Bayesian Estimation of DSGE Models[1]
Chapter 8: Particle Filters

Ed Herbst[1] Frank Schorfheide[2]

[1]Federal Reserve Board

[2]University of Pennsylvania

February 9, 2016

[1]The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Board of Governors or the Federal Reserve System.

Particle Filters

- There are many particle filters...
- We will focus on three types:
  - Bootstrap PF
  - A generic PF
  - A conditionally-optimal PF

Filtering - General Idea

- State-space representation of linearized DSGE model

$$\begin{aligned} y_t &= \Psi_0(\theta) + \Psi_1(\theta)t + \Psi_2(\theta)s_t(+u_t) \quad \text{measurement} \\ s_t &= \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t \quad \text{state transition} \end{aligned}$$

- Likelihood function:

$$p(Y_{1:T}|\theta) = \prod_{t=1}^{T} p(y_t|Y_{1:t-1}, \theta)$$

- A filter generates a sequence of conditional distributions $s_t|Y_{1:t}$.
- Iterations:
  - Initialization at time $t-1$: $p(s_{t-1}|Y_{1:t-1}, \theta)$
  - Forecasting t given $t-1$:
    1. Transition equation: $p(s_t|Y_{1:t-1}, \theta) = \int p(s_t|s_{t-1}, Y_{1:t-1}, \theta)p(s_{t-1}|Y_{1:t-1}, \theta)ds_{t-1}$
    2. Measurement equation: $p(y_t|Y_{1:t-1}, \theta) = \int p(y_t|s_t, Y_{1:t-1}, \theta)p(s_t|Y_{1:t-1}, \theta)ds_t$
  - Updating with Bayes theorem. Once y_t becomes available:

$$p(s_t|Y_{1:t}, \theta) = p(s_t|y_t, Y_{1:t-1}, \theta) = \frac{p(y_t|s_t, Y_{1:t-1}, \theta)p(s_t|Y_{1:t-1}, \theta)}{p(y_t|Y_{1:t-1}, \theta)}$$

Bootstrap Particle Filter

1. **Initialization.** Draw the initial particles from the distribution $s_0^j \stackrel{iid}{\sim} p(s_0)$ and set $W_0^j = 1$, $j = 1, \ldots, M$.
2. **Recursion.** For $t = 1, \ldots, T$:
    1. **Forecasting s_t.** Propagate the period $t-1$ particles $\{s_{t-1}^j, W_{t-1}^j\}$ by iterating the state-transition equation forward:

    $$\tilde{s}_t^j = \Phi(s_{t-1}^j, \epsilon_t^j; \theta), \quad \epsilon_t^j \sim F_\epsilon(\cdot; \theta). \tag{1}$$

    An approximation of $\mathbb{E}[h(s_t)|Y_{1:t-1}, \theta]$ is given by

    $$\hat{h}_{t,M} = \frac{1}{M} \sum_{j=1}^{M} h(\tilde{s}_t^j) W_{t-1}^j. \tag{2}$$

Bootstrap Particle Filter

1. **Initialization.**
2. **Recursion.** For $t = 1, \ldots, T$:
   1. **Forecasting s_t.**
   2. **Forecasting y_t.** Define the incremental weights

$$\tilde{w}_t^j = p(y_t|\tilde{s}_t^j, \theta). \tag{3}$$

The predictive density $p(y_t|Y_{1:t-1}, \theta)$ can be approximated by

$$\hat{p}(y_t|Y_{1:t-1}, \theta) = \frac{1}{M} \sum_{j=1}^{M} \tilde{w}_t^j W_{t-1}^j. \tag{4}$$

If the measurement errors are $N(0, \Sigma_u)$ then the incremental weights take the form

$$\tilde{w}_t^j = (2\pi)^{-n/2} |\Sigma_u|^{-1/2} \exp\left\{ -\frac{1}{2} \big(y_t - \Psi(\tilde{s}_t^j, t; \theta)\big)' \Sigma_u^{-1} \big(y_t - \Psi(\tilde{s}_t^j, t; \theta)\big) \right\}, \tag{5}$$

where n here denotes the dimension of y_t.

