State Space Models and Filtering

Jesús Fernández-Villaverde
University of Pennsylvania
State Space Form

- What is a state space representation?
- States versus observables.
- Why is it useful?
- Relation with filtering.
- Relation with optimal control.
- Linear versus nonlinear, Gaussian versus nongaussian.
State Space Representation

• Let the following system:
  
  – Transition equation
  \[ x_{t+1} = Fx_t + G\omega_{t+1}, \quad \omega_{t+1} \sim \mathcal{N}(0, Q) \]

  – Measurement equation
  \[ z_t = H'x_t + \nu_t, \quad \nu_t \sim \mathcal{N}(0, R) \]

  – where \( x_t \) are the states and \( z_t \) are the observables.

• Assume we want to write the likelihood function of \( z^T = \{z_t\}_{t=1}^T \).
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 $x_t^* = Bx_t$, $F^* = BFB^{-1}$, $G^* = BG$, and $H^* = (H'B)'$, we can write a new, equivalent, representation:

  - Transition equation
    \[
    x_{t+1}^* = F^* x_t^* + G^* \omega_{t+1}, \quad \omega_{t+1} \sim \mathcal{N}(0, Q)
    \]
  
  - Measurement equation
    \[
    z_t = H^* x_t^* + v_t, \quad v_t \sim \mathcal{N}(0, R)
    \]
Example 1

- Assume the following AR(2) process:

\[ z_t = \rho_1 z_{t-1} + \rho_2 z_{t-2} + \nu_t, \quad \nu_t \sim \mathcal{N}(0, \sigma^2) \]

- Model is not apparently not Markovian.

- Can we write this model in different state space forms?

- Yes!
State Space Representation I

- Transition equation:

\[ x_t = \begin{bmatrix} \rho_1 & 1 \\ \rho_2 & 0 \end{bmatrix} x_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} v_t \]

where \( x_t = \begin{bmatrix} y_t & \rho_2 y_{t-1} \end{bmatrix}' \)

- Measurement equation:

\[ z_t = \begin{bmatrix} 1 & 0 \end{bmatrix} x_t \]
State Space Representation II

- Transition equation:

\[
x_t = \begin{bmatrix} \rho_1 & \rho_2 \\ 1 & 0 \end{bmatrix} x_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \nu_t
\]

where \( x_t = \begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} \).

- Measurement equation:

\[
z_t = \begin{bmatrix} 1 & 0 \end{bmatrix} x_t
\]

- Try \( B = \begin{bmatrix} 1 & 0 \\ 0 & \rho_2 \end{bmatrix} \) on the second system to get the first system.
Example II

- Assume the following MA(1) process:

\[ z_t = v_t + \theta v_{t-1}, \quad v_t \sim \mathcal{N}(0, \sigma_v^2) , \quad \text{and} \quad E v_t v_s = 0 \text{ for } s \neq t. \]

- Again, we have a more complicated structure than a simple Markovian process.

- However, it will again be straightforward to write a state space representation.
State Space Representation 1

- Transition equation:

\[ x_t = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x_{t-1} + \begin{bmatrix} 1 \\ \theta \end{bmatrix} v_t \]

where \( x_t = \begin{bmatrix} y_t \\ \theta v_t \end{bmatrix} \)

- Measurement equation:

\[ z_t = \begin{bmatrix} 1 & 0 \end{bmatrix} x_t \]
State Space Representation II

- Transition equation:

\[ x_t = v_{t-1} \]

- where \( x_t = [v_{t-1}]' \)

- Measurement equation:

\[ z_t = \theta x_t + v_t \]

- Again both representations are equivalent!
Example III

- Assume the following random walk plus drift process:
  \[ z_t = z_{t-1} + \beta + v_t, \quad v_t \sim \mathcal{N}(0, \sigma_v^2) \]

- This is even more interesting.

- We have a unit root.

- We have a constant parameter (the drift).
State Space Representation

- Transition equation:

\[ x_t = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} v_t \]

where \( x_t = \begin{bmatrix} y_t & \beta \end{bmatrix}' \)

- Measurement equation:

\[ z_t = \begin{bmatrix} 1 & 0 \end{bmatrix} x_t \]
Some Conditions on the State Space Representation

• We only consider Stable Systems.

• A system is stable if for any initial state $x_0$, the vector of states, $x_t$, converges to some unique $x^*$.

• 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.

- Why the name “filter”? 
Some Definitions

- Let $x_t|t-1 = E(x_t|z^{t-1})$ be the best linear predictor of $x_t$ given the history of observables until $t - 1$, i.e. $z^{t-1}$.

- Let $z_t|t-1 = E(z_t|z^{t-1}) = H'x_t|t-1$ be the best linear predictor of $z_t$ given the history of observables until $t - 1$, i.e. $z^{t-1}$.

- Let $x_t|t = E(x_t|z^t)$ be the best linear predictor of $x_t$ given the history of observables until $t$, i.e. $z^t$. 
What is the Kalman Filter trying to do?

- Let assume we have $x_{t|t-1}$ and $z_{t|t-1}$.

- We observe a new $z_t$.

- We need to obtain $x_{t|t}$.

- Note that $x_{t+1|t} = Fx_{t|t}$ and $z_{t+1|t} = H'x_{t+1|t}$, so we can go back to the first step and wait for $z_{t+1}$.

- Therefore, the key question is how to obtain $x_{t|t}$ from $x_{t|t-1}$ and $z_t$. 

16
A Minimization Approach to the Kalman Filter I

- Assume we use the following equation to get $x_{t|t}$ from $z_t$ and $x_{t|t-1}$:
  \[
  x_{t|t} = x_{t|t-1} + K_t (z_t - z_{t|t-1}) = x_{t|t-1} + K_t (z_t - H' x_{t|t-1})
  \]

- This formula will have some probabilistic justification (to follow).

- What is $K_t$?
A Minimization Approach to the Kalman Filter II

- \( K_t \) is called the Kalman filter gain and it measures how much we update \( x_{t|t-1} \) as a function in our error in predicting \( z_t \).

- The question is how to find the optimal \( K_t \).

- The Kalman filter is about how to build \( K_t \) such that optimally update \( x_{t|t} \) from \( x_{t|t-1} \) and \( z_t \).

- How do we find the optimal \( K_t \)?
Some Additional Definitions

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

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

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

19
Finding the optimal $K_t$

- We want $K_t$ such that $\min \Sigma_t|t$.

- It can be shown that, if that is the case:

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

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

$$x_{t|t} = x_{t|t-1} + K_t \left( z_t - H' x_{t|t-1} \right)$$

- We will provide some intuition later.
Example 1

Assume the following model in State Space form:

• Transition equation

\[ x_t = \mu + \nu_t, \quad \nu_t \sim N\left(0, \sigma^2_\nu\right) \]

• Measurement equation

\[ z_t = x_t + \xi_t, \quad \xi_t \sim N\left(0, \sigma^2_\xi\right) \]

• Let \( \sigma^2_\xi = q\sigma^2_\nu \).
Example II

• Then, if $\Sigma_{1|0} = \sigma_v^2$, what it means that $x_1$ was drawn from the ergodic distribution of $x_t$.

