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

$$x_{t+1} = Fx_t + G\omega_{t+1},\ \omega_{t+1} \sim \mathcal{N}(0, Q)$$

  - Measurement equation

$$z_t = H'x_t + \upsilon_t,\ \upsilon_t \sim \mathcal{N}(0, R)$$

  - 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

$$x_{t+1}^* = F^* x_t^* + G^* \omega_{t+1},\ \omega_{t+1} \sim \mathcal{N}(0, Q)$$

  – Measurement equation

$$z_t = H^{*\prime} x_t^* + \upsilon_t,\ \upsilon_t \sim \mathcal{N}(0, R)$$

Example I

- Assume the following AR(2) process:

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

- Model is not apparently not Markovian.

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

- Yes!

State Space Representation I

- Transition equation:

$$x_t = \begin{bmatrix} \rho_1 & 1 \\ \rho_2 & 0 \end{bmatrix} x_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \upsilon_t$$

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:

$$x_t = \begin{bmatrix} \rho_1 & \rho_2 \\ 1 & 0 \end{bmatrix} x_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \upsilon_t$$

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:

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

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

- However, it will again be straightforward to write a state space representation.

State Space Representation I

- Transition equation:

$$x_t = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x_{t-1} + \begin{bmatrix} 1 \\ \theta \end{bmatrix} \upsilon_t$$

where $x_t = \begin{bmatrix} y_t & \theta \upsilon_t \end{bmatrix}'$

- Measurement equation:

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

State Space Representation II

- Transition equation:

$$x_t = \upsilon_{t-1}$$

- where $x_t = [\upsilon_{t-1}]'$

- Measurement equation:

$$z_t = \theta x_t + \upsilon_t$$

- Again both representations are equivalent!

Example III

- Assume the following random walk plus drift process:

$$z_t = z_{t-1} + \beta + \upsilon_t, \; \upsilon_t \sim \mathcal{N}\left(0, \sigma_\upsilon^2\right)$$

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

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

- Let $x_{t|t} = E\left(x_t | z^t\right)$ 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 $x_{t+1|t} = F x_{t|t}$ 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.

A Minimization Approach to the Kalman Filter I

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

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

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

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

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

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 I

Assume the following model in State Space form:

- Transition equation

$$x_t = \mu + \upsilon_t,\ \upsilon_t \sim N\left(0, \sigma_\upsilon^2\right)$$

- Measurement equation

$$z_t = x_t + \xi_t,\ \xi_t \sim N\left(0, \sigma_\xi^2\right)$$

- Let $\sigma_\xi^2 = q\sigma_\upsilon^2$.

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 σ_ξ^2 relative to σ_v^2 (the bigger q) the lower K_1 and the less we trust z_1.

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:

$$\begin{aligned}
\Omega_{t|t-1} &\equiv E\left(\left(z_t - z_{t|t-1}\right)\left(z_t - z_{t|t-1}\right)' |z^{t-1}\right) \\
&= E\left(\begin{array}{c} H'\left(x_t - x_{t|t-1}\right)\left(x_t - x_{t|t-1}\right)' H + \\ \upsilon_t\left(x_t - x_{t|t-1}\right)' H + H'\left(x_t - x_{t|t-1}\right)\upsilon_t' + \\ \upsilon_t \upsilon_t' |z^{t-1} \end{array}\right) \\
&= H'\Sigma_{t|t-1} H + R
\end{aligned}$$

The Kalman Filter Algorithm II

Let $\mathbf{\Sigma}_{t|t-1}$, then we compute:

$$
\begin{aligned}
E\left(\left(z_t - z_{t|t-1}\right)\left(x_t - x_{t|t-1}\right)' |z^{t-1}\right) &= \\
E\left(H'\left(x_t - x_{t|t-1}\right)\left(x_t - x_{t|t-1}\right)' + \upsilon_t\left(x_t - x_{t|t-1}\right)' |z^{t-1}\right) &= H'\mathbf{\Sigma}_{t|t-1}
\end{aligned}
$$

Let $\mathbf{\Sigma}_{t|t-1}$, then we compute:

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

Let $\mathbf{\Sigma}_{t|t-1}$, $x_{t|t-1}$, 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:

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

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

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

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 $\Sigma_{t|t-1}$.

The, 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:

$$\Sigma_{t|t-1} \equiv E\left(\left(x_t - x_{t|t-1}\right)\left(x_t - x_{t|t-1}\right)' |z^{t-1}\right)$$

Some Derivations I

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

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

- Let

$$\Omega_{t|t-1} \equiv E\left(\left(z_t - z_{t|t-1}\right)\left(z_t - z_{t|t-1}\right)'|z^{t-1}\right) =$$

$$E\left(\begin{array}{c} H'\left(x_t - x_{t|t-1}\right)\left(x_t - x_{t|t-1}\right)'H+ \\ \upsilon_t\left(x_t - x_{t|t-1}\right)'H+ \\ H'\left(x_t - x_{t|t-1}\right)\upsilon_t'+ \\ \upsilon_t\upsilon_t'|z^{t-1} \end{array}\right) = H'\Sigma_{t|t-1}H + R$$

Some Derivations II

Finally, let

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

$$E\left(H'\left(x_t - x_{t|t-1}\right)\left(x_t - x_{t|t-1}\right)' + \upsilon_t\left(x_t - x_{t|t-1}\right)' | z^{t-1}\right) =$$

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

The Kalman Filter First Iteration I

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

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

- Remember that:

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

The Kalman Filter First Iteration II

Then, we can write:

$$x_1|z_1, z^0 = x_1|z^1 \sim N\left(x_{1|1}, \boldsymbol{\Sigma}_{1|1}\right)$$

where

$$x_{1|1} = x_{1|0} + \boldsymbol{\Sigma}_{1|0} H \left(H'\boldsymbol{\Sigma}_{1|0} H + R\right)^{-1} \left(z_1 - H' x_{1|0}\right)$$

and

$$\boldsymbol{\Sigma}_{1|1} = \boldsymbol{\Sigma}_{1|0} - \boldsymbol{\Sigma}_{1|0} H \left(H'\boldsymbol{\Sigma}_{1|0} H + R\right)^{-1} H'\boldsymbol{\Sigma}_{1|0}$$

- 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 $z_{2|1} = H' x_{2|1} + \upsilon_{2|1}$:
  - $x_{2|1} = F x_{1|1}$
  - $\Sigma_{2|1} = F \Sigma_{1|1} F' + G Q G'$

The Kalman Filter th Iteration I

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

$$\left(\begin{array}{c} x_t \\ z_t \end{array} |z^{t-1}\right) N\left(\left[\begin{array}{c} x_{t|t-1} \\ H'x_{t|t-1} \end{array}\right], \left[\begin{array}{cc} \Sigma_{t|t-1} & \Sigma_{t|t-1}H \\ H'\Sigma_{t|t-1} & H'\Sigma_{t|t-1}H + R \end{array}\right]\right)$$

- 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 th Iteration II

Then, we can write:

$$x_t|z_t, z^{t-1} = x_t|z^t \sim N\left(x_{t|t}, \boldsymbol{\Sigma}_{t|t}\right)$$

where

$$x_{t|t} = x_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1}H\left(H'\boldsymbol{\Sigma}_{t|t-1}H + R\right)^{-1}\left(z_t - H'x_{t|t-1}\right)$$

and

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

The Kalman Filter Algorithm

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

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

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

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

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:

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

- But since:

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

- We can also write in the probabilistic approach:

$$\begin{aligned} x_{t|t} &= x_{t|t-1} + \boldsymbol{\Sigma}_{t|t-1} H \left(H' \boldsymbol{\Sigma}_{t|t-1} H + R \right)^{-1} \left(z_t - H' x_{t|t-1} \right) = \\ &= x_{t|t-1} + K_t \left(z_t - H' x_{t|t-1} \right) \end{aligned}$$

- Therefore, both approaches are equivalent.

Writing the Likelihood Function

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

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

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

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

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

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

Initial conditions for the Kalman Filter

- An important step in the Kalman Fitler 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

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

How do we find Σ^*?