Bootstrap Particle Filter

1. **Initialization.**
2. **Recursion.** For $t = 1, \ldots, T$:
   1. **Forecasting s_t.**
   2. **Forecasting y_t.** Define the incremental weights

$$\tilde{w}_t^j = p(y_t|\tilde{s}_t^j, \theta). \tag{6}$$

   3. **Updating.** Define the normalized weights

$$\tilde{W}_t^j = \frac{\tilde{w}_t^j W_{t-1}^j}{\frac{1}{M}\sum_{j=1}^M \tilde{w}_t^j W_{t-1}^j}. \tag{7}$$

An approximation of $\mathbb{E}[h(s_t)|Y_{1:t}, \theta]$ is given by

$$\tilde{h}_{t,M} = \frac{1}{M}\sum_{j=1}^M h(\tilde{s}_t^j)\tilde{W}_t^j. \tag{8}$$

Bootstrap Particle Filter

1. **Initialization.**
2. **Recursion.** For $t = 1, \ldots, T$:
   1. **Forecasting s_t.**
   2. **Forecasting y_t.**
   3. **Updating.**
   4. **Selection (Optional).** Resample the particles via multinomial resampling. Let $\{s_t^j\}_{j=1}^M$ denote M iid draws from a multinomial distribution characterized by support points and weights $\{\tilde{s}_t^j, \tilde{W}_t^j\}$ and set $W_t^j = 1$ for $j =, 1 \ldots, M$.

      An approximation of $\mathbb{E}[h(s_t)|Y_{1:t}, \theta]$ is given by

$$\bar{h}_{t,M} = \frac{1}{M} \sum_{j=1}^{M} h(s_t^j) W_t^j. \tag{9}$$

3. **Likelihood Approximation.** The approximation of the log likelihood function is given by

$$\ln \hat{p}(Y_{1:T}|\theta) = \sum_{t=1}^{T} \ln \left(\frac{1}{M} \sum_{j=1}^{M} \tilde{w}_t^j W_{t-1}^j \right). \tag{10}$$

The Role of Measurement Errors

- Measurement errors may not be intrinsic to DSGE model.
- Bootstrap filter needs non-degenerate $p(y_t|s_t, \theta)$ for incremental weights to be well defined.
- Decreasing the measurement error variance Σ_u, holding everything else fixed, increases the variance of the particle weights, and reduces the accuracy of Monte Carlo approximation.

Generic Particle Filter

1. **Initialization.** Same as BS PF
2. **Recursion.** For $t = 1, \dots, T$:
   1. **Forecasting s_t.** Draw $\tilde{s}_t^j$ from density $g_t(\tilde{s}_t | s_{t-1}^j, \theta)$ and define
   $$\omega_t^j = \frac{p(\tilde{s}_t^j | s_{t-1}^j, \theta)}{g_t(\tilde{s}_t^j | s_{t-1}^j, \theta)}. \tag{11}$$
   An approximation of $\mathbb{E}[h(s_t) | Y_{1:t-1}, \theta]$ is given by
   $$\hat{h}_{t,M} = \frac{1}{M} \sum_{j=1}^{M} h(\tilde{s}_t^j) \omega_t^j W_{t-1}^j. \tag{12}$$
   2. **Forecasting y_t.** Define the incremental weights
   $$\tilde{w}_t^j = p(y_t | \tilde{s}_t^j, \theta) \omega_t^j. \tag{13}$$
   The predictive density $p(y_t | Y_{1:t-1}, \theta)$ can be approximated by
   $$\hat{p}(y_t | Y_{1:t-1}, \theta) = \frac{1}{M} \sum_{j=1}^{M} \tilde{w}_t^j W_{t-1}^j. \tag{14}$$
   3. **Updating.** Same as BS PF
   4. **Selection.** Same as BS PF
3. **Likelihood Approximation.** Same as BS PF

- Conditionally-optimal importance distribution:

$$g_t(\tilde{s}_t|s^j_{t-1}) = p(\tilde{s}_t|y_t, s^j_{t-1}).$$

  This is the posterior of s_t given s^j_{t-1}. Typically infeasible, but a good benchmark.

- Approximately conditionally-optimal distributions: from linearize version of DSGE model or approximate nonlinear filters.