• We have:

$$K_1 = \sigma_v^2 \frac{1}{1 + q} \propto \frac{1}{1 + q}.$$  

• Therefore, the bigger $\sigma_\xi^2$ relative to $\sigma_v^2$ (the bigger $q$) the lower $K_1$ and the less we trust $z_1$.  

22
The Kalman Filter Algorithm I

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

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

$$
\Omega_{t|t-1} \equiv E\left( (z_t - z_{t|t-1}) (z_t - z_{t|t-1})' | z^{t-1} \right)
\equiv E \left( H' (x_t - x_{t|t-1}) (x_t - x_{t|t-1})' H + v_t (x_t - x_{t|t-1})' H + H' (x_t - x_{t|t-1}) v_t' + v_t v_t' | z^{t-1} \right)
= H' \Sigma_{t|t-1} H + R
$$
The Kalman Filter Algorithm II

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

\[
E \left( \left( z_t - z_{t|t-1} \right) \left( x_t - x_{t|t-1} \right)' \mid z^{t-1} \right) =
\]
\[
E \left( H' \left( x_t - x_{t|t-1} \right) \left( x_t - x_{t|t-1} \right)' + v_t \left( x_t - x_{t|t-1} \right)' \mid z^{t-1} \right) = H' \Sigma_{t|t-1}
\]

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 $\Sigma_{t|t-1}$, $x_{t|t-1}$, $K_t$, and $z_t$ then we compute:

\[
x_{t|t} = x_{t|t-1} + K_t \left( z_t - H' x_{t|t-1} \right)
\]
The Kalman Filter Algorithm III

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

$$
\Sigma_{t|t} \equiv E \left( (x_t - x_{t|t}) (x_t - x_{t|t})' | z^t \right) =
$$

$$
E \left( \begin{pmatrix}
(x_t - x_{t|t-1}) (x_t - x_{t|t-1})' - \\
(x_t - x_{t|t-1}) (z_t - H' x_{t|t-1})' K'_t - \\
K_t (z_t - H' x_{t|t-1}) (x_t - x_{t|t-1})' + \\
K_t (z_t - H' x_{t|t-1}) (z_t - H' x_{t|t-1})' K'_t z^t
\end{pmatrix} \right) = \Sigma_{t|t-1} - K_t H' \Sigma_{t|t-1}
$$

where, you have to notice that $x_t - x_{t|t} = x_t - x_{t|t-1} - K_t \left( z_t - H' x_{t|t-1} \right)$. 

25
The Kalman Filter Algorithm IV

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

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

Let $x_{t|t}$, then we compute:

1. $x_{t+1|t} = Fx_{t|t}$

2. $z_{t+1|t} = H'x_{t+1|t}$

Therefore, from $x_{t|t-1}$, $\Sigma_{t|t-1}$, and $z_t$ we compute $x_{t|t}$ and $\Sigma_{t|t}$. 
The Kalman Filter Algorithm V

We also compute $z_{t|t-1}$ and $\Omega_{t|t-1}$.

Why?

To calculate the likelihood function of $z^T = \{z_t\}_{t=1}^T$ (to follow).
The Kalman Filter Algorithm: A Review

We start with $x_{t|t-1}$ and $\Sigma_{t|t-1}$.

Then, we observe $z_t$ and:

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

- $z_{t|t-1} = H'x_{t|t-1}$

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

- $\Sigma_{t|t} = \Sigma_{t|t-1} - K_tH'\Sigma_{t|t-1}$
• $x_{t|t} = x_{t|t-1} + K_t \left( z_t - H' x_{t|t-1} \right)$

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

• $x_{t+1|t} = F x_{t|t}$

We finish with $x_{t+1|t}$ and $\Sigma_{t+1|t}$. 
Some Intuition about the optimal $K_t$

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

- Notice that we can rewrite $K_t$ in the following way:

$$K_t = \Sigma_{t|t-1} H \Omega_{t|t-1}^{-1}$$

- If we did a big mistake forecasting $x_{t|t-1}$ using past information ($\Sigma_{t|t-1}$ large) we give a lot of weight to the new information ($K_t$ large).

- If the new information is noise ($R$ large) we give a lot of weight to the old prediction ($K_t$ small).
A Probabilistic Approach to the Kalman Filter

- Assume:

\[ Z|w = [X'|w Y'|w]' \sim N\left( \begin{bmatrix} x^* \\ y^* \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \right) \]

- then:

\[ X|y, w \sim N\left( x^* + \Sigma_{xy} \Sigma_{yy}^{-1} (y - y^*), \Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \right) \]

- Also \( x_{t|t-1} \equiv E\left(x_t|z^{t-1}\right) \) and:

\[ \Sigma_{t|t-1} \equiv E\left( (x_t - x_{t|t-1}) (x_t - x_{t|t-1})' | z^{t-1} \right) \]
Some Derivations I

If $z_t|z^{t-1}$ is the random variable $z_t$ (observable) conditional on $z^{t-1}$, then:

- Let $z_t|t-1 \equiv E(z_t|z^{t-1}) = E(H'x_t + \nu_t|z^{t-1}) = H'x_{t|t-1}$

- Let

$$
\Omega_{t|t-1} \equiv E\left(\left(z_t - z_{t|t-1}\right)\left(z_t - z_{t|t-1}\right)' \Big| z^{t-1}\right) = \\
E\left( \\
H' \left( x_t - x_{t|t-1} \right) \left( x_t - x_{t|t-1} \right)' H + \\
v_t \left( x_t - x_{t|t-1} \right)' H + \\
H' \left( x_t - x_{t|t-1} \right) v_t' + \\
v_t v_t'|z^{t-1} \right) = H' \Sigma_{t|t-1} H + R
$$
Some Derivations II

Finally, let

\[ E \left( \left( z_t - z_{t|t-1} \right) \left( x_t - x_{t|t-1} \right)' \bigg| z^{t-1} \right) = \]

\[ E \left( H' \left( x_t - x_{t|t-1} \right) \left( x_t - x_{t|t-1} \right)' + v_t \left( x_t - x_{t|t-1} \right)' \bigg| z^{t-1} \right) = \]

\[ = H' \Sigma_{t|t-1} \]
The Kalman Filter First Iteration I

- Assume we know $x_{1|0}$ and $\Sigma_{1|0}$, then

$$
\begin{pmatrix}
x_1 \\
z_1
\end{pmatrix}
\mid z^0
\sim
\mathcal{N}
\left(
\begin{bmatrix}
x_{1|0} \\
H' x_{1|0}
\end{bmatrix},
\begin{bmatrix}
\Sigma_{1|0} & \Sigma_{1|0} H \\
H' \Sigma_{1|0} & H' \Sigma_{1|0} H + R
\end{bmatrix}
\right)
$$

- Remember that:

$$X \mid y, w \sim \mathcal{N}
\left(x^* + \Sigma_{xy} \Sigma_y^{-1} (y - y^*), \Sigma_{xx} - \Sigma_{xy} \Sigma_y^{-1} \Sigma_{yx}\right)$$
The Kalman Filter First Iteration II