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

Initial conditions for the Kalman Filter II

Under the following conditions:

1. The system is stable, i.e. all eigenvalues of F are strictly less than one in absolute value.

2. GQG' and R are p.s.d. symmetric

3. $\Sigma_{1|0}$ is p.s.d. symmetric

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

Remarks

1. There are more general theorems than the one just described.

2. Those theorems are based on non-stable systems.

3. Since we are going to work with stable system the former theorem is enough.

4. Last theorem gives us a way to find Σ as $\Sigma_{t+1|t} \rightarrow \Sigma$ for any $\Sigma_{1|0}$ we start with.

The Kalman Filter and DSGE models

- Basic Real Business Cycle model

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

$$\begin{aligned} c_t + k_{t+1} &= k_t^{\alpha} \left(e^{z_t} l_t \right)^{1-\alpha} + (1-\delta) k \\ z_t &= \rho z_{t-1} + \varepsilon_t, \ \varepsilon_t \sim \mathcal{N}(0, \sigma) \end{aligned}$$

- Parameters: $\gamma = \{\alpha, \beta, \rho, \xi, \eta, \sigma\}$

Equilibrium Conditions

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

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

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

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

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 $k_{t+1} = \alpha\beta e^{z_t} k_t^{\alpha} l^{1-\alpha}$.

A Special Case II

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

- Let us try if the Euler Equation holds:

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

$$\frac{1}{(1 - \alpha\beta)\, e^{z_t} k_t^{\alpha} l^{1-\alpha}} = \beta E_t \left\{ \frac{1}{(1 - \alpha\beta)\, e^{z_t+1} k_{t+1}^{\alpha} l^{1-\alpha}} \left(\alpha e^{z_{t+1}} k_{t+1}^{\alpha-1} l_{t+1}^{1-\alpha} \right) \right\}$$

$$\frac{1}{(1 - \alpha\beta)\, e^{z_t} k_t^{\alpha} l^{1-\alpha}} = \beta E_t \left\{ \frac{\alpha}{(1 - \alpha\beta)\, k_{t+1}} \right\}$$

$$\frac{\alpha\beta}{(1 - \alpha\beta)} = \frac{\beta\alpha}{(1 - \alpha\beta)}$$

- 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:

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

- 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 Λ is a diagonal matrix with σ_1^2 and σ_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

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

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

$$H' = \begin{pmatrix} -\log \alpha\beta\lambda l^{1-\alpha} & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}, R = \Lambda$$

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:

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

- Croos-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.

- Fernández-Villaverde, Rubio-Ramírez, and Santos (2005).

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

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

and

$$\widehat{l}_t = R\widehat{k}_t + Sz_t$$

Writing the Likelihood Function I

- The transitions Equation:

$$\begin{pmatrix} 1 \\ \widehat{k}_{t+1} \\ z_{t+1} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & P & Q \\ 0 & 0 & \rho \end{pmatrix} \begin{pmatrix} 1 \\ \widehat{k}_{t} \\ z_{t} \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \epsilon_t .$$

- The Measurement Equation requires some care.

Writing the Likelihood Function II

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

- Therefore, using $\widehat{l}_t = R\widehat{k}_t + Sz_t$

$$\begin{gathered} \widehat{y}_t = z_t + \alpha \widehat{k}_t + (1 - \alpha)(R\widehat{k}_t + Sz_t) = \\ (\alpha + (1 - \alpha)R)\,\widehat{k}_t + (1 + (1 - \alpha)S)\,z_t \end{gathered}$$

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

$$\begin{gathered} \widehat{c}_t = z_t + \alpha \widehat{k}_t - \alpha_5(R\widehat{k}_t + Sz_t) = \\ (\alpha - \alpha_5 R)\,\widehat{k}_t + (1 - \alpha_5 S)\,z_t \end{gathered}$$

Writing the Likelihood Function III

Therefore the measurement equation is:

$$\begin{pmatrix} \log y_t \\ \log l_t \\ \log c_t \end{pmatrix} = \begin{pmatrix} \log y & \alpha + (1-\alpha)R & 1 + (1-\alpha)S \\ \log l & R & S \\ \log c & \alpha - \alpha_5 R & 1 - \alpha_5 S \end{pmatrix} \begin{pmatrix} 1 \\ \widehat{k}_t \\ z_t \end{pmatrix} + \begin{pmatrix} v_{1,t} \\ v_{2,t} \\ v_{3,t} \end{pmatrix}.$$

The Likelihood Function of a General Dynamic Equilibrium Economy

- Transition equation:

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

- Measurement equation:

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

- Interpretation.

Some Assumptions

1. We can partition $\{W_t\}$ into two independent sequences $\left\{W_{1,t}\right\}$ and $\left\{W_{2,t}\right\}$, s.t. $W_t = \left(W_{1,t}, W_{2,t}\right)$ and $\dim\left(W_{2,t}\right) + \dim\left(V_t\right) \geq \dim\left(Y_t\right)$.

2. We can always evaluate the conditional densities $p\left(y_t | W_1^t, y^{t-1}, S_0; \gamma\right)$. 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 γ:

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

- We factorize it as:

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

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

A Law of Large Numbers

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

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

...thus

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

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

Introducing Particles

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

- Each $s_0^{t-1,i}, w_1^{t-1,i}$ is a *particle* and $\left\{s_0^{t-1,i}, w_1^{t-1,i}\right\}_{i=1}^{N}$ a *swarm of particles*.

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

- Each $s_0^{t|t-1,i}, w_1^{t|t-1,i}$ is a *proposed particle* and $\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^{N}$ a *swarm of proposed particles*.

... and Weights

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

A Proposition

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

Importance of the Proposition

1. It shows how a draw $\left\{s_0^{t|t-1,i}, w_1^{t|t-1,i}\right\}_{i=1}^{N}$ from $p\left(W_1^t, S_0|y^{t-1};\gamma\right)$ can be used to draw $\left\{s_0^{t,i}, w_1^{t,i}\right\}_{i=1}^{N}$ from $p\left(W_1^t, S_0|y^t;\gamma\right)$.

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

Sequential Monte Carlo I: Filtering

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

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

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

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

Sequential Monte Carlo II: Likelihood

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

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

A “Trivial” Application

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

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

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

1. Let $s_0^{0,i} = s_0$ for all i.

2. Generate N i.i.d. draws $\left\{ s_0^{1|0,i}, w^{1|0,i} \right\}_{i=1}^{N}$ from $\mathcal{N}(0, \sigma)$.

3. Evaluate $p\left(y_1 | w_1^{1|0,i}, y^0, s_0^{1|0,i} \right) = p_{t(2)}\left(y_1 - \left(\alpha + \beta \frac{s_0^{1|0,i}}{1+s_0^{1|0,i}} + w^{1|0,i} \right) \right)$.

4. Evaluate the relative weights $q_1^i = \frac{p_{t(2)}\left(y_1 - \left(\alpha + \beta \frac{s_0^{1|0,i}}{1+s_0^{1|0,i}} + w^{1|0,i} \right) \right)}{\sum_{i=1}^{N} p_{t(2)}\left(y_1 - \left(\alpha + \beta \frac{s_0^{1|0,i}}{1+s_0^{1|0,i}} + w^{1|0,i} \right) \right)}$.

5. Resample with replacement N values of $\left\{ s_0^{1|0,i}, w^{1|0,i} \right\}_{i=1}^{N}$ with relative weights q_1^i. Call those sampled values $\left\{ s_0^{1,i}, w^{1,i} \right\}_{i=1}^{N}$.

6. Go to step 1, and iterate 1-4 until the end of the sample.

A Law of Large Numbers

A law of the large numbers delivers:

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

and consequently:

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

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).

- Simulation algorithms: Kitagawa (1987), Gordon, Salmond and Smith (1993), Mariano and Tanizaki (1995) and Geweke and Tanizaki (1999).

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 $U = E_0 \sum_{t=0}^{\infty} \beta^t \frac{\left(c_t^\theta (1-l_t)^{1-\theta}\right)^{1-\tau}}{1-\tau}$.

- 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:

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

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

- 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.