- Conditionally-linear models: do Kalman filter updating on a subvector of s_t. Example:

$$\begin{aligned} y_t &= \Psi_0(m_t) + \Psi_1(m_t)t + \Psi_2(m_t)s_t + u_t, \quad u_t \sim N(0, \Sigma_u), \\ s_t &= \Phi_0(m_t) + \Phi_1(m_t)s_{t-1} + \Phi_\epsilon(m_t)\epsilon_t, \quad \epsilon_t \sim N(0, \Sigma_\epsilon), \end{aligned}$$

  where m_t follows a discrete Markov-switching process.

Nonlinear and Partially Deterministic State Transitions

- Example:

$$s_{1,t} = \Phi_1(s_{t-1}, \epsilon_t), \quad s_{2,t} = \Phi_2(s_{t-1}), \quad \epsilon_t \sim N(0,1).$$

- Generic filter requires evaluation of $p(s_t|s_{t-1})$.
- Define $\varsigma_t = [s_t', \epsilon_t']'$ and add identity $\epsilon_t = \epsilon_t$ to state transition.
- Factorize the density $p(\varsigma_t|\varsigma_{t-1})$ as

$$p(\varsigma_t|\varsigma_{t-1}) = p^{\epsilon}(\epsilon_t)p(s_{1,t}|s_{t-1}, \epsilon_t)p(s_{2,t}|s_{t-1}).$$

where $p(s_{1,t}|s_{t-1}, \epsilon_t)$ and $p(s_{2,t}|s_{t-1})$ are pointmasses.

- Sample innovation ϵ_t from $g_t^{\epsilon}(\epsilon_t|s_{t-1})$.
- Then

$$\omega_t^j = \frac{p(\tilde{\varsigma}_t^j|\varsigma_{t-1}^j)}{g_t(\tilde{\varsigma}_t^j|\varsigma_{t-1}^j)} = \frac{p^{\epsilon}(\tilde{\epsilon}_t^j)p(\tilde{s}_{1,t}^j|s_{t-1}^j, \tilde{\epsilon}_t^j)p(\tilde{s}_{2,t}^j|s_{t-1}^j)}{g_t^{\epsilon}(\tilde{\epsilon}_t^j|s_{t-1}^j)p(\tilde{s}_{1,t}^j|s_{t-1}^j, \tilde{\epsilon}_t^j)p(\tilde{s}_{2,t}^j|s_{t-1}^j)} = \frac{p^{\epsilon}(\tilde{\epsilon}_t^j)}{g_t^{\epsilon}(\tilde{\epsilon}_t^j|s_{t-1}^j)}.$$

Degenerate Measurement Error Distributions

- Our discussion of the conditionally-optimal importance distribution suggests that in the absence of measurement errors, one has to solve the system of equations

$$y_t = \Psi\big(\Phi(s^j_{t-1}, \tilde{\epsilon}^j_t)\big),$$

to determine $\tilde{\epsilon}^j_t$ as a function of s^j_{t-1} and the current observation y_t.

- Then define

$$\omega^j_t = p^{\epsilon}(\tilde{\epsilon}^j_t) \quad \text{and} \quad \tilde{s}^j_t = \Phi(s^j_{t-1}, \tilde{\epsilon}^j_t).$$

- Difficulty: one has to find all solutions to a nonlinear system of equations.

- While resampling duplicates particles, the duplicated particles do not mutate, which can lead to a degeneracy.

Next Steps

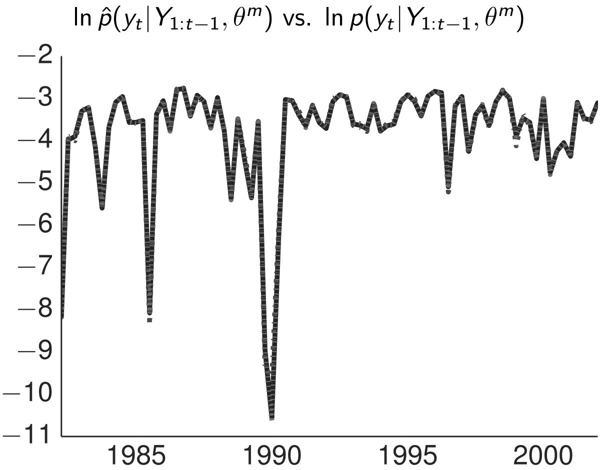
- We will now apply PFs to linearized DSGE models.
- This allows us to compare the Monte Carlo approximation to the "truth."
- Small-scale New Keynesian DSGE model
- Smets-Wouters model

