol{\theta}^{(j)}]}{g[\boldsymbol{\theta}^{(j)}]}$$

is infinite.

Practical observations:

- works best if $g(\boldsymbol{\theta})$ has fatter tails than $p(\mathbf{Y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$
- works best when $g(\boldsymbol{\theta})$ is good approximation to $p(\boldsymbol{\theta}|\mathbf{Y})$

Always produces an answer, good idea to check it.

(1) Try special cases where result is known analytically.

(2) Try different $g(.)$ to see if get the same result.

(3) Use analytic results for components of $\boldsymbol{\theta}$ 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:

$$\underset{r\times 1}{\boldsymbol{\xi}_{t+1}} = \underset{r\times 1}{\boldsymbol{\phi}_t(\boldsymbol{\xi}_t, \mathbf{v}_{t+1})}$$

Observation equation:

$$\underset{n\times 1}{\mathbf{y}_t} = \underset{n\times 1}{\mathbf{h}_t(\boldsymbol{\xi}_t, \mathbf{w}_t)}$$

$\boldsymbol{\phi}_t(.)$ and $\mathbf{h}_t(.)$ known functions (may depend on unknown $\boldsymbol{\theta}$)

$\{\mathbf{w}_t, \mathbf{v}_t\}$ have known distribution (e.g., i.i.d., perhaps depend on $\boldsymbol{\theta}$)

$$\Omega_t = \{\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1\}$$

$$\Lambda_t = \{\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}, \ldots, \boldsymbol{\xi}_0\}$$

output for step t:

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

represented by a series of particles:

$$\{\boldsymbol{\xi}_t^{(i)}, \boldsymbol{\xi}_{t-1}^{(i)}, \ldots, \boldsymbol{\xi}_0^{(i)}\}_{i=1}^{D}$$

Goal: use observations on history of $\mathbf{y}_t$ through date t,

$$\Omega_t = \{\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1\}$$

to form inference about states

$$\Lambda_t = \{\boldsymbol{\xi}_t, \boldsymbol{\xi}_{t-1}, \ldots, \boldsymbol{\xi}_0\}$$

and also evaluate likelihood

$$p(\mathbf{y}_t | \Omega_{t-1})$$

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.

$$E(\boldsymbol{\xi}_{t-1} | \Omega_t) = \sum_{i=1}^{D} \boldsymbol{\xi}_{t-1}^{(i)} \hat{\omega}_t^{(i)}$$

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 $\{\xi_{t+1}^{(i)}\}_{i=1}^{D}$ and recalculate weights $\hat{\omega}_{t+1}^{(i)}$

byproduct:

$p(\mathbf{y}_{t+1}|\Omega_t)$

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 ξ_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 $t+1$:

$$p_{t+1}(\Lambda_{t+1}|\Omega_{t+1}) = \frac{p(\mathbf{y}_{t+1}|\boldsymbol{\xi}_{t+1})p(\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t)p_t(\Lambda_t|\Omega_t)}{p(\mathbf{y}_{t+1}|\Omega_t)}$$

$$\propto \underbrace{p(\mathbf{y}_{t+1}|\boldsymbol{\xi}_{t+1})}_{\text{known from obs eq}} \; \underbrace{p(\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t)}_{\text{known from state eq}} \; \underbrace{p_t(\Lambda_t|\Omega_t)}_{\text{known at } t}$$

$$\begin{aligned}
\omega_{t+1}^{(i)} &= \frac{p_{t+1}(\Lambda_{t+1}^{(i)}|\Omega_{t+1})}{g_{t+1}(\Lambda_{t+1}^{(i)}|\Omega_{t+1})} \\
&\propto \frac{p(\mathbf{y}_{t+1}|\boldsymbol{\xi}_{t+1}^{(i)})p(\boldsymbol{\xi}_{t+1}^{(i)}|\boldsymbol{\xi}_t^{(i)})p_t(\Lambda_t^{(i)}|\Omega_t)}{\tilde{g}_{t+1}(\boldsymbol{\xi}_{t+1}^{(i)}|\Lambda_t^{(i)},\Omega_{t+1})g_t(\Lambda_t^{(i)}|\Omega_t)} \\
&= \frac{p(\mathbf{y}_{t+1}|\boldsymbol{\xi}_{t+1}^{(i)})p(\boldsymbol{\xi}_{t+1}^{(i)}|\boldsymbol{\xi}_t^{(i)})}{\tilde{g}_{t+1}(\boldsymbol{\xi}_{t+1}^{(i)}|\Lambda_t^{(i)},\Omega_{t+1})} \frac{p_t(\Lambda_t^{(i)}|\Omega_t)}{g_t(\Lambda_t^{(i)}|\Omega_t)} \\
&= \tilde{\omega}_{t+1}^{(i)}\omega_t^{(i)}
\end{aligned}$$