- We use a Finite Element Method. Why? Aruoba, Fernández-Villaverde and Rubio-Ramírez (2003):

  1. Speed: sparse system.

  2. Accuracy: flexible grid generation.

  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

$$
\begin{aligned}
k_t &= f_1(S_{t-1}, W_t; \gamma) = e^{\tanh^{-1}(\lambda_{t-1})} k_{t-1}^{\alpha} l\left(k_{t-1}, \tanh^{-1}(\lambda_{t-1}); \gamma\right)^{1-\alpha} * \\
&\left(1 - \frac{\theta}{1-\theta}(1-\alpha)\frac{\left(1 - l\left(k_{t-1}, \tanh^{-1}(\lambda_{t-1}); \gamma\right)\right)}{l\left(k_{t-1}, \tanh^{-1}(\lambda_{t-1}); \gamma\right)}\right) + (1-\delta)\, k_{t-1} \\
\lambda_t &= f_2(S_{t-1}, W_t; \gamma) = \tanh(\rho \tanh^{-1}(\lambda_{t-1}) + \epsilon_t) \\
gdp_t &= g_1(S_t, V_t; \gamma) = e^{\tanh^{-1}(\lambda_t)} k_t^{\alpha} l\left(k_t, \tanh^{-1}(\lambda_t); \gamma\right)^{1-\alpha} + V_{1,t} \\
hours_t &= g_2(S_t, V_t; \gamma) = l\left(k_t, \tanh^{-1}(\lambda_t); \gamma\right) + V_{2,t} \\
inv_t &= g_3(S_t, V_t; \gamma) = e^{\tanh^{-1}(\lambda_t)} k_t^{\alpha} l\left(k_t, \tanh^{-1}(\lambda_t); \gamma\right)^{1-\alpha} * \\
&\left(1 - \frac{\theta}{1-\theta}(1-\alpha)\frac{\left(1 - l\left(k_t, \tanh^{-1}(\lambda_t); \gamma\right)\right)}{l\left(k_t, \tanh^{-1}(\lambda_t); \gamma\right)}\right) + V_{3,t}
\end{aligned}
$$

Likelihood Function

Since our measurement equation implies that

$$p\left(y_t|S_t;\gamma\right) = (2\pi)^{-\frac{3}{2}}\left|\mathbf{\Sigma}\right|^{-\frac{1}{2}} e^{-\frac{\omega(S_t;\gamma)}{2}}$$

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

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

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

$$\simeq (2\pi)^{-\frac{3T}{2}}\left|\mathbf{\Sigma}\right|^{\frac{-T}{2}} \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} e^{-\frac{\omega(s_t^i;\gamma)}{2}}$$

Priors for the Parameters

Priors for the Parameters of the Model		
Parameters	Distribution	Hyperparameters
θ	Uniform	0,1
ρ	Uniform	0,1
τ	Uniform	0,100
α	Uniform	0,1
δ	Uniform	0,0.05
β	Uniform	0.75,1
σ_ϵ	Uniform	0,0.1
σ_1	Uniform	0,0.1
σ_2	Uniform	0,0.1
σ_3	Uniform	0,0.1

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

Parameter	θ	ρ	τ	α	δ
Value	0.357	0.95	2.0	0.4	0.02
Parameter	β	σ_ϵ	σ_1	σ_2	σ_3
Value	0.99	0.007	$1.58*10^{-4}$	0.0011	$8.66*10^{-4}$

- Sensitivity: $\tau = 50$ and $\sigma_\epsilon = 0.035$.