Illustration 1: Small-Scale DSGE Model

Parameter Values For Likelihood Evaluation

Parameter	θ^m	θ^l	Parameter	θ^m	θ^l
τ	2.09	3.26	κ	0.98	0.89
ψ_1	2.25	1.88	ψ_2	0.65	0.53
ρ_r	0.81	0.76	ρ_g	0.98	0.98
ρ_z	0.93	0.89	$r^{(A)}$	0.34	0.19
$\pi^{(A)}$	3.16	3.29	$\gamma^{(Q)}$	0.51	0.73
σ_r	0.19	0.20	σ_g	0.65	0.58
σ_z	0.24	0.29	$\ln p(Y\|\theta)$	-306.5	-313.4

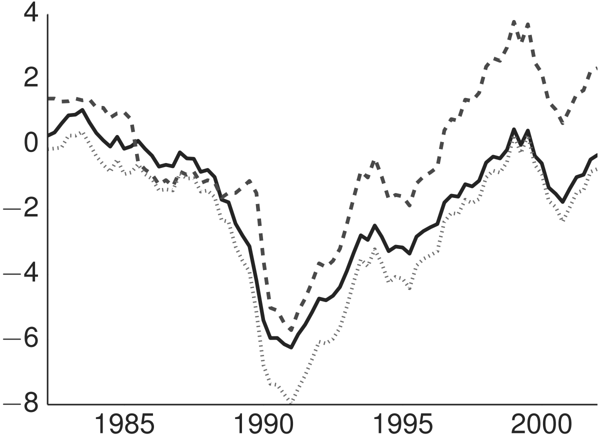
Likelihood Approximation



Notes: The results depicted in the figure are based on a single run of the bootstrap PF (dashed), the conditionally-optimal PF (dotted), and the Kalman filter (solid).

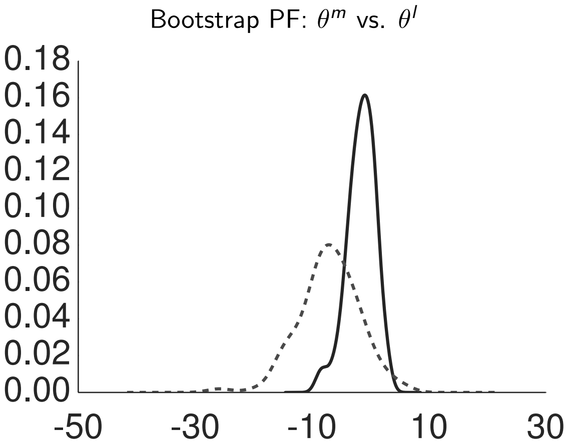
Filtered State

$\widehat{\mathbb{E}}[\hat{g}_t | Y_{1:t}, \theta^m]$ vs. $\mathbb{E}[\hat{g}_t | Y_{1:t}, \theta^m]$



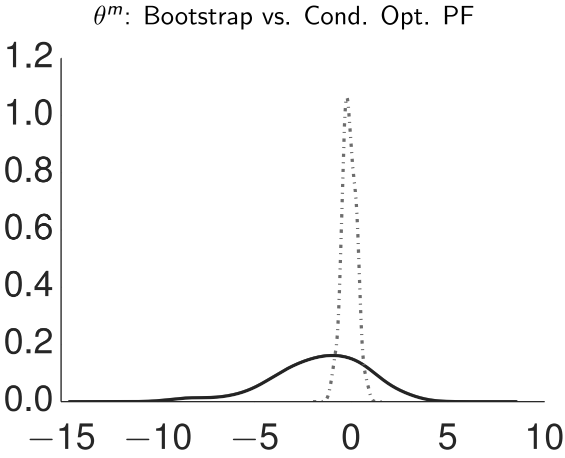
Notes: The results depicted in the figure are based on a single run of the bootstrap PF (dashed), the conditionally-optimal PF (dotted), and the Kalman filter (solid).

Distribution of Log-Likelihood Approximation Errors



Notes: Density estimate of $\hat{\Delta}_1 = \ln \hat{p}(Y_{1:T}|\theta) - \ln p(Y_{1:T}|\theta)$ based on $N_{run} = 100$ runs of the PF. Solid line is $\theta = \theta^m$; dashed line is $\theta = \theta^l$ ($M = 40,000$).