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

$$\hat{E}(\boldsymbol{\xi}_{t-1}|\Omega_t) = \sum_{i=1}^{D} \hat{\omega}_t^{(i)} \boldsymbol{\xi}_{t-1}^{(i)}$$

$$\hat{P}(\xi_{1,t} > 0|\Omega_t) = \sum_{i=1}^{D} \hat{\omega}_t^{(i)} \delta_{[\xi_{1t}>0]}$$

$$\tilde{\omega}_{t+1}^{(i)} = \frac{p(\mathbf{y}_{t+1}|\boldsymbol{\xi}_{t+1}^{(i)})p(\boldsymbol{\xi}_{t+1}^{(i)}|\boldsymbol{\xi}_{t}^{(i)})}{\tilde{g}_{t+1}(\boldsymbol{\xi}_{t+1}^{(i)}|\Lambda_t^{(i)},\Omega_{t+1})}$$

$$\hat{p}(\mathbf{y}_{t+1}|\Omega_t) = \sum_{i=1}^{D} \tilde{\omega}_{t+1}^{(i)}\hat{\omega}_t^{(i)}$$

$$\hat{\mathcal{L}}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log \hat{p}(\mathbf{y}_t|\Omega_{t-1})$$

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

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

Flury and Shephard, Econometric Theory, 2011

How start algorithm for $t = 0$?

Draw $\boldsymbol{\xi}_0^{(i)}$ from $p(\boldsymbol{\xi}_0)$

(prior distribution or hypothesized unconditional distribution)

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

(1) Bootstrap filter

$\tilde{g}_{t+1}(\boldsymbol{\xi}_{t+1}|\Lambda_t,\Omega_{t+1}) = p(\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t)$

known from state equation

$\boldsymbol{\xi}_{t+1} = \boldsymbol{\phi}_t(\boldsymbol{\xi}_t,\mathbf{v}_{t+1})$

But better performance from adaptive filters that also use $\mathbf{y}_{t+1}$

Note that for bootstrap filter

$$\tilde{g}_{t+1}(\boldsymbol{\xi}_{t+1}|\Lambda_t,\Omega_{t+1}) = p(\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t)$$

$$\tilde{\omega}_{t+1}^{(i)} = \frac{p(\mathbf{y}_{t+1}|\boldsymbol{\xi}_{t+1}^{(i)})p(\boldsymbol{\xi}_{t+1}^{(i)}|\boldsymbol{\xi}_t^{(i)})}{\tilde{g}_{t+1}(\boldsymbol{\xi}_{t+1}^{(i)}|\Lambda_t^{(i)},\Omega_{t+1})}$$

$$= p(\mathbf{y}_{t+1}|\boldsymbol{\xi}_{t+1}^{(i)})$$

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

Partial solution to degeneracy problem:
Sequential Importance Sampling
with Resampling
Before finishing step t, now resample $\{\Lambda_t^{(j)}\}_{j=1}^{D}$ with replacement
by drawing from the distribution

$$\Lambda_t^{(j)} = \begin{cases} \Lambda_t^{(1)} & \text{with probability } \hat{\omega}_t^{(1)} \\ \vdots & \\ \Lambda_t^{(D)} & \text{with probability } \hat{\omega}_t^{(D)} \end{cases}$$

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 $\{\Lambda_t^{(i)}\}$ and weights $\{\hat{\omega}_t^{(i)}\}$ rather than $\{\Lambda_t^{(j)}\}$ and weights $\{1/D\}$

(6) Better not to resample every t

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

(1) Bootstrap filter:

$\tilde{g}_{t+1}(\boldsymbol{\xi}_{t+1}|\Lambda_t, \Omega_{t+1}) = p(\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t)$