Then, we can write:

\[ x_{1|z_1}, z^0 = x_{1|z}^1 \sim N \left( x_{1|1}, \Sigma_{1|1} \right) \]

where

\[ x_{1|1} = x_{1|0} + \Sigma_{1|0} H \left( H' \Sigma_{1|0} H + R \right)^{-1} \left( z_1 - H' x_{1|0} \right) \]

and

\[ \Sigma_{1|1} = \Sigma_{1|0} - \Sigma_{1|0} H \left( H' \Sigma_{1|0} H + R \right)^{-1} H' \Sigma_{1|0} \]
• Therefore, we have that:

- \( z_{1|0} = H' x_{1|0} \)

- \( \Omega_{1|0} = H' \Sigma_{1|0} H + R \)

- \( x_{1|1} = x_{1|0} + \Sigma_{1|0} H \left( H' \Sigma_{1|0} H + R \right)^{-1} \left( z_{1} - H' x_{1|0} \right) \)

- \( \Sigma_{1|1} = \Sigma_{1|0} - \Sigma_{1|0} H \left( H' \Sigma_{1|0} H + R \right)^{-1} H' \Sigma_{1|0} \)

• Also, since \( x_{2|1} = F x_{1|1} + G \omega_{2|1} \) and \( z_{2|1} = H' x_{2|1} + \nu_{2|1} \):

- \( x_{2|1} = F x_{1|1} \)

- \( \Sigma_{2|1} = F \Sigma_{1|1} F' + G Q G' \)
The Kalman Filter $th$ Iteration I

• Assume we know $x_{t|t-1}$ and $\Sigma_{t|t-1}$, then

\[
\begin{pmatrix}
  x_t \\
  z_t \\
\end{pmatrix} \mid z^{t-1} \sim N \left( \begin{bmatrix}
  x_{t|t-1} \\
  H' x_{t|t-1} \\
\end{bmatrix} , \begin{bmatrix}
  \Sigma_{t|t-1} & \Sigma_{t|t-1} H \\
  H' \Sigma_{t|t-1} & H' \Sigma_{t|t-1} H + R \\
\end{bmatrix} \right)
\]

• Remember that:

\[
X | y, w \sim N \left( x^* + \Sigma_{xy} \Sigma_{yy}^{-1} (y - y^*) , \Sigma_{xx} - \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx} \right)
\]
The Kalman Filter \( n \)th Iteration II

Then, we can write:

\[
x_t \mid z_t, z^{t-1} = x_t \mid z^t \sim N \left( x_t \mid t, \Sigma_t \mid t \right)
\]

where

\[
x_t \mid t = x_t \mid t-1 + \Sigma_t \mid t-1 H \left( H' \Sigma_t \mid t-1 H + R \right)^{-1} \left( z_t - H' x_t \mid t-1 \right)
\]

and

\[
\Sigma_t \mid t = \Sigma_t \mid t-1 - \Sigma_t \mid t-1 H \left( H' \Sigma_t \mid t-1 H + R \right)^{-1} H' \Sigma_t \mid t-1
\]
The Kalman Filter Algorithm

Given $x_{t|t-1}$, $\Sigma_{t|t-1}$ and observation $z_t$

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

- $z_{t|t-1} = H'x_{t|t-1}$

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

- $x_{t|t} = x_{t|t-1} + \Sigma_{t|t-1}H\left(H'\Sigma_{t|t-1}H + R\right)^{-1}(z_t - H'x_{t|t-1})'$
\[ \Sigma_{t+1|t} = F \Sigma_{t|t} F' + G Q G_{t|t-1} \]

\[ x_{t+1|t} = F x_{t|t-1} \]
Putting the Minimization and the Probabilistic Approaches Together

- From the Minimization Approach we know that:

\[ x_{t|t} = x_{t|t-1} + K_t \left( z_t - H' x_{t|t-1} \right) \]

- From the Probability Approach we know that:

\[ x_{t|t} = x_{t|t-1} + \Sigma_t H^{-1} \left( H' \Sigma_t H + R \right)^{-1} \left( z_t - H' x_{t|t-1} \right) \]
• But since:

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

• We can also write in the probabilistic approach:

\[ x_{t|t} = x_{t|t-1} + \Sigma_{t|t-1} H \left( H' \Sigma_{t|t-1} H + R \right)^{-1} \left( z_t - H' x_{t|t-1} \right) = \]
\[ = x_{t|t-1} + K_t \left( z_t - H' x_{t|t-1} \right) \]

• Therefore, both approaches are equivalent.
Writing the Likelihood Function

We want to write the likelihood function of $z^T = \{z_t\}^T_{t=1}$:

$$\log \ell \left( z^T \mid F, G, H, Q, R \right) =$$

$$\sum_{t=1}^{T} \log \ell \left( z_t \mid z^{t-1} F, G, H, Q, R \right) =$$

$$-\sum_{t=1}^{T} \left[ \frac{N}{2} \log 2\pi + \frac{1}{2} \log |\Omega_t|_{t-1} | + \frac{1}{2} \sum_{t=1}^{T} v_t^T \Omega^{-1}_{t|t-1} v_t \right]$$

$$v_t = z_t - z_{t|t-1} = z_t - H' x_{t|t-1}$$

$$\Omega_{t|t-1} = H_t^' \Sigma_{t|t-1} H_t + R$$
Initial conditions for the Kalman Filter

• An important step in the Kalman Filter is to set the initial conditions.

• Initial conditions:

1. $x_{1|0}$

2. $\Sigma_{1|0}$

• Where do they come from?
Since we only consider stable system, the standard approach is to set:

- \( x_{1|0} = x^* \)

- \( \Sigma_{1|0} = \Sigma^* \)

where \( x \) solves

\[
\begin{align*}
x^* &= Fx^* \\
\Sigma^* &= F\Sigma^*F' + GQG'
\end{align*}
\]

How do we find \( \Sigma^* \)?

\[
\Sigma^* = \left[ I - F \otimes F \right]^{-1} \text{vec}(GQG')
\]
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 $\Sigma_{t+1|t} \to \Sigma^*$. 
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 $\Sigma$ as $\Sigma_{t+1|t} \rightarrow \Sigma$ for any $\Sigma_{1|0}$ we start with.
The Kalman Filter and DSGE models

- Basic Real Business Cycle model

\[
\max E_0 \sum_{t=0}^{\infty} \beta^t \{ \xi \log c_t + (1 - \xi) \log (1 - l_t) \}
\]

\[
c_t + k_{t+1} = k_t^{\alpha} (e^{zt} l_t)^{1-\alpha} + (1 - \delta) k
\]

\[
z_t = \rho z_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma)
\]

- Parameters: \( \gamma = \{ \alpha, \beta, \rho, \xi, \eta, \sigma \} \)
Equilibrium Conditions

\[
\frac{1}{c_t} = \beta E_t \left\{ \frac{1}{c_{t+1}} \left( 1 + \alpha e^{z_{t+1}} k_{t+1}^{\alpha-1} l_{t+1}^{1-\alpha} - \eta \right) \right\}
\]

\[
\frac{1 - \xi}{1 - l_t} = \frac{\xi}{c_t} (1 - \alpha) e^{z_t} k_t^\alpha l_t^{1-\alpha}
\]

\[
c_t + k_{t+1} = e^{z_t} k_t^\alpha l_t^{1-\alpha} + (1 - \eta) k_t
\]

\[
z_t = \rho z_{t-1} + \varepsilon_t
\]
A Special Case

• We set, unrealistically but rather useful for our point, $\eta = 1$.