Figure 5.1: Likelihood Function Benchmark Calibration

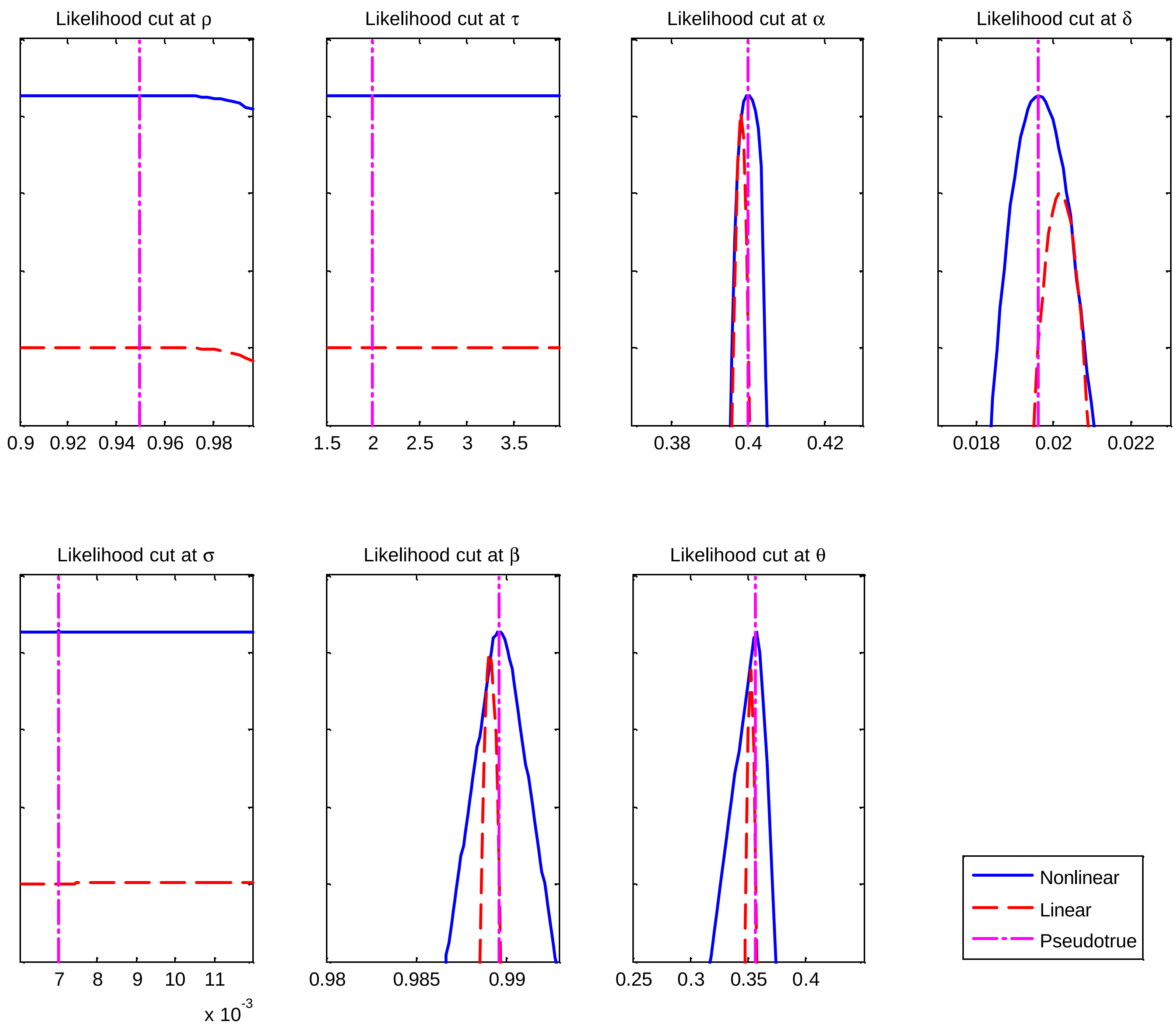


Figure 5.2: Posterior Distribution Benchmark Calibration

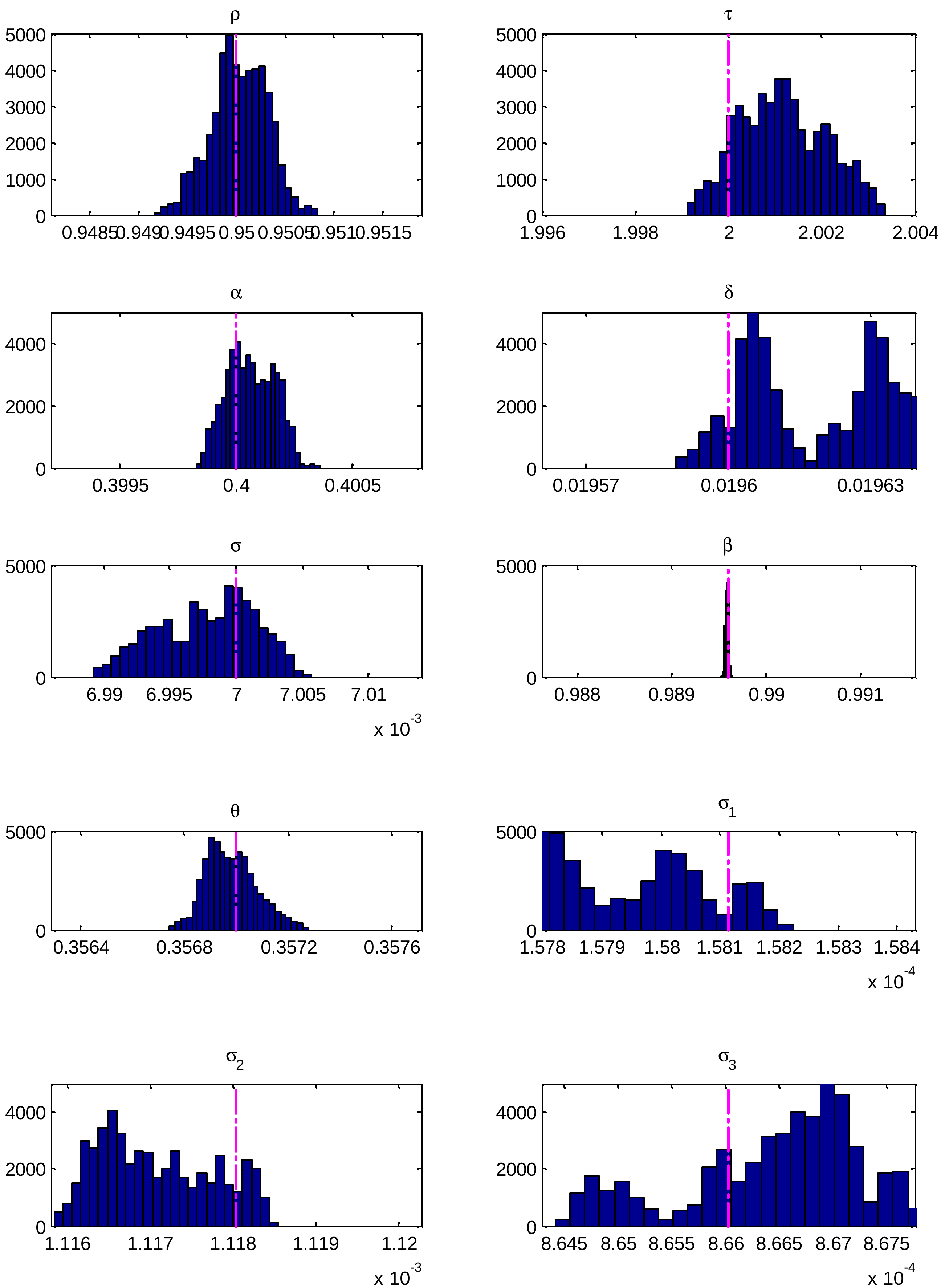