(2) Auxiliary particle filter:

use $\mathbf{y}_{t+1}$ to get better proposal density for $\boldsymbol{\xi}_{t+1}$

Example: from state equation

$\boldsymbol{\xi}_{t+1} = \boldsymbol{\phi}_t(\boldsymbol{\xi}_t, \mathbf{v}_{t+1})$

we have guess for likely value for $\boldsymbol{\xi}_{t+1}$ associated with particle i,

e.g. $\hat{\boldsymbol{\xi}}_{t+1}^{(i)} = \boldsymbol{\phi}_t(\boldsymbol{\xi}_t^{(i)}, \mathbf{0})$

And from observation equation we can calculate how likely it would be to observe $\mathbf{y}_{t+1}$ if ξ_{t+1} took on this value:

$$p(\mathbf{y}_{t+1}|\hat{\boldsymbol{\xi}}_{t+1}^{(i)}) = p[\mathbf{y}_{t+1}|\boldsymbol{\phi}_t(\boldsymbol{\xi}_t^{(i)}, \mathbf{0})]$$
$$= \tilde{\tau}(\boldsymbol{\xi}_t^{(i)}, \mathbf{y}_{t+1})$$

Calculate

$$\tau_t^{(i)} = \tilde{\tau}_t^{(i)} \omega_t^{(i)}$$
$$\tilde{\tau}_t^{(i)} = \tilde{\tau}(\boldsymbol{\xi}_t^{(i)}, \mathbf{y}_{t+1})$$
$$\hat{\tau}_t^{(i)} = \frac{\tau_t^{(i)}}{\sum_{i=1}^{D} \tau_t^{(i)}}$$

Resample historical particles with prob $\hat{\tau}_t^{(i)}$

$$\Lambda_t^{(j)} = \begin{cases} \Lambda_t^{(1)} & \text{with probability } \hat{\tau}_t^{(1)} \\ \vdots & \\ \Lambda_t^{(D)} & \text{with probability } \hat{\tau}_t^{(D)} \end{cases}$$

Draw $\xi_{t+1}^{(j)}$ from proposal density $\tilde{g}_{t+1}(\xi_{t+1}|\Lambda_t^{(j)}, \Omega_{t+1}) = p(\xi_{t+1}|\xi_t^{(j)})$
From what importance density did we generate proposed $\Lambda_{t+1}^{(j)}$?

If we had resampled using original weights propotional to $\omega_t^{(i)}$, then $\Lambda_t^{(i)}$ 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 $\tilde{\tau}(\xi_t^{(i)}, \mathbf{y}_{t+1})\omega_t^{(i)}$, $\Lambda_t^{(j)}$ represents an i.i.d. sample with density proportional to $\tilde{\tau}(\Lambda_t, \mathbf{y}_{t+1})p_t(\Lambda_t|\Omega_t)$.

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

Target density is

$$p_{t+1}(\Lambda_{t+1}|\Omega_{t+1}) \propto p(\mathbf{y}_{t+1}|\boldsymbol{\xi}_{t+1})p(\boldsymbol{\xi}_{t+1}|\boldsymbol{\xi}_t)p_t(\Lambda_t|\Omega_t)$$

Desired weights are thus proportional to

$$\omega_{t+1}^{(j)} = \frac{p(\mathbf{y}_{t+1}|\boldsymbol{\xi}_{t+1}^{(j)})p(\boldsymbol{\xi}_{t+1}^{(j)}|\boldsymbol{\xi}_t^{(j)})p_t(\Lambda_t^{(j)}|\Omega_t)}{p(\boldsymbol{\xi}_{t+1}^{(j)}|\boldsymbol{\xi}_t^{(j)})\tilde{\tau}(\Lambda_t^{(j)},\mathbf{y}_{t+1})p_t(\Lambda_t^{(j)}|\Omega_t)} = \frac{p(\mathbf{y}_{t+1}|\boldsymbol{\xi}_{t+1}^{(j)})}{\tilde{\tau}(\Lambda_t^{(j)},\mathbf{y}_{t+1})}$$

Summary of auxiliary particle filter:

(1) Calculate measure of how useful $\Lambda_t^{(i)}$ is for predicting $\mathbf{y}_{t+1}$, e.g.