• In this case, the model has two important and useful features:

1. First, the income and the substitution effect from a productivity shock to labor supply exactly cancel each other. Consequently, $l_t$ is constant and equal to:

$$l_t = l = \frac{(1 - \alpha) \xi}{(1 - \alpha) \xi + (1 - \xi)(1 - \alpha \beta)}$$

2. Second, the policy function for capital is

$$k_{t+1} = \alpha \beta e^{\gamma t} k_t^\alpha l^{1-\alpha}.$$
A Special Case II

- The definition of \( k_{t+1} \) implies that \( c_t = (1 - \alpha \beta) e^{zt} k_t^{\alpha} l^{1-\alpha} \).

- Let us try if the Euler Equation holds:

\[
\frac{1}{c_t} = \beta E_t \left\{ \frac{1}{c_{t+1}} \left( \alpha e^{zt+1} k_{t+1}^{\alpha} l^{1-\alpha} \right) \right\}
\]

\[
\frac{1}{(1 - \alpha \beta) e^{zt} k_t^{\alpha} l^{1-\alpha}} = \beta E_t \left\{ \frac{1}{(1 - \alpha \beta) e^{zt+1} k_{t+1}^{\alpha} l^{1-\alpha}} \left( \alpha e^{zt+1} k_{t+1}^{\alpha} l^{1-\alpha} \right) \right\}
\]

\[
\frac{1}{(1 - \alpha \beta) e^{zt} k_t^{\alpha} l^{1-\alpha}} = \beta E_t \left\{ \frac{\alpha}{(1 - \alpha \beta) k_{t+1}} \right\}
\]

\[
\frac{\alpha \beta}{(1 - \alpha \beta)} = \frac{\beta \alpha}{(1 - \alpha \beta)}
\]
• Let us try if the Intratemporal condition holds

\[
\frac{1 - \xi}{1 - l} = \frac{\xi}{(1 - \alpha \beta)} e^{z_t k_t^\alpha l^{1-\alpha}} (1 - \alpha) e^{z_t k_t^\alpha l^{-\alpha}}
\]

\[
\frac{1 - \xi}{1 - l} = \frac{\xi}{(1 - \alpha \beta)} \frac{(1 - \alpha)}{l}
\]

\[
(1 - \alpha \beta)(1 - \xi) l = \xi (1 - \alpha) (1 - l)
\]

\[
((1 - \alpha \beta)(1 - \xi) + (1 - \alpha) \xi) l = (1 - \alpha) \xi
\]

• Finally, the budget constraint holds because of the definition of $c_t$. 

Transition Equation

• Since this policy function is linear in logs, we have the transition equation for the model:

\[
\begin{pmatrix}
1 \\
\log k_{t+1} \\
z_t
\end{pmatrix} = 
\begin{pmatrix}
1 & 0 & 0 \\
\log \alpha \beta \lambda^{1-\alpha} & \alpha & \rho \\
0 & 0 & \rho
\end{pmatrix} 
\begin{pmatrix}
1 \\
\log k_t \\
z_{t-1}
\end{pmatrix} + 
\begin{pmatrix}
0 \\
1 \\
1
\end{pmatrix} \epsilon_t.
\]

• Note constant.

• Alternative formulations.
Measurement Equation

- As observables, we assume \( \log y_t \) and \( \log i_t \) subject to a linearly additive measurement error \( V_t = \begin{pmatrix} v_{1,t} & v_{2,t} \end{pmatrix} \).

- Let \( V_t \sim N(0, \Lambda) \), where \( \Lambda \) is a diagonal matrix with \( \sigma_1^2 \) and \( \sigma_2^2 \), as diagonal elements.

- Why measurement error? Stochastic singularity.

- Then:

\[
\begin{pmatrix} \log y_t \\ \log i_t \end{pmatrix} = \begin{pmatrix} -\log \alpha \beta \lambda l^{1-\alpha} & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ \log k_{t+1} \\ z_t \end{pmatrix} + \begin{pmatrix} v_{1,t} \\ v_{2,t} \end{pmatrix}.
\]
The Solution to the Model in State Space Form

\[ x_t = \begin{pmatrix} 1 \\ \log k_t \\ z_{t-1} \end{pmatrix}, \quad z_t = \begin{pmatrix} \log y_t \\ \log i_t \end{pmatrix} \]

\[ F = \begin{pmatrix} 1 & 0 & 0 \\ \log \alpha \beta \lambda^{1-\alpha} & \alpha & \rho \\ 0 & 0 & \rho \end{pmatrix}, \quad G = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \quad Q = \sigma^2 \]

\[ H' = \begin{pmatrix} -\log \alpha \beta \lambda^{1-\alpha} & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad R = \Lambda \]
The Solution to the Model in State Space Form III

• Now, using $z^T$, $F$, $G$, $H$, $Q$, and $R$ as defined in the last slide...

• ...we can use the Ricatti equations to compute the likelihood function of the model:

$$\log l \left( z^T | F, G, H, Q, R \right)$$

• Cross-equations restrictions implied by equilibrium solution.

• With the likelihood, we can do inference!
What do we Do if $\eta \neq 1$?

We have two options:

- First, we could linearize or log-linearize the model and apply the Kalman filter.

- Second, we could compute the likelihood function of the model using a non-linear filter (particle filter).

- Advantages and disadvantages.

The Kalman Filter and linearized DSGE Models

- We linearize (or loglinerize) around the steady state.

- We assume that we have data on log output (log $y_t$), log hours (log $l_t$), and log investment (log $c_t$) subject to a linearly additive measurement error $V_t = \begin{pmatrix} v_1,t & v_2,t & v_3,t \end{pmatrix}'.$

- We need to write the model in state space form. Remember that

$$\hat{k}_{t+1} = P\hat{k}_t + Qz_t$$

and

$$\hat{l}_t = R\hat{k}_t + Sz_t$$
Writing the Likelihood Function I

- The transitions Equation:

\[
\begin{pmatrix}
1 \\
\hat{k}_{t+1} \\
\hat{z}_{t+1}
\end{pmatrix} = \begin{pmatrix}
1 & 0 & 0 \\
0 & P & Q \\
0 & 0 & \rho
\end{pmatrix} \begin{pmatrix}
1 \\
\hat{k}_t \\
\hat{z}_t
\end{pmatrix} + \begin{pmatrix}
0 \\
0 \\
1
\end{pmatrix} \epsilon_t.
\]

- The Measurement Equation requires some care.
Writing the Likelihood Function II

• Notice that $\hat{y}_t = z_t + \alpha \hat{k}_t + (1 - \alpha)\hat{l}_t$

• Therefore, using $\hat{l}_t = R\hat{k}_t + S z_t$

\[
\hat{y}_t = z_t + \alpha \hat{k}_t + (1 - \alpha)(R\hat{k}_t + S z_t) = \\
(\alpha + (1 - \alpha)R) \hat{k}_t + (1 + (1 - \alpha)S) z_t
\]

