Trend-Cycle Decompositions of Real GDP Revisited: Classical and Bayesian Perspectives on an Unsolved Puzzle

June. 28. 2016

Macro Forum
Korea Development Institute

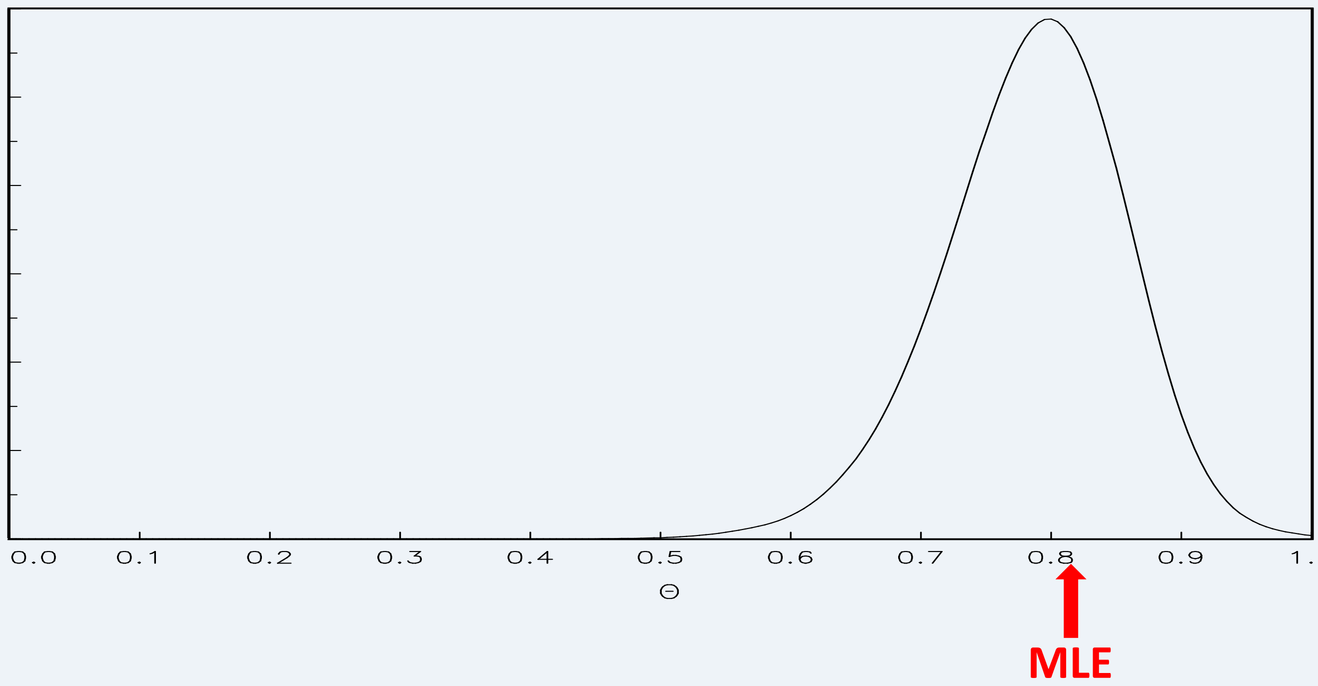
Chang-Jin Kim
University of Washington

Jaeho Kim
University of Oklahoma

I.1. Conventional Wisdom in the Case of a Flat Prior

- Bayesian inference may not be very different from classical inference as the likelihood dominates the posterior density.

- The posterior density would be centered on the ML estimate.



Purposes of this Paper

1. **We show that the conventional wisdom does not apply to the empirical model of real GDP.**

ARIMA model of real GDP (Perron and Wada (JME, 2009))

Result from ML Approach: "Real GDP is a Trend Stationary Process (TSP)"
Result from Bayesian Approach: "Real GDP is a Difference Stationary Process(DSP)"

2. **We show that the Maximum Likelihood approach suffers severely from the 'pile-up' problem, while the Bayesian approach is relatively free from it. We explain why the two approaches produce different results.**
   >> Difference in how to handle the nuisance parameters

3. **We employ a new multi-move Markov-Chain Monte Carlo (MCMC) algorithm proposed by Kim and Kim (JBES, 2015) to estimate a structural break ARIMA model of real GDP.**

4. **We provide convincing evidence that the cycle measure from a DSP model has out-of-sample predictive power for future output growth at short horizons.**

I.2. Pile-up Problem

MA (1) without Intercept

$$\Delta y_t = e_t - \theta e_{t-1}, \qquad e_t \sim i.i.d.N\ (0, \sigma^2),$$

where the first-order autocorrelation (ρ_1) is given by:

$$\rho_1 = -\frac{\theta}{(1+\theta^2)}$$

- It can be shown that two parameter set, i.e., (θ, σ^2) and $(\frac{1}{\theta}, \sigma^2)$, induce an identical autocorrelation structure and thus an identical log likelihood value.

- This **identification problem** can be handled by restricting the parameter space to $|\theta| \leq 1$, including an 'invertibility region' and unity.

I.2. Pile-up Problem

MA (1) without Intercept

$$\Delta y_t = e_t - \theta e_{t-1}, \qquad e_t \sim i.i.d.N\ (0, \sigma^2),$$

where the first-order autocorrelation (ρ_1) is given by:

$$\rho_1 = -\frac{\theta}{(1+\theta^2)}$$

- Kang (1975), Sargan and