Figure 5.3: Likelihood Function Extreme Calibration

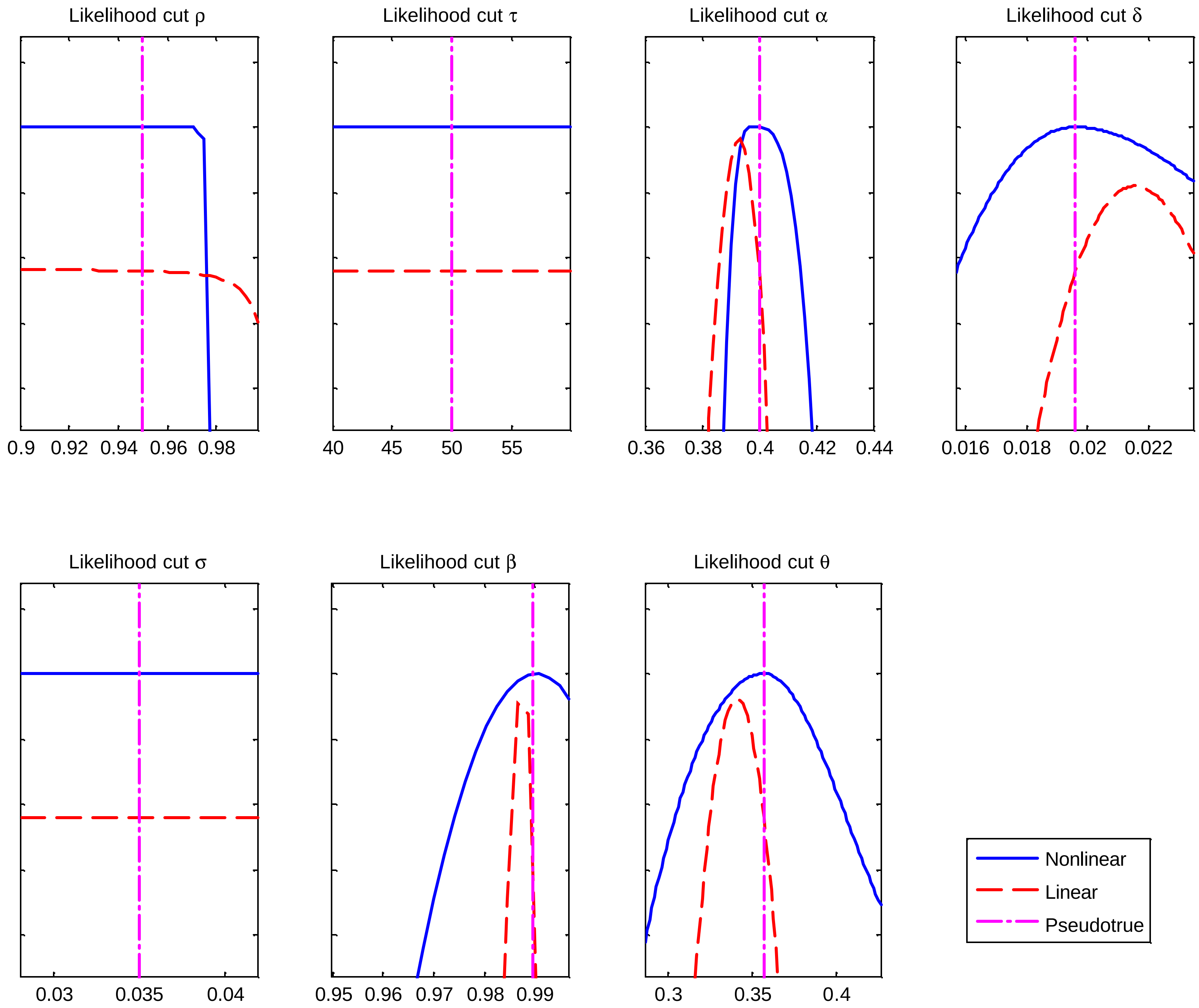


Figure 5.4: Posterior Distribution Extreme Calibration

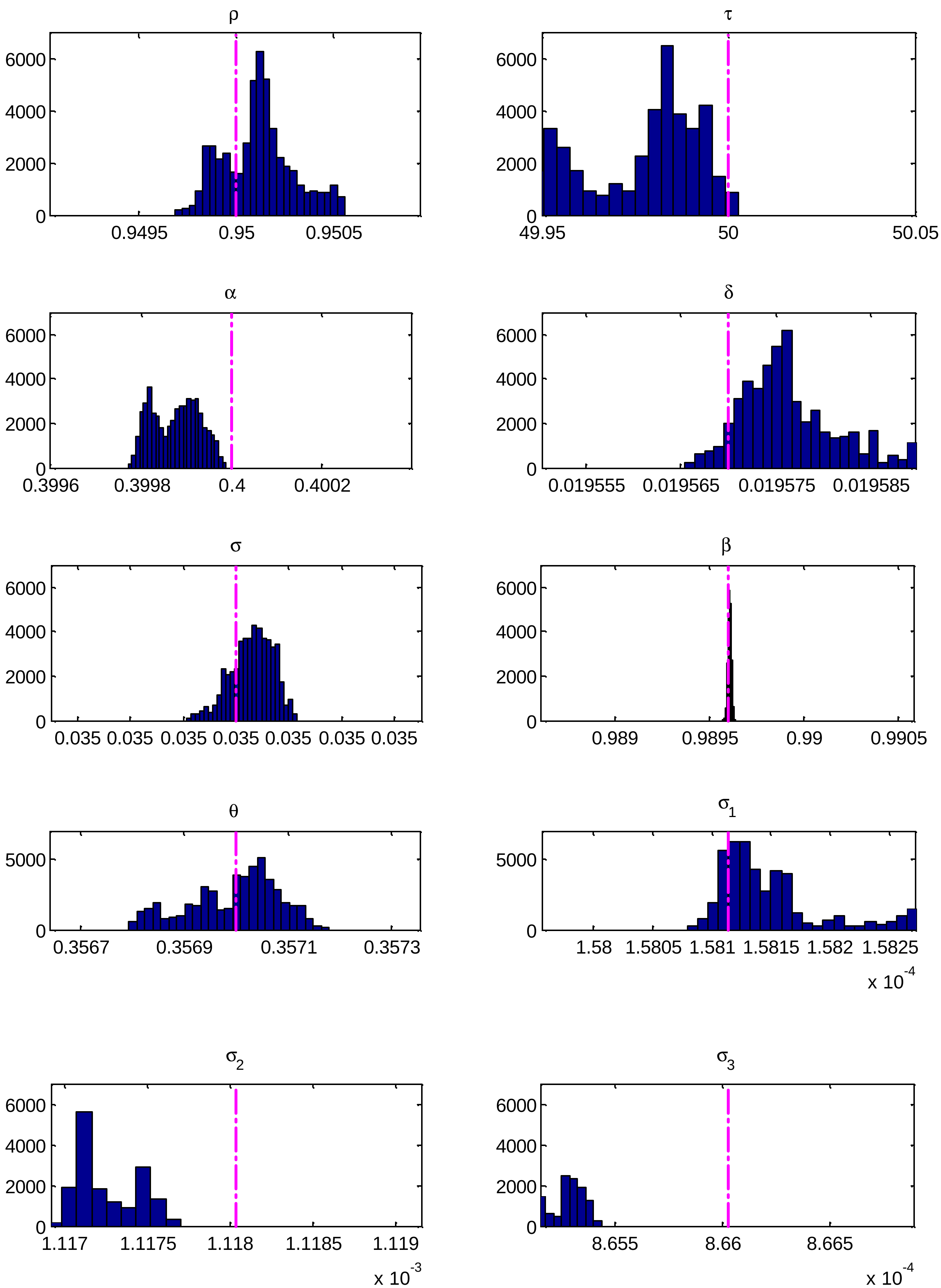


Figure 5.5: Converge of Posteriors Extreme Calibration