• Also since $\hat{c}_t = -\alpha_5 \hat{l}_t + z_t + \alpha \hat{k}_t$ and using again $\hat{l}_t = R\hat{k}_t + S z_t$

\[
\hat{c}_t = z_t + \alpha \hat{k}_t - \alpha_5(R\hat{k}_t + S z_t) = \\
(\alpha - \alpha_5 R) \hat{k}_t + (1 - \alpha_5 S) z_t
\]
Writing the Likelihood Function III

Therefore the measurement equation is:

\[
\begin{pmatrix}
\log y_t \\
\log l_t \\
\log c_t
\end{pmatrix} = \begin{pmatrix}
\log y & \alpha + (1 - \alpha)R & 1 + (1 - \alpha)S \\
\log l & R & S \\
\log c & \alpha - \alpha_5R & 1 - \alpha_5S
\end{pmatrix} \begin{pmatrix}
1 \\
\hat{k}_t \\
z_t
\end{pmatrix} + \begin{pmatrix}
v_{1,t} \\
v_{2,t} \\
v_{3,t}
\end{pmatrix}.
\]
The Likelihood Function of a General Dynamic Equilibrium Economy

- Transition equation:
  \[ S_t = f (S_{t-1}, W_t; \gamma) \]

- Measurement equation:
  \[ Y_t = g (S_t, V_t; \gamma) \]

- Interpretation.
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 \( p(y_t|W_{1,t}^{t}, y_{t-1}, S_0; \gamma) \).
   Lubick and Schorfheide (2003).

3. The model assigns positive probability to the data.
Our Goal: Likelihood Function

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

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

- We factorize it as:

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

$$ = \prod_{t=1}^{T} \int \int p(y_t|W_1^t, y^{t-1}, S_0; \gamma) p(W_1^t, S_0|y^{t-1}; \gamma) \, dW_1^t \, dS_0 $$
A Law of Large Numbers

If \( \left\{ \left\{ s_{0}^{t-1,i}, w_{1}^{t-1,i} \right\} \right\}_{i=1}^{N} \) \( \overset{T}{\underset{t=1}{\text{N i.i.d. draws from}}} \) \( \left\{ p\left( W_{1}^{t}, S_{0} | y^{t-1}; \gamma \right) \right\}_{t=1}^{T} \),

then:

\[
p \left( y^{T}; \gamma \right) \backsimeq \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} p \left( y_{t} | w_{1}^{t-1,i}, y^{t-1}, s_{0}^{t-1,i}; \gamma \right)
\]
...thus

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

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

- \( \{s_{0}^{t-1,i}, w_{1}^{t-1,i}\}_{i=1}^{N} \) \( N \) 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 \( \{s_{0}^{t-1,i}, w_{1}^{t-1,i}\}_{i=1}^{N} \) a swarm of particles.

- \( \{s_{0}^{t|t-1,i}, w_{1}^{t|t-1,i}\}_{i=1}^{N} \) \( N \) 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 \( \{s_{0}^{t|t-1,i}, w_{1}^{t|t-1,i}\}_{i=1}^{N} \) a swarm of proposed particles.
... and Weights

\[
q_t^i = \frac{p\left(y_t \mid w_{1 \mid t-1,i}^t, y_{t-1}, s_{0 \mid t-1,i}^t; \gamma\right)}{\sum_{i=1}^{N} p\left(y_t \mid w_{1 \mid t-1,i}^t, y_{t-1}, s_{0 \mid t-1,i}^t; \gamma\right)}
\]
A Proposition

Let \( \{\tilde{s}_0^i, \tilde{w}_1^i\}_{i=1}^N \) be a draw with replacement from \( \{s_0^{t-1,i}, w_1^{t-1,i}\}_{i=1}^N \) and probabilities \( q_t^i \). Then \( \{\tilde{s}_0^i, \tilde{w}_1^i\}_{i=1}^N \) is a draw from \( p(W_t^t, S_0|y^t; \gamma) \).
Importance of the Proposition

1. It shows how a draw \( \left\{ s_{0|t-1,i}, w_{1|t-1,i} \right\}_{i=1}^{N} \) from \( p(W_t, S_0|y^{t-1}; \gamma) \)

   can be used to draw \( \left\{ s_{t,i}, w_{t,i} \right\}_{i=1}^{N} \) from \( p(W_t, S_0|y^{t}; \gamma) \).

2. With a draw \( \left\{ s_{0,t}, w_{1,t} \right\}_{i=1}^{N} \) from \( p(W_1, S_0|y^{t}; \gamma) \) we can use \( p(W_{1,t+1}; \gamma) \)

   to get a draw \( \left\{ s_{0|t+1,i}, w_{1|t+1,i} \right\}_{i=1}^{N} \) and iterate the procedure.
Sequential Monte Carlo I: Filtering

Step 0, Initialization: Set $t \sim 1$ and initialize $p(W_{t-1}^t, S_0|y^{t-1}; \gamma) = p(S_0; \gamma)$.

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

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

Step 3, Sampling: Draw $\left\{ s_0^i, w_1^i \right\}_{i=1}^N$ with rep. from $\left\{ s_0^{t-1,i}, w_1^{t-1,i} \right\}_{i=1}^N$ with probabilities $\left\{ q_i^t \right\}_{i=1}^N$. If $t < T$ set $t \sim t + 1$ and go to step 1. Otherwise stop.
Sequential Monte Carlo II: Likelihood

Use \( \left\{ \left\{ s_{0|t-1,i}, w_{1|t-1,i} \right\} \right\}^{N}_{i=1} \) to compute:

\[
p \left( y^T; \gamma \right) \approx \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} p \left( y_t | w_{1|t-1,i}, y_{t-1}, s_{0|t-1,i}; \gamma \right)
\]
A “Trivial” Application

How do we evaluate the likelihood function \( p(y^T|\alpha, \beta, \sigma) \) of the nonlinear, nonnormal process:

\[
\begin{align*}
  s_t &= \alpha + \beta \frac{s_{t-1}}{1 + s_{t-1}} + w_t \\
  y_t &= s_t + v_t
\end{align*}
\]

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

2. Generate $N$ i.i.d. draws $\left\{ s_{0,1|i}, w_{1,0|i} \right\}_{i=1}^N$ from $\mathcal{N}(0, \sigma)$.

3. Evaluate $p \left( y_1 \mid w_{1,0|i}, y_0, s_{0,1|i} \right) = p_t(2) \left( y_1 - \left( \alpha + \beta \frac{s_{0,1|i}}{1 + s_{0,1|i}} + w_{1,0|i} \right) \right)$.

4. Evaluate the relative weights $q_1^i = \frac{p_t(2) \left( y_1 - \left( \alpha + \beta \frac{s_{0,1|i}}{1 + s_{0,1|i}} + w_{1,0|i} \right) \right)}{\sum_{i=1}^N p_t(2) \left( y_1 - \left( \alpha + \beta \frac{s_{0,1|i}}{1 + s_{0,1|i}} + w_{1,0|i} \right) \right)}$. 