$$\tilde{\tau}_t^{(i)} = p[\mathbf{y}_{t+1}|\boldsymbol{\xi}_{t+1} = \boldsymbol{\phi}_t(\boldsymbol{\xi}_t^{(i)}, \mathbf{0})]$$

where $p(\mathbf{y}_{t+1}|\boldsymbol{\xi}_{t+1})$ comes from obs eq and $\boldsymbol{\xi}_{t+1} = \boldsymbol{\phi}_t(\boldsymbol{\xi}_t, \mathbf{v}_{t+1})$ is state eq

(2) Resample $\Lambda_t^{(j)}$ from $\Lambda_t^{(i)}$ with probabilities proportional to $\tilde{\tau}_t^{(i)}\omega_t^{(i)}$

(3) Generate $\boldsymbol{\xi}_{t+1}^{(j)}$ from $\boldsymbol{\phi}_t(\boldsymbol{\xi}_t^{(j)}, \mathbf{v}_{t+1})$

(4) Calculate weights

$$\omega_{t+1}^{(j)} = \frac{p(\mathbf{y}_{t+1}|\boldsymbol{\xi}_{t+1}^{(j)})}{\tilde{\tau}_t^{(j)}}$$

$$\omega_{t+1}^{(j)} = \frac{p(\mathbf{y}_{t+1}|\boldsymbol{\xi}_{t+1}^{(j)})}{\tilde{\tau}_t^{(j)}}$$

$$\hat{p}(\mathbf{y}_{t+1}|\Omega_t) = \sum_{i=1}^{D} \omega_{t+1}^{(j)}\hat{\omega}_t^{(j)}$$

$$\hat{\omega}_{t+1}^{(j)} = \frac{\omega_{t+1}^{(j)}}{\sum_{j=1}^{D} \omega_{t+1}^{(j)}}$$

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

$$\hat{\mathcal{L}}(\boldsymbol{\theta}) = \sum_{t=0}^{T-1} \log \hat{p}(\mathbf{y}_{t+1}|\Omega_t)$$

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:

$$\max_{\{c_t,k_{t+1}\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t \log c_t$$

s.t. $c_t + k_{t+1} = e^{z_t} k_t^{\alpha} + (1-\delta)k_t \quad t = 1,2,\ldots$

$z_t = \rho z_{t-1} + \sigma \varepsilon_t \quad t = 1,2,\ldots$

k_0, z_0 given

$\varepsilon_t \sim N(0,1)$

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

We seek decision rules of the form

$c_t = c(k_t, z_t; \sigma)$

$k_{t+1} = k(k_t, z_t; \sigma)$

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

$$a_1(k_t, z_t; \sigma, \varepsilon_{t+1}) = \frac{1}{c(k_t,z_t;\sigma)} - \beta \frac{\alpha k(k_t,z_t;\sigma)^{\alpha-1} \exp(\rho z_t + \sigma \varepsilon_{t+1})}{c(k(k_t,z_t;\sigma), \rho z_t + \sigma \varepsilon_{t+1}; \sigma)}$$

$$a_2(k_t, z_t; \sigma, \varepsilon_{t+1}) = c(k_t, z_t; \sigma) + k(k_t, z_t; \sigma) - e^{z_t} k_t^{\alpha} - (1-\delta) k_t$$

Zero-order approximation (deterministic steady state)

$\sigma = 0$

$z_t = z = 0$

$k_t = k$

$\mathbf{a}(k, 0; 0) = \mathbf{0}$

$a_1(k, 0; 0) = 0$

$\Rightarrow \frac{1}{c} - \beta \frac{\alpha k^{\alpha-1}}{c} = 0$

$\Rightarrow 1 = \beta \alpha k^{\alpha-1}$

$a_2(k, 0; 0) = 0$

$\Rightarrow c + k - \ k^{\alpha} \ - (1-\delta)k$

$\Rightarrow c = k^{\alpha} - \delta k$

First-order approximation:

Since $E_t \mathbf{a}(k_t, z_t; \sigma, \varepsilon_{t+1}) = \mathbf{0}$ for all $k_t, z_t; \sigma$, it follows that