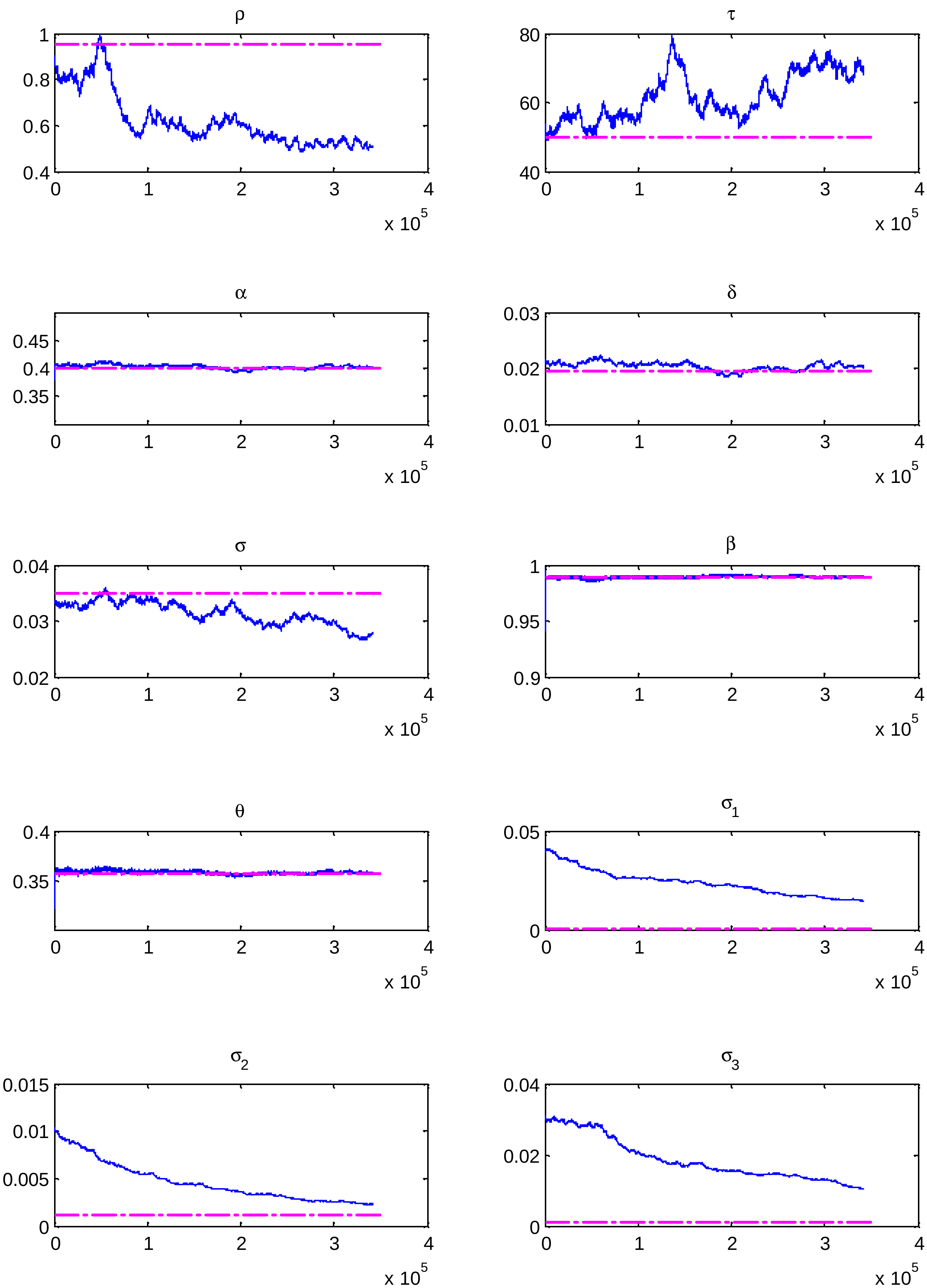


Figure 5.6: Posterior Distribution Real Data

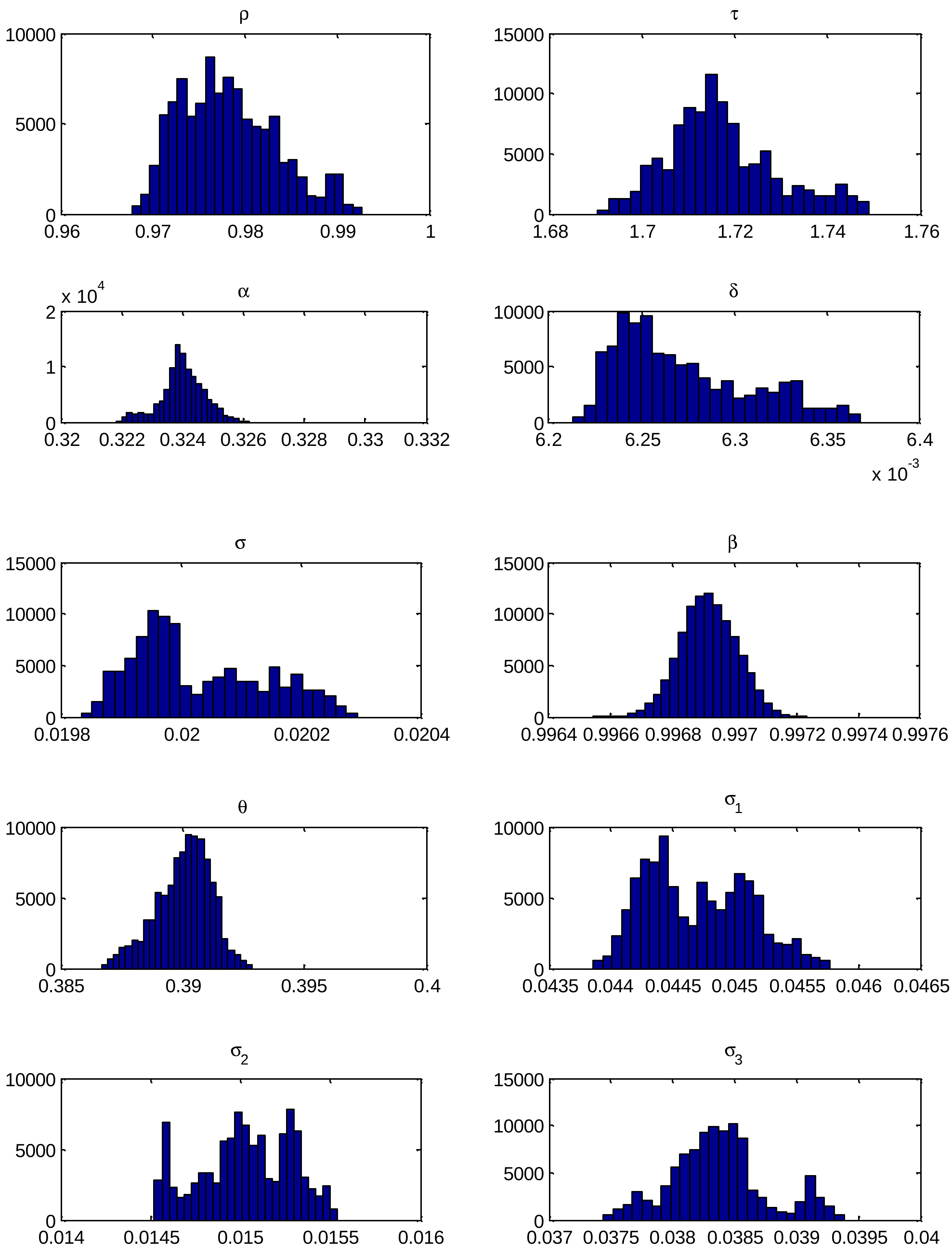


Figure 6.1: Likelihood Function

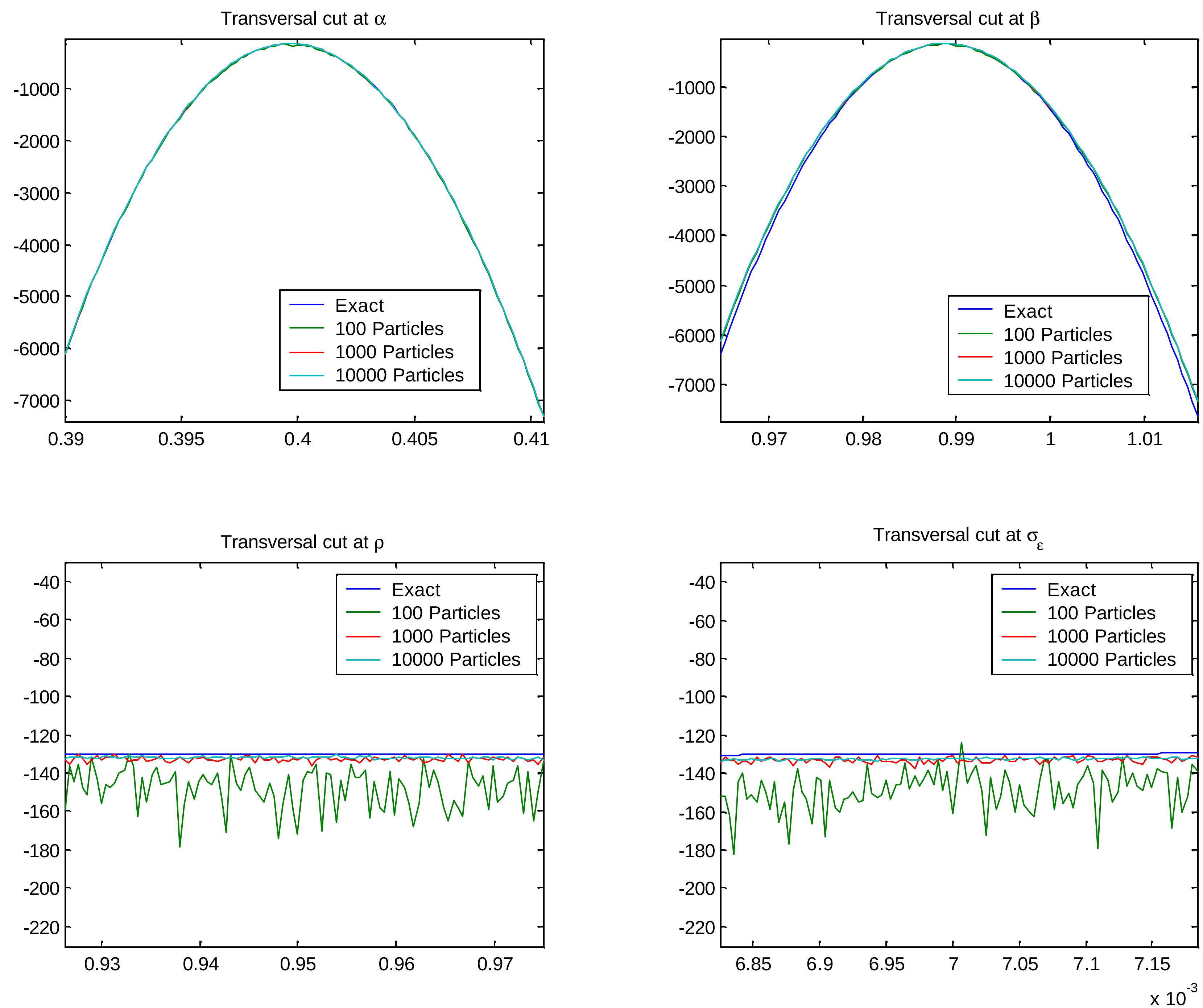