74
5. Resample with replacement \( N \) values of \( \left\{ s_{0}^{1|0,i}, w_{0}^{1|0,i} \right\}_{i=1}^{N} \) with relative weights \( q_{i}^{1} \). Call those sampled values \( \left\{ s_{0}^{1,i}, w_{0}^{1,i} \right\}_{i=1}^{N} \).

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:

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

and consequently:

\[ p\left( y^T | \alpha, \beta, \sigma \right) \approx \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} p\left( y_t | w_1^{t|t-1,i}, y^{t-1}, s_0^{t|t-1,i} \right) \]
Comparison with Alternative Schemes

- Deterministic algorithms: Extended Kalman Filter and derivations (Jazwinski, 1973), Gaussian Sum approximations (Alspach and Sorenson, 1972), grid-based filters (Bucy and Senne, 1974), Jacobian of the transform (Miranda and Rui, 1997).

Tanizaki (1996).

A “Real” Application: the Stochastic Neoclassical Growth Model

• Standard model.

• Isn’t the model nearly linear?

• Yes, but:
  1. Better to begin with something easy.
  2. We will learn something nevertheless.
The Model

- Representative agent with utility function $U = E_0 \sum_{t=0}^{\infty} \beta^t \frac{(c_t^\theta (1-l_t)^{1-\theta})^{1-\tau}}{1-\tau}$.

- One good produced according to $y_t = e^{z_t} A k_t^{\alpha} l_t^{1-\alpha}$ with $\alpha \in (0, 1)$.

- Productivity evolves $z_t = \rho z_{t-1} + \epsilon_t$, $|\rho| < 1$ and $\epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon)$.

- Law of motion for capital $k_{t+1} = i_t + (1-\delta)k_t$.

- Resource constrain $c_t + i_t = y_t$. 
• Solve for $c(\cdot, \cdot)$ and $l(\cdot, \cdot)$ given initial conditions.

• Characterized by:

$$U_c(t) = \beta E_t \left[ U_c(t + 1) \left( 1 + \alpha A e^{zt+1} k^{\alpha-1}_{t+1} l(k_{t+1}, z_{t+1})^\alpha - \delta \right) \right]$$

$$\frac{1 - \theta}{\theta} \frac{c(k_t, z_t)}{1 - l(k_t, z_t)} = (1 - \alpha) e^{zt} A k^\alpha_t l(k_t, z_t)^{-\alpha}$$

• A system of functional equations with no known analytical solution.
Solving the Model

- We need to use a numerical method to solve it.

- Different nonlinear approximations: value function iteration, perturbation, projection methods.

  1. Speed: sparse system.
  3. Scalable.
Building the Likelihood Function

• Time series:

  1. Quarterly real output, hours worked and investment.

  2. Main series from the model and keep dimensionality low.

• Measurement error. Why?

• \( \gamma = (\theta, \rho, \tau, \alpha, \delta, \beta, \sigma_\epsilon, \sigma_1, \sigma_2, \sigma_3) \)
State Space Representation

\[ k_t = f_1(S_{t-1}, W_t; \gamma) = e^{\tanh^{-1}(\lambda_{t-1}) k_{t-1}^\alpha} l \left( k_{t-1}, \tanh^{-1}(\lambda_{t-1}); \gamma \right)^{1-\alpha} \]

\[ \left( 1 - \frac{\theta}{1 - \theta} (1 - \alpha) \frac{1 - l \left( k_{t-1}, \tanh^{-1}(\lambda_{t-1}); \gamma \right)}{l \left( k_{t-1}, \tanh^{-1}(\lambda_{t-1}); \gamma \right)} \right) + (1 - \delta) k_{t-1} \]

\[ \lambda_t = f_2(S_{t-1}, W_t; \gamma) = \tanh(\rho \tanh^{-1}(\lambda_{t-1}) + \epsilon_t) \]

\[ gdp_t = g_1(S_t, V_t; \gamma) = e^{\tanh^{-1}(\lambda_t) k_t^\alpha} l \left( k_t, \tanh^{-1}(\lambda_t); \gamma \right)^{1-\alpha} + V_{1,t} \]

\[ hours_t = g_2(S_t, V_t; \gamma) = l \left( k_t, \tanh^{-1}(\lambda_t); \gamma \right) + V_{2,t} \]

\[ inv_t = g_3(S_t, V_t; \gamma) = e^{\tanh^{-1}(\lambda_t) k_t^\alpha} l \left( k_t, \tanh^{-1}(\lambda_t); \gamma \right)^{1-\alpha} \]

\[ \left( 1 - \frac{\theta}{1 - \theta} (1 - \alpha) \frac{1 - l \left( k_t, \tanh^{-1}(\lambda_t); \gamma \right)}{l \left( k_t, \tanh^{-1}(\lambda_t); \gamma \right)} \right) + V_{3,t} \]

Likelihood Function
Since our measurement equation implies that

\[ p(y_t|S_t; \gamma) = (2\pi)^{-\frac{3T}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{\omega(S_t;\gamma)}{2}} \]

where \( \omega(S_t;\gamma) = (y_t - x(S_t;\gamma))' \Sigma^{-1} (y_t - x(S_t;\gamma)) \) \( \forall t \), we have

\[
p(y^T; \gamma) =
(2\pi)^{-\frac{3T}{2}} |\Sigma|^{-\frac{T}{2}} \int \left( \prod_{t=1}^{T} \int e^{-\frac{\omega(S_t;\gamma)}{2}} p(S_t|y^{t-1}, S_0; \gamma) dS_t \right) p(S_0; \gamma) dS_1
\]

\[
\simeq (2\pi)^{-\frac{3T}{2}} |\Sigma|^{-\frac{T}{2}} \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} e^{-\frac{\omega(s^i_t;\gamma)}{2}}
\]
### Priors for the Parameters of the Model

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Distribution</th>
<th>Hyperparameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>Uniform</td>
<td>0,1</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Uniform</td>
<td>0,1</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Uniform</td>
<td>0,100</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Uniform</td>
<td>0,1</td>
</tr>
<tr>
<td>$\delta$</td>
<td>Uniform</td>
<td>0,0.05</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Uniform</td>
<td>0.75,1</td>
</tr>
<tr>
<td>$\sigma_{\epsilon}$</td>
<td>Uniform</td>
<td>0,0.1</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>Uniform</td>
<td>0,0.1</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>Uniform</td>
<td>0,0.1</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>Uniform</td>
<td>0,0.1</td>
</tr>
</tbody>
</table>
Likelihood-Based Inference I: a Bayesian Perspective

- Define priors over parameters: truncated uniforms.

- Use a Random-walk Metropolis-Hastings to draw from the posterior.

- Find the Marginal Likelihood.
Likelihood-Based Inference II: a Maximum Likelihood Perspective

• We only need to maximize the likelihood.

• Difficulties to maximize with Newton type schemes.

• Common problem in dynamic equilibrium economies.

• We use a simulated annealing scheme.
An Exercise with Artificial Data

- First simulate data with our model and use that data as sample.

- Pick “true” parameter values. Benchmark calibration values for the stochastic neoclassical growth model (Cooley and Prescott, 1995).