$E_t \mathbf{a}_k(k_t, z_t; \sigma, \varepsilon_{t+1}) = \mathbf{0}$

for $\mathbf{a}_k(k_t, z_t; \sigma, \varepsilon_{t+1}) = \frac{\partial \mathbf{a}(k_t, z_t; \sigma, \varepsilon_{t+1})}{\partial k_t}$

likewise

$E_t \mathbf{a}_z(k_t, z_t; \sigma, \varepsilon_{t+1}) = E_t \mathbf{a}_\sigma(k_t, z_t; \sigma, \varepsilon_{t+1}) = \mathbf{0}$

$$E_t\left\{\frac{\partial a_1(k_t,z_t;\sigma,\varepsilon_{t+1})}{\partial k_t}\Big|_{k_t=k,z_t=0,\sigma=0}\right\} =$$
$$\frac{-1}{c^2}c_k - \frac{\beta\alpha(\alpha-1)k^{\alpha-2}}{c}k_k + \frac{\beta\alpha k^{\alpha-1}}{c^2}c_k k_k$$

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

$$c_k = \frac{\partial c(k_t,z_t;\sigma)}{\partial k_t}\Big|_{k_t=k,z_t=0,\sigma=0}$$

$$\frac{\partial a_2(k_t,z_t;\sigma)}{\partial k_t}\Big|_{k_t=k,z_t=0,\sigma=0} =$$
$$c_k + k_k - \alpha k^{\alpha-1} - (1-\delta)$$

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

$$E_t\left\{\frac{\partial a_1(k_t,z_t;\sigma,\varepsilon_{t+1})}{\partial z_t}\Big|_{k_t=k,z_t=0,\sigma=0}\right\} =$$
$$\frac{-1}{c^2}c_z - \frac{\beta\alpha(\alpha-1)k^{\alpha-2}}{c}k_z - \frac{\beta\alpha k^{\alpha-1}\rho}{c}$$
$$+ \frac{\beta\alpha k^{\alpha-1}}{c^2}(c_k k_z + \rho c_z)$$
$$\frac{\partial a_2(k_t,z_t;\sigma)}{\partial z_t}\Big|_{k_t=k,z_t=0,\sigma=0} =$$
$$c_z + k_z - k^\alpha$$

setting these to zero allows us to solve for c_z, k_z

$$\frac{\partial a_1(k_t,z_t;\sigma,\varepsilon_{t+1})}{\partial\sigma}\Bigg|_{k_t=k,z_t=0,\sigma=0} =$$

$$\frac{-1}{c^2}c_\sigma - \frac{\beta\alpha(\alpha-1)k^{\alpha-2}}{c}k_\sigma - \frac{\beta\alpha k^{\alpha-1}\varepsilon_{t+1}}{c}$$

$$+ \frac{\beta\alpha k^{\alpha-1}}{c^2}(c_k k_\sigma + \varepsilon_{t+1}c_z + c_\sigma)$$

$$\frac{\partial a_2(k_t,z_t;\sigma)}{\partial\sigma}\Bigg|_{k_t=k,z_t=0,\sigma=0} =$$

$$c_\sigma + k_\sigma$$

Taking expectations and setting to zero yields

$$\frac{-1}{c^2}c_\sigma - \frac{\beta\alpha(\alpha-1)k^{\alpha-2}}{c}k_\sigma + \frac{\beta\alpha k^{\alpha-1}}{c^2}(c_k k_\sigma + c_\sigma) = 0$$

$$c_\sigma + k_\sigma = 0$$

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

$$c(k_t,z_t;\sigma) \simeq c + c_k(k_t - k) + c_z z_t + c_\sigma\sigma$$

$$k(k_t,z_t;\sigma) \simeq k + k_k(k_t - k) + k_z z_t + k_\sigma\sigma$$

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 $E_t \mathbf{a}(k_t, z_t; \sigma, \varepsilon_{t+1}) = \mathbf{0}$ 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

$E_t \mathbf{a}_k(k_t, z_t; \sigma, \varepsilon_{t+1}) = \mathbf{0}$

$E_t \mathbf{a}_z(k_t, z_t; \sigma, \varepsilon_{t+1}) = \mathbf{0}$

$E_t \mathbf{a}_\sigma(k_t, z_t; \sigma, \varepsilon_{t+1}) = \mathbf{0}$

with respect to $k_t, z_t,$ and σ.