Figure 6.2: C.D.F. Benchmark Calibration

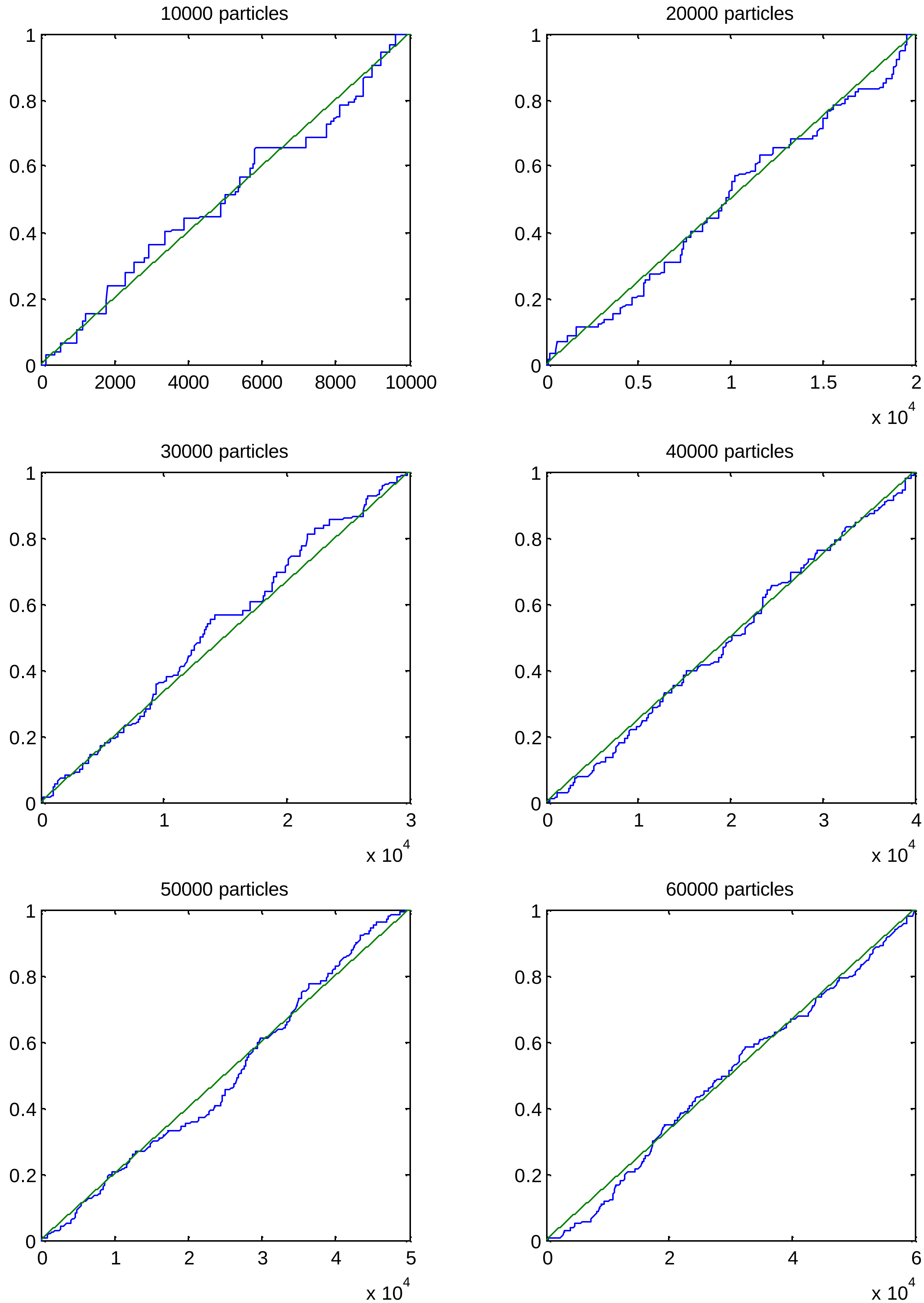


Figure 6.3: C.D.F. Extreme Calibration

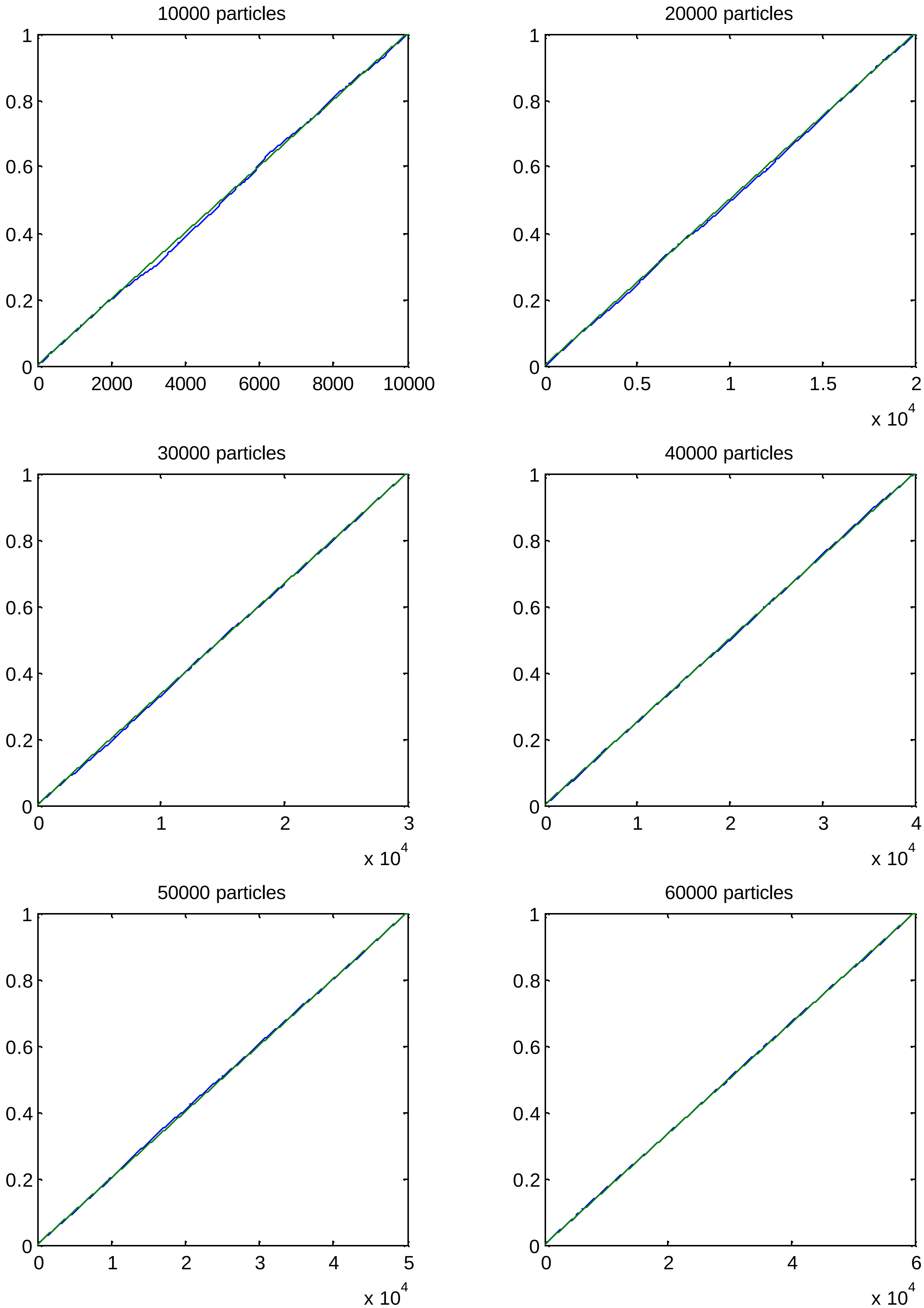
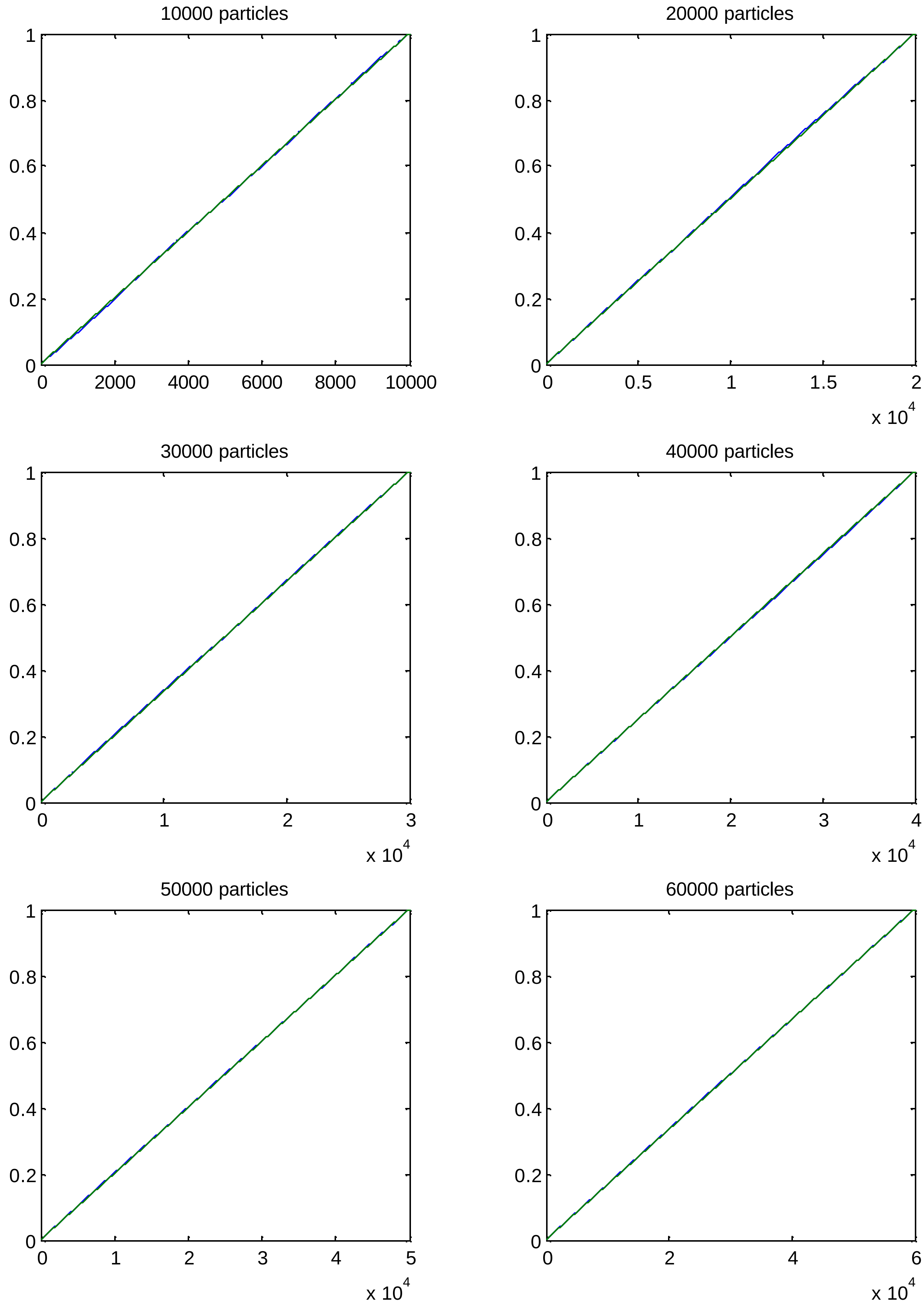