### Calibrated Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\theta$</th>
<th>$\rho$</th>
<th>$\tau$</th>
<th>$\alpha$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.357</td>
<td>0.95</td>
<td>2.0</td>
<td>0.4</td>
<td>0.02</td>
</tr>
<tr>
<td>Parameter</td>
<td>$\beta$</td>
<td>$\sigma_\epsilon$</td>
<td>$\sigma_1$</td>
<td>$\sigma_2$</td>
<td>$\sigma_3$</td>
</tr>
<tr>
<td>Value</td>
<td>0.99</td>
<td>0.007</td>
<td>$1.58\times10^{-4}$</td>
<td>0.0011</td>
<td>$8.66\times10^{-4}$</td>
</tr>
</tbody>
</table>

- Sensitivity: $\tau = 50$ and $\sigma_\epsilon = 0.035$. 
Figure 5.1: Likelihood Function Benchmark Calibration

Likelihood cut at $\rho$

Likelihood cut at $\tau$

Likelihood cut at $\alpha$

Likelihood cut at $\delta$

Likelihood cut at $\sigma$

Likelihood cut at $\beta$

Likelihood cut at $\theta$

- Nonlinear
- Linear
- Pseudotrue
Figure 5.2: Posterior Distribution Benchmark Calibration

- $\rho$
- $\tau$
- $\alpha$
- $\delta$
- $\sigma$
- $\beta$
- $\theta$
- $\sigma_1$
- $\sigma_2$
- $\sigma_3$
Figure 5.3: Likelihood Function Extreme Calibration

Likelihood cut $\rho$

Likelihood cut $\tau$

Likelihood cut $\alpha$

Likelihood cut $\delta$

Likelihood cut $\sigma$

Likelihood cut $\beta$

Likelihood cut $\theta$
Figure 5.4: Posterior Distribution Extreme Calibration
Figure 5.5: Converge of Posteriors Extreme Calibration
Figure 5.6: Posterior Distribution Real Data
Figure 6.1: Likelihood Function

Transversal cut at $\alpha$

Transversal cut at $\beta$

Transversal cut at $\rho$

Transversal cut at $\sigma_x$
Figure 6.3: C.D.F. Extreme Calibration
Figure 6.4: C.D.F. Real Data
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Mean</th>
<th>s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>0.357</td>
<td>$6.72 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.950</td>
<td>$3.40 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\tau$</td>
<td>2.000</td>
<td>$6.78 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.400</td>
<td>$8.60 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.020</td>
<td>$1.34 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.989</td>
<td>$1.54 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\sigma_\epsilon$</td>
<td>0.007</td>
<td>$9.29 \times 10^{-6}$</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>$1.58 \times 10^{-4}$</td>
<td>$5.75 \times 10^{-8}$</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>$1.12 \times 10^{-2}$</td>
<td>$6.44 \times 10^{-7}$</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>$8.64 \times 10^{-4}$</td>
<td>$6.49 \times 10^{-7}$</td>
</tr>
<tr>
<td>Parameters</td>
<td>MLE</td>
<td>s.d.</td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>--------------</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.357</td>
<td>8.19×10^{-6}</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.950</td>
<td>0.001</td>
</tr>
<tr>
<td>$\tau$</td>
<td>2.000</td>
<td>0.020</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.400</td>
<td>2.02×10^{-6}</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.002</td>
<td>2.07×10^{-5}</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.990</td>
<td>1.00×10^{-6}</td>
</tr>
<tr>
<td>$\sigma_\epsilon$</td>
<td>0.007</td>
<td>0.004</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>1.58×10^{-4}</td>
<td>0.007</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>1.12×10^{-3}</td>
<td>0.007</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>8.63×10^{-4}</td>
<td>0.005</td>
</tr>
<tr>
<td>Parameters</td>
<td>Mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>------------</td>
<td>----------</td>
<td>---------------</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.357</td>
<td>$7.19 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.950</td>
<td>$1.88 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\tau$</td>
<td>50.00</td>
<td>$7.12 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.400</td>
<td>$4.80 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.020</td>
<td>$3.52 \times 10^{-6}$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.989</td>
<td>$8.69 \times 10^{-6}$</td>
</tr>
<tr>
<td>$\sigma_\epsilon$</td>
<td>0.035</td>
<td>$4.47 \times 10^{-6}$</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>$1.58 \times 10^{-4}$</td>
<td>$1.87 \times 10^{-8}$</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>$1.12 \times 10^{-2}$</td>
<td>$2.14 \times 10^{-7}$</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>$8.65 \times 10^{-4}$</td>
<td>$2.33 \times 10^{-7}$</td>
</tr>
<tr>
<td>Parameters</td>
<td>MLE</td>
<td>s.d.</td>
</tr>
<tr>
<td>------------</td>
<td>-------</td>
<td>----------------</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.357</td>
<td>$2.42 \times 10^{-6}$</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.950</td>
<td>$6.12 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\tau$</td>
<td>50.000</td>
<td>0.022</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.400</td>
<td>$3.62 \times 10^{-7}$</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.019</td>
<td>$7.43 \times 10^{-6}$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.990</td>
<td>$1.00 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\sigma_\epsilon$</td>
<td>0.035</td>
<td>0.015</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>$1.58 \times 10^{-4}$</td>
<td>0.017</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>$1.12 \times 10^{-3}$</td>
<td>0.014</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>$8.66 \times 10^{-4}$</td>
<td>0.023</td>
</tr>
</tbody>
</table>
## Convergence on Number of Particles

<table>
<thead>
<tr>
<th>N</th>
<th>Mean</th>
<th>s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>10000</td>
<td>1014.558</td>
<td>0.3296</td>
</tr>
<tr>
<td>20000</td>
<td>1014.600</td>
<td>0.2595</td>
</tr>
<tr>
<td>30000</td>
<td>1014.653</td>
<td>0.1829</td>
</tr>
<tr>
<td>40000</td>
<td>1014.666</td>
<td>0.1604</td>
</tr>
<tr>
<td>50000</td>
<td>1014.688</td>
<td>0.1465</td>
</tr>
<tr>
<td>60000</td>
<td>1014.664</td>
<td>0.1347</td>
</tr>
</tbody>
</table>
## Posterior Distributions Real Data

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Mean</th>
<th>s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>0.323</td>
<td>$7.976 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.969</td>
<td>0.008</td>
</tr>
<tr>
<td>$\tau$</td>
<td>1.825</td>
<td>0.011</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.388</td>
<td>0.001</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.006</td>
<td>$3.557 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.997</td>
<td>$9.221 \times 10^{-5}$</td>
</tr>
<tr>
<td>$\sigma_\epsilon$</td>
<td>0.023</td>
<td>$2.702 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.039</td>
<td>$5.346 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.018</td>
<td>$4.723 \times 10^{-4}$</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>0.034</td>
<td>$6.300 \times 10^{-4}$</td>
</tr>
<tr>
<td>Parameters</td>
<td>MLE</td>
<td>s.d.</td>
</tr>
<tr>
<td>------------</td>
<td>-----</td>
<td>------</td>
</tr>
<tr>
<td>$\theta$</td>
<td>0.390</td>
<td>0.044</td>
</tr>
<tr>
<td>$\rho$</td>
<td>0.987</td>
<td>0.708</td>
</tr>
<tr>
<td>$\tau$</td>
<td>1.781</td>
<td>1.398</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.324</td>
<td>0.019</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.006</td>
<td>0.160</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.997</td>
<td>$8.67 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\sigma_\epsilon$</td>
<td>0.023</td>
<td>0.224</td>
</tr>
<tr>
<td>$\sigma_1$</td>
<td>0.038</td>
<td>0.060</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.016</td>
<td>0.061</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>0.035</td>
<td>0.076</td>
</tr>
</tbody>
</table>
Logmarginal Likelihood Difference: Nonlinear-Linear