Distribution of Log-Likelihood Approximation Errors



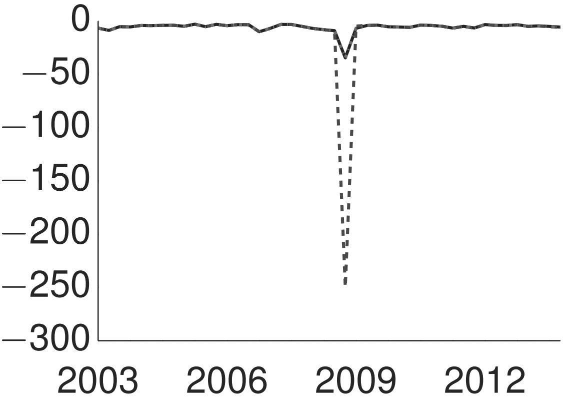
Notes: Density estimate of $\hat{\Delta}_1 = \ln \hat{p}(Y_{1:T}|\theta) - \ln p(Y_{1:T}|\theta)$ based on $N_{run} = 100$ runs of the PF. Solid line is bootstrap particle filter $(M = 40,000)$; dotted line is conditionally optimal particle filter $(M = 400)$.

Summary Statistics for Particle Filters

	Bootstrap	Cond. Opt.	Auxiliary
Number of Particles M	40,000	400	40,000
Number of Repetitions	100	100	100
High Posterior Density: $\theta = \theta^m$			
Bias $\hat{\Delta}_1$	-1.39	-0.10	-2.83
StdD $\hat{\Delta}_1$	2.03	0.37	1.87
Bias $\hat{\Delta}_2$	0.32	-0.03	-0.74
Low Posterior Density: $\theta = \theta^l$			
Bias $\hat{\Delta}_1$	-7.01	-0.11	-6.44
StdD $\hat{\Delta}_1$	4.68	0.44	4.19
Bias $\hat{\Delta}_2$	-0.70	-0.02	-0.50

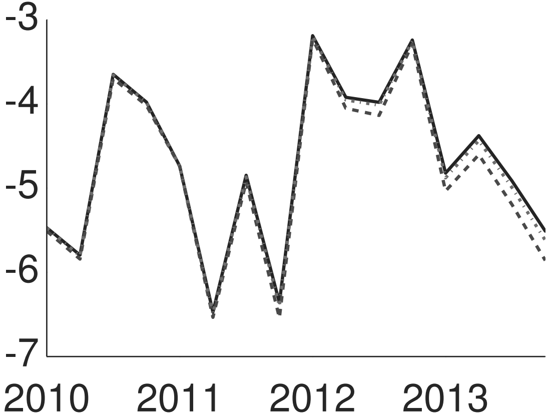
Notes: $\hat{\Delta}_1 = \ln \hat{p}(Y_{1:T}|\theta) - \ln p(Y_{1:T}|\theta)$ and $\hat{\Delta}_2 = \exp[\ln \hat{p}(Y_{1:T}|\theta) - \ln p(Y_{1:T}|\theta)] - 1$. Results are based on $N_{run} = 100$ runs of the particle filters.

Mean of Log-likelihood Increments $\ln \hat{p}(y_t|Y_{1:t-1}, \theta^m)$

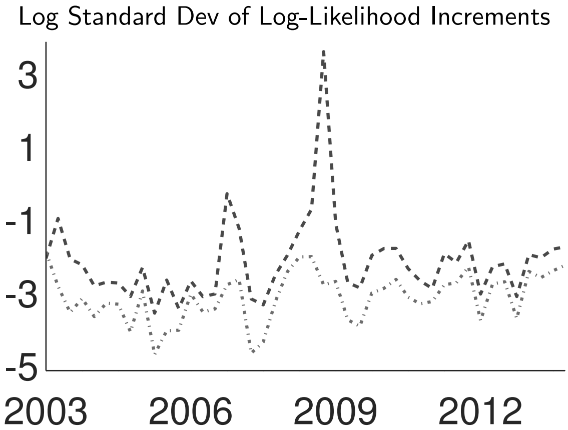


Notes: Solid lines represent results from Kalman filter. Dashed lines correspond to bootstrap particle filter ($M = 40,000$) and dotted lines correspond to conditionally-optimal particle filter ($M = 400$). Results are based on $N_{run} = 100$ runs of the filters.