Figure 6.4: C.D.F. Real Data



Posterior Distributions Benchmark Calibration		
Parameters	Mean	s.d.
θ	0.357	6.72×10^{-5}
ρ	0.950	3.40×10^{-4}
τ	2.000	6.78×10^{-4}
α	0.400	8.60×10^{-5}
δ	0.020	1.34×10^{-5}
β	0.989	1.54×10^{-5}
σ_ϵ	0.007	9.29×10^{-6}
σ_1	1.58×10^{-4}	5.75×10^{-8}
σ_2	1.12×10^{-2}	6.44×10^{-7}
σ_3	8.64×10^{-4}	6.49×10^{-7}

Maximum Likelihood Estimates Benchmark Calibration		
Parameters	MLE	s.d.
θ	0.357	8.19×10^{-6}
ρ	0.950	0.001
τ	2.000	0.020
α	0.400	2.02×10^{-6}
δ	0.002	2.07×10^{-5}
β	0.990	1.00×10^{-6}
σ_ϵ	0.007	0.004
σ_1	1.58×10^{-4}	0.007
σ_2	1.12×10^{-3}	0.007
σ_3	8.63×10^{-4}	0.005

Posterior Distributions Extreme Calibration		
Parameters	Mean	s.d.
θ	0.357	7.19×10^{-4}
ρ	0.950	1.88×10^{-4}
τ	50.00	7.12×10^{-3}
α	0.400	4.80×10^{-5}
δ	0.020	3.52×10^{-6}
β	0.989	8.69×10^{-6}
σ_ϵ	0.035	4.47×10^{-6}
σ_1	1.58×10^{-4}	1.87×10^{-8}
σ_2	1.12×10^{-2}	2.14×10^{-7}
σ_3	8.65×10^{-4}	2.33×10^{-7}

Maximum Likelihood Estimates Extreme Calibration		
Parameters	MLE	s.d.
θ	0.357	2.42×10^{-6}
ρ	0.950	6.12×10^{-3}
τ	50.000	0.022
α	0.400	3.62×10^{-7}
δ	0.019	7.43×10^{-6}
β	0.990	1.00×10^{-5}
σ_ϵ	0.035	0.015
σ_1	1.58×10^{-4}	0.017
σ_2	1.12×10^{-3}	0.014
σ_3	8.66×10^{-4}	0.023

Convergence on Number of Particles

Convergence Real Data		
N	Mean	s.d.
10000	1014.558	0.3296
20000	1014.600	0.2595
30000	1014.653	0.1829
40000	1014.666	0.1604
50000	1014.688	0.1465
60000	1014.664	0.1347

Posterior Distributions Real Data		
Parameters	Mean	s.d.
θ	0.323	7.976×10^{-4}
ρ	0.969	0.008
τ	1.825	0.011
α	0.388	0.001
δ	0.006	3.557×10^{-5}
β	0.997	9.221×10^{-5}
σ_ϵ	0.023	2.702×10^{-4}
σ_1	0.039	5.346×10^{-4}
σ_2	0.018	4.723×10^{-4}
σ_3	0.034	6.300×10^{-4}

Maximum Likelihood Estimates Real Data		
Parameters	MLE	s.d.
θ	0.390	0.044
ρ	0.987	0.708
τ	1.781	1.398
α	0.324	0.019
δ	0.006	0.160
β	0.997	8.67×10^{-3}
σ_ϵ	0.023	0.224
σ_1	0.038	0.060
σ_2	0.016	0.061
σ_3	0.035	0.076

Logmarginal Likelihood Difference: Nonlinear-Linear

p	Benchmark Calibration	Extreme Calibration	Real Data
0.1	73.631	117.608	93.65
0.5	73.627	117.592	93.55
0.9	73.603	117.564	93.55

Nonlinear versus Linear Moments Real Data						
	Real Data		Nonlinear (SMC filter)		Linear (Kalman filter)	
	Mean	s.d	Mean	s.d	Mean	s.d
$output$	1.95	0.073	1.91	0.129	1.61	0.068
$hours$	0.36	0.014	0.36	0.023	0.34	0.004
inv	0.42	0.066	0.44	0.073	0.28	0.044

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?

  1. Good luck: Sims (1999), Bernanke and Mihov (1998a and 1998b) and Stock and Watson (2002).

  2. Good policy: Clarida, Gertler and Galí (2000), Cogley and Sargent (2001 and 2003), De Long (1997) and Romer and Romer (2002).

  3. Long run trend: Blanchard and Simon (2001).

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?

  3. Welfare analysis.

Why Not a Dynamic Equilibrium Model?

- New generation equilibrium models: Christiano, Eichebaum and Evans (2003) and Smets and Wouters (2003).

- 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 $\mu\left(s^t\right)$.

The Final Good Producer

- Perfectly Competitive Final Good Producer that solves

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

- Demand function for each input of the form

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

with price aggregator:

$$p \left(s^t\right) = \left(\int p_i \left(s^t\right)^{\frac{\theta}{\theta - 1}} di \right)^{\frac{\theta - 1}{\theta}}.$$

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:

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

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

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

Consumers Problem

$$\underset{s^t}{E} \sum_{t=0}^{\infty} \beta^t \left\{ \varepsilon_c \left(s^t\right) \frac{\left(c\left(s^t\right) - dc\left(s^{t-1}\right)\right)^{\sigma_c}}{\sigma_c} - \varepsilon_l \left(s^t\right) \frac{l\left(s^t\right)^{\sigma_l}}{\sigma_l} + \varepsilon_m \left(s^t\right) \frac{m\left(s^t\right)^{\sigma_m}}{\sigma_m} \right\}$$

$$p\left(s^t\right)\left(c\left(s^t\right) + x\left(s^t\right)\right) + M\left(s^t\right) + \int_{s^{t+1}} q\left(s^{t+1} \middle| s^t\right) B\left(s^{t+1}\right) ds_{t+1} =$$
$$p\left(s^t\right)\left(w\left(s^t\right) l\left(s^t\right) + r\left(s^t\right) k\left(s^{t-1}\right)\right) + M\left(s^{t-1}\right) + B\left(s^t\right) + \Pi\left(s^t\right) + T\left(s^t\right)$$
$$B\left(s^{t+1}\right) \geq B$$

$$k\left(s^t\right) = (1-\delta)\, k\left(s^{t-1}\right) - \phi\left(\frac{x\left(s^t\right)}{k\left(s^{t-1}\right)}\right) + x\left(s^t\right).$$

Government Policy

- Monetary Policy: Taylor rule

$$
\begin{aligned}
i\left(s^{t}\right) &= r_{g}\pi_{g}\left(s^{t}\right) \\
&\quad +a\left(s^{t}\right)\left(\pi\left(s^{t}\right)-\pi_{g}\left(s^{t}\right)\right) \\
&\quad +b\left(s^{t}\right)\left(y\left(s^{t}\right)-y_{g}\left(s^{t}\right)\right)+\varepsilon_{i}\left(s^{t}\right) \\
\pi_{g}\left(s^{t}\right) &= \pi_{g}\left(s^{t-1}\right)+\varepsilon_{\pi}\left(s^{t}\right) \\
a\left(s^{t}\right) &= a\left(s^{t-1}\right)+\varepsilon_{a}\left(s^{t}\right) \\
b\left(s^{t}\right) &= b\left(s^{t-1}\right)+\varepsilon_{b}\left(s^{t}\right)
\end{aligned}
$$

- Fiscal Policy.

Stochastic Volatility I

- We can stack all shocks in one vector:

$$\varepsilon\left(s^t\right)=\left(\varepsilon_z\left(s^t\right),\varepsilon_c\left(s^t\right),\varepsilon_l\left(s^t\right),\varepsilon_m\left(s^t\right),\varepsilon_i\left(s^t\right),\varepsilon_\pi\left(s^t\right),\varepsilon_a\left(s^t\right),\varepsilon_b\left(s^t\right)\right)'$$

- Stochastic volatility:

$$\varepsilon\left(s^t\right)=R\left(s^t\right)^{0.5}\vartheta\left(s^t\right).$$

- The matrix $R\left(s^t\right)$ can be decomposed as:

$$R\left(s^t\right)=G\left(s^t\right)^{-1}H\left(s^t\right)G\left(s^t\right).$$

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 $H\left(s^t\right)$, and the 28 elements of the matrix $G\left(s^t\right)$.

- Perturbation.

- We are making good progress.