<table>
<thead>
<tr>
<th>p</th>
<th>Benchmark Calibration</th>
<th>Extreme Calibration</th>
<th>Real Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>73.631</td>
<td>117.608</td>
<td>93.65</td>
</tr>
<tr>
<td>0.5</td>
<td>73.627</td>
<td>117.592</td>
<td>93.55</td>
</tr>
<tr>
<td>0.9</td>
<td>73.603</td>
<td>117.564</td>
<td>93.55</td>
</tr>
<tr>
<td></td>
<td>Real Data</td>
<td>Nonlinear (SMC filter)</td>
<td>Linear (Kalman filter)</td>
</tr>
<tr>
<td>----------------</td>
<td>---------------------</td>
<td>------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>s.d</td>
<td>Mean</td>
</tr>
<tr>
<td><strong>output</strong></td>
<td>1.95</td>
<td>0.073</td>
<td>1.91</td>
</tr>
<tr>
<td><strong>hours</strong></td>
<td>0.36</td>
<td>0.014</td>
<td>0.36</td>
</tr>
<tr>
<td><strong>inv</strong></td>
<td>0.42</td>
<td>0.066</td>
<td>0.44</td>
</tr>
</tbody>
</table>
A “Future” Application: Good Luck or Good Policy?

- U.S. economy has become less volatile over the last 20 years (Stock and Watson, 2002).

- Why?


How Has the Literature Addressed this Question?

• So far: mostly with reduced form models (usually VARs).

• But:

  1. Results difficult to interpret.

  2. How to run counterfactuals?

Why Not a Dynamic Equilibrium Model?


• Linear and Normal.

• But we can do it!!!
Environment

- Discrete time $t = 0, 1, ...$

- Stochastic process $s \in S$ with history $s^t = (s_0, ..., s_t)$ and probability $\mu(s^t)$. 
The Final Good Producer

- Perfectly Competitive Final Good Producer that solves

\[
\max_{y_i(s^t)} \left( \int y_i(s^t)^\theta \, di \right)^{\frac{1}{\theta}} - \int p_i(s^t) y_i(s^t) \, di.
\]

- Demand function for each input of the form

\[
y_i(s^t) = \left( \frac{p_i(s^t)}{p(s^t)} \right)^{\frac{1}{\theta-1}} y(s^t),
\]

with price aggregator:

\[
p(s^t) = \left( \int p_i(s^t)^{\frac{\theta}{\theta-1}} \, di \right)^{\frac{\theta-1}{\theta}}.
\]
The Intermediate Good Producer

- Continuum of intermediate good producers, each of one behaving as monopolistic competitor.

- The producer of good $i$ has access to the technology:

$$y_i(s^t) = \max \left\{ e^{z(s^t) k_i^{\alpha} (s^{t-1})} l_i^{1-\alpha} (s^t) - \phi, 0 \right\}.$$ 

- Productivity $z(s^t) = \rho z(s^{t-1}) + \varepsilon z(s^t)$.

- Calvo pricing with indexing. Probability of changing prices (before observing current period shocks) $1 - \zeta$. 103
Consumers Problem

\[ E \sum_{t=0}^{\infty} \beta^t \left\{ \epsilon_c (s^t) \frac{(c(s^t) - dc(s^{t-1}))}{\sigma_c} - \epsilon_l (s^t) \frac{l(s^t)}{\sigma_l} + \epsilon_m (s^t) \frac{m(s^t)}{\sigma_m} \right\} \]

\[ p(s^t) (c(s^t) + x(s^t)) + M(s^t) + \int_{s^{t+1}} q(s^{t+1} | s^t) B(s^{t+1}) ds_{t+1} = \]

\[ p(s^t) (w(s^t) l(s^t) + r(s^t) k(s^{t-1})) + M(s^{t-1}) + B(s^t) + \Pi(s^t) + T(s^t) \]

\[ B(s^{t+1}) \geq B \]

\[ k(s^t) = (1 - \delta) k(s^{t-1}) - \phi \left( \frac{x(s^t)}{k(s^{t-1})} \right) + x(s^t). \]
Government Policy

- Monetary Policy: Taylor rule

\[
   i(s^t) = r_g \pi_g(s^t) \\
   + a(s^t) \left( \pi(s^t) - \pi_g(s^t) \right) \\
   + b(s^t) \left( y(s^t) - y_g(s^t) \right) + \varepsilon_i(s^t)
\]

\[
   \pi_g(s^t) = \pi_g(s^{t-1}) + \varepsilon_{\pi}(s^t)
\]

\[
   a(s^t) = a(s^{t-1}) + \varepsilon_a(s^t)
\]

\[
   b(s^t) = b(s^{t-1}) + \varepsilon_b(s^t)
\]

- Fiscal Policy.
Stochastic Volatility I

- We can stack all shocks in one vector:
  \[
  \varepsilon(s^t) = (\varepsilon_z(s^t), \varepsilon_c(s^t), \varepsilon_l(s^t), \varepsilon_m(s^t), \varepsilon_i(s^t), \varepsilon_{\pi}(s^t), \varepsilon_a(s^t), \varepsilon_b(s^t))'
  \]

- Stochastic volatility:
  \[
  \varepsilon(s^t) = R(s^t)^{0.5} \vartheta(s^t).
  \]

- The matrix \(R(s^t)\) can be decomposed as:
  \[
  R(s^t) = G(s^t)^{-1} H(s^t) G(s^t).
  \]
Stochastic Volatility II

- $H\left(s^t\right)$ (instantaneous shocks variances) is diagonal with nonzero elements $h_i\left(s^t\right)$ that evolve:

  \[
  \log h_i\left(s^t\right) = \log h_i\left(s^{t-1}\right) + \varsigma_i \eta_i\left(s^t\right).
  \]

- $G\left(s^t\right)$ (loading matrix) is lower triangular, with unit entries in the diagonal and entries $\gamma_{ij}\left(s^t\right)$ that evolve:

  \[
  \gamma_{ij}\left(s^t\right) = \gamma_{ij}\left(s^{t-1}\right) + \omega_{ij} \nu_{ij}\left(s^t\right).
  \]
Where Are We Now?

- Solving the model: problem with 45 state variables: physical capital, the aggregate price level, 7 shocks, 8 elements of matrix \( H(s^t) \), and the 28 elements of the matrix \( G(s^t) \).

- Perturbation.

- We are making good progress.