Gives 18 linear equations in the 12 unknowns $\{c_{ij}, k_{ij}\}_{i,j \in \{k,z,\sigma\}}$ 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,

$$c(k_t, z_t; \sigma) \simeq c + \mathbf{c}_1' \mathbf{s}_t + (1/2)\mathbf{s}_t' \mathbf{C}_2 \mathbf{s}_t$$

$$k(k_t, z_t; \sigma) \simeq k + \mathbf{k}_1' \mathbf{s}_t + (1/2)\mathbf{s}_t' \mathbf{K}_2 \mathbf{s}_t$$

$$\mathbf{c}_1' = \begin{bmatrix} c_k & c_z & 0 \end{bmatrix}$$

$$\mathbf{s}_t = \begin{bmatrix} (k_t - k) & z_t & \sigma \end{bmatrix}$$

$$\mathbf{k}_1' = \begin{bmatrix} k_k & k_z & 0 \end{bmatrix}$$

$$\mathbf{C}_2 = \begin{bmatrix} c_{kk} & c_{kz} & c_{k\sigma} \\ c_{zk} & c_{zz} & c_{z\sigma} \\ c_{\sigma k} & c_{\sigma z} & c_{\sigma\sigma} \end{bmatrix}$$

$$\mathbf{K}_2 = \begin{bmatrix} k_{kk} & k_{kz} & k_{k\sigma} \\ k_{zk} & k_{zz} & k_{z\sigma} \\ k_{\sigma k} & k_{\sigma z} & k_{\sigma\sigma} \end{bmatrix}$$

$$c(k_t, z_t; \sigma) \simeq c + \mathbf{c}_1' \mathbf{s}_t + (1/2)\mathbf{s}_t' \mathbf{C}_2 \mathbf{s}_t$$

$$\mathbf{s}_t = \begin{bmatrix} (k_t - k) & z_t & \sigma \end{bmatrix}$$

Note: term on σ^2 in $\mathbf{s}_t' \mathbf{C}_2 \mathbf{s}_t$ 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

$$
\begin{aligned}
C_t + I_t &= A_t K_t^{\alpha} L_t^{1-\alpha} \\
K_{t+1} &= (1-\delta)K_t + U_t I_t \\
\log A_t &= \zeta + \log A_{t-1} + \sigma_{at}\varepsilon_{at} \\
\log U_t &= \theta + \log U_{t-1} + \sigma_{vt}\varepsilon_{vt} \\
\log \sigma_{at} &= (1-\lambda_a)\log \overline{\sigma}_a \\
&\quad + \lambda_a \log \sigma_{a,t-1} + \tau_a \eta_{at} \\
\log \sigma_{vt} &= (1-\lambda_v)\log \overline{\sigma}_v \\
&\quad + \lambda_v \log \sigma_{v,t-1} + \tau_v \eta_{vt}
\end{aligned}
$$

$$
E_0 \sum_{t=0}^{\infty} \beta^t \{ e^{d_t} \log C_t + \psi \log(1-L_t) \}
$$

$$
\begin{aligned}
d_t &= \rho d_{t-1} + \sigma_{dt}\varepsilon_{dt} \\
\log \sigma_{dt} &= (1-\lambda_d)\log \overline{\sigma}_d \\
&\quad + \lambda_d \log \sigma_{d,t-1} + \tau_d \eta_{dt}
\end{aligned}
$$

$\mathbf{v}_t = (\varepsilon_{at}, \varepsilon_{vt}, \varepsilon_{dt}, \eta_{at}, \eta_{vt}, \eta_{dt})'$

$\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{I}_6)$

$\boldsymbol{\Omega} = \text{diag}\{\overline{\sigma}_a^2, \overline{\sigma}_v^2, \overline{\sigma}_d^2, \tau_a^2, \tau_v^2, \tau_d^2\}$

perturbation method: Continuum of economies with variance $\chi\boldsymbol{\Omega}$, take expansion around $\chi = 0$

Transformations to find steady-state representation:

$Z_t = A_{t-1}^{1/(1-\alpha)} U_{t-1}^{\alpha/(1-\alpha)}$

$\tilde{Y}_t = Y_t/Z_t,\ \tilde{C}_t = C_t/Z_t,\ \tilde{I}_t = I_t/Z_t$

$\tilde{U}_t = U_t/U_{t-1},\ \tilde{A}_t = A_t/A_{t-1},\ \tilde{K}_t = K_t/Z_t U_{t-1}$

$\tilde{k}$ = log of steady-state value for $\tilde{K}$

$\hat{\tilde{k}}_t = \log \tilde{K}_t - \tilde{k}$

state vector for economic model:

$\tilde{\mathbf{s}}_t = (\hat{\tilde{k}}_t, \varepsilon_{at}, \varepsilon_{vt}, \varepsilon_{dt}, d_{t-1},$
$\sigma_{at} - \overline{\sigma}_a, \sigma_{vt} - \overline{\sigma}_v, \sigma_{dt} - \overline{\sigma}_d)'$

second-order perturbation:

$\hat{\tilde{k}}_{t+1} = \boldsymbol{\psi}_{k1}' \tilde{\mathbf{s}}_t + (1/2)\tilde{\mathbf{s}}_t' \boldsymbol{\Psi}_{k2} \tilde{\mathbf{s}}_t + \psi_{k0}$

$\hat{\tilde{i}}_t = \boldsymbol{\psi}_{i1}' \tilde{\mathbf{s}}_t + (1/2)\tilde{\mathbf{s}}_t' \boldsymbol{\Psi}_{i2} \tilde{\mathbf{s}}_t + \psi_{i0}$

$\hat{\tilde{\ell}}_t = \boldsymbol{\psi}_{\ell 1}' \tilde{\mathbf{s}}_t + (1/2)\tilde{\mathbf{s}}_t' \boldsymbol{\Psi}_{\ell 2} \tilde{\mathbf{s}}_t + \psi_{\ell 0}$

ψ_{j0} reflects precautionary effects

However, we will observe actual GDP growth per capita

$$\Delta \log Y_t = \Delta \log \tilde{Y}_t + \tfrac{1}{1-\alpha}(\Delta \log A_{t-1} + \alpha \Delta \log U_{t-1}) + \sigma_{y\varepsilon}\varepsilon_{yt}$$
$$= h_y(\tilde{\mathbf{s}}_t, \tilde{\mathbf{s}}_{t-1}) + \sigma_{y\varepsilon}\varepsilon_{yt}$$

ε_{yt} = 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

$$\Delta \log I_t = h_i(\tilde{\mathbf{s}}_t, \tilde{\mathbf{s}}_{t-1}) + \sigma_{i\varepsilon}\varepsilon_{it}$$
$$\log \ell_t = h_\ell(\tilde{\mathbf{s}}_t, \tilde{\mathbf{s}}_{t-1}) + \sigma_{\ell\varepsilon}\varepsilon_{\ell t}$$
$$\Delta \log P_t = -\Delta \log U_t$$

$$\tilde{\mathbf{s}}_t = (\hat{\tilde{k}}_t, \varepsilon_{at}, \varepsilon_{vt}, \varepsilon_{dt}, d_{t-1}, \sigma_{at} - \overline{\sigma}_a, \sigma_{vt} - \overline{\sigma}_v, \sigma_{dt} - \overline{\sigma}_d)'$$
$$\mathbf{v}_t = (\varepsilon_{at}, \varepsilon_{vt}, \varepsilon_{dt}, \eta_{at}, \eta_{vt}, \eta_{dt})'$$
$$\mathbf{S}_t = (\tilde{\mathbf{s}}_t', \tilde{\mathbf{s}}_{t-1}')$$

state equation

$$\mathbf{S}_t = \mathbf{f}(\mathbf{S}_{t-1}, \mathbf{v}_t)$$
$$f_1(\mathbf{S}_{t-1}, \mathbf{v}_t) = \boldsymbol{\psi}_{k1}'\tilde{\mathbf{s}}_t + (1/2)\tilde{\mathbf{s}}_t'\boldsymbol{\Psi}_{k2}\tilde{\mathbf{s}}_t + \psi_{k0}$$

$$f_2(\mathbf{S}_{t-1}, \mathbf{v}_t) = \varepsilon_{at}$$

$$\vdots$$

$$f_5(\mathbf{S}_{t-1}, \mathbf{v}_t) = \rho d_{t-2} + \sigma_{d,t-1}\varepsilon_{d,t-1}$$