Mean of Log-likelihood Increments $\ln \hat{p}(y_t | Y_{1:t-1}, \theta^m)$

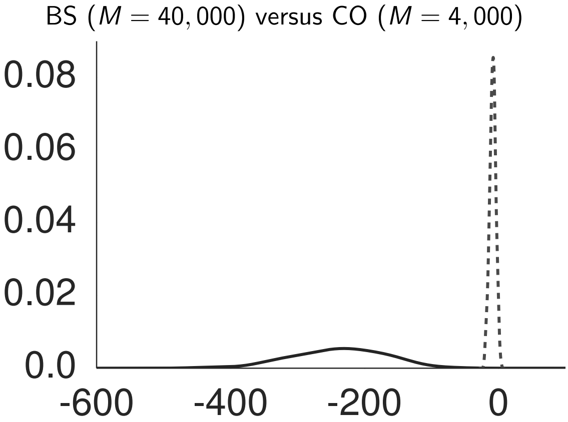


Notes: Solid lines represent results from Kalman filter. Dashed lines correspond to bootstrap particle filter ($M = 40,000$) and dotted lines correspond to conditionally-optimal particle filter ($M = 400$). Results are based on $N_{run} = 100$ runs of the filters.



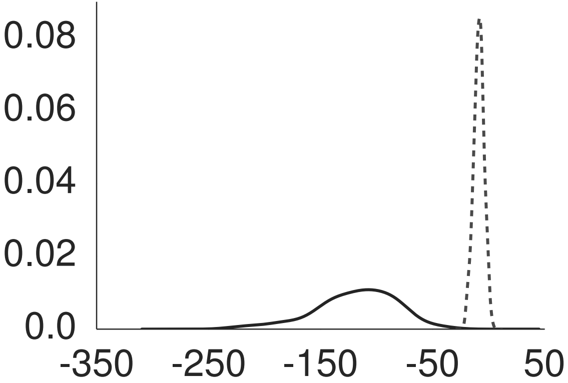
Notes: Solid lines represent results from Kalman filter. Dashed lines correspond to bootstrap particle filter ($M = 40,000$) and dotted lines correspond to conditionally-optimal particle filter ($M = 400$). Results are based on $N_{run} = 100$ runs of the filters.

SW Model: Distr. of Log-Likelihood Approximation Errors



Notes: Density estimates of $\hat{\Delta}_1 = \ln \hat{p}(Y|\theta) - \ln p(Y|\theta)$ based on $N_{run} = 100$. Solid densities summarize results for the bootstrap (BS) particle filter; dashed densities summarize results for the conditionally-optimal (CO) particle filter.

BS ($M = 400,000$) versus CO ($M = 4,000$)



Notes: Density estimates of $\hat{\Delta}_1 = \ln \hat{p}(Y|\theta) - \ln p(Y|\theta)$ based on $N_{run} = 100$. Solid densities summarize results for the bootstrap (BS) particle filter; dashed densities summarize results for the conditionally-optimal (CO) particle filter.

SW Model: Summary Statistics for Particle Filters

	Bootstrap		Cond. Opt.	
Number of Particles M	40,000	400,000	4,000	40,000
Number of Repetitions	100	100	100	100
High Posterior Density: $\theta = \theta^m$				
Bias $\hat{\Delta}_1$	-238.49	-118.20	-8.55	-2.88
StdD $\hat{\Delta}_1$	68.28	35.69	4.43	2.49
Bias $\hat{\Delta}_2$	-1.00	-1.00	-0.87	-0.41
Low Posterior Density: $\theta = \theta^l$				
Bias $\hat{\Delta}_1$	-253.89	-128.13	-11.48	-4.91
StdD $\hat{\Delta}_1$	65.57	41.25	4.98	2.75
Bias $\hat{\Delta}_2$	-1.00	-1.00	-0.97	-0.64

Notes: $\hat{\Delta}_1 = \ln \hat{p}(Y_{1:T}|\theta) - \ln p(Y_{1:T}|\theta)$ and $\hat{\Delta}_2 = \exp[\ln \hat{p}(Y_{1:T}|\theta) - \ln p(Y_{1:T}|\theta)] - 1$. Results are based on $N_{run} = 100$.