$$f_6(\mathbf{S}_{t-1}, \mathbf{v}_t) = \exp[(1-\lambda_a)\log \overline{\sigma}_a + \lambda_a \log \sigma_{a,t-1} + \tau_a \eta_{at}] - \overline{\sigma}_a$$

$$\vdots$$

$$\mathbf{f}_{9-16}(\mathbf{S}_{t-1}, \mathbf{v}_t) = \tilde{\mathbf{s}}_{t-1}$$

$$\mathbf{y}_t = (\Delta \log Y_t, \Delta \log I_t, \log \ell_t, \Delta \log P_t)'$$

observation equation:

$$\mathbf{y}_t = \mathbf{h}(\mathbf{S}_t) + \mathbf{w}_t$$

According to the set-up, ε_{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
$\mathbf{v}_{1t} = (\varepsilon_{at}, \varepsilon_{dt}, \eta_{at}, \eta_{vt}, \eta_{dt})'$
in order to have a value for σ_{vt} and value for ε_{vt}
$\varepsilon_{vt} = -\frac{\Delta \log P_t + \theta}{\sigma_{vt}}$

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 $N(\mathbf{0}, \mathbf{I}_6)$ 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 $\mathbf{v}_{1,t+1}^{(i)} \sim N(\mathbf{0}, \mathbf{I}_5)$ for $i = 1, \ldots, D$

Step 2: generate $\mathbf{S}_{t+1}^{(i)} = \mathbf{f}\left(\mathbf{S}_t^{(i)}, \mathbf{v}_{t+1}^{(i)}\right)$ except for the third element $\varepsilon_{v,t+1}^{(i)}$

Step 3: calculate

$\mathbf{w}_{t+1}^{(i)} = \mathbf{y}_{t+1} - \mathbf{h}(\mathbf{S}_{t+1}^{(i)})$

and set third element of $\mathbf{S}_{t+1}^{(i)}$ equal to fourth element of $\mathbf{w}_{t+1}^{(i)}$, $\varepsilon_{v,t+1}^{(i)} = -\frac{\Delta \log P_{t+1} + \theta}{\sigma_{v,t+1}^{(i)}}$

Step 4: calculate

$$\tilde{\omega}_{t+1}^{(i)} = (2\pi)^{-4/2} |\mathbf{D}_{t+1}^{(i)}| \times \exp\left(-(1/2)\left[\mathbf{w}_{t+1}^{(i)}\right][\mathbf{D}_{t+1}^{(i)}]^{-1}\left[\mathbf{w}_{t+1}^{(i)}\right]\right)$$

$$\mathbf{D}_{t+1}^{(i)} = \begin{bmatrix} \sigma_{y\varepsilon}^2 & 0 & 0 & 0 \\ 0 & \sigma_{i\varepsilon}^2 & 0 & 0 \\ 0 & 0 & \sigma_{\ell\varepsilon}^2 & 0 \\ 0 & 0 & 0 & \left[\sigma_{v,t+1}^{(i)}\right]^2 \end{bmatrix}$$

Step 5: Contribution to likelihood is

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

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

$$\Lambda_{t+1}^{(j)} = \begin{cases} \Lambda_{t+1}^{(1)} & \text{with probability } \hat{\omega}_{t+1}^{(1)} \\ \vdots & \\ \Lambda_{t+1}^{(D)} & \text{with probability } \hat{\omega}_{t+1}^{(D)} \end{cases}$$

Structural parameters:

$$\boldsymbol{\theta} = (\alpha, \delta, \rho, \beta, \psi, \theta, \zeta, \tau_a, \tau_v, \tau_d, \overline{\sigma}_a, \overline{\sigma}_v, \overline{\sigma}_d, \lambda_a, \lambda_v, \lambda_d, \sigma_{y\varepsilon}, \sigma_{i\varepsilon}, \sigma_{\ell\varepsilon})'$$

Fernandez-Villaverde and Rubio-Ramirez estimate $\boldsymbol{\theta}$ by maximizing

$$\hat{\mathcal{L}}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \hat{p}(\mathbf{y}_t | \Omega_{t-1}; \boldsymbol{\theta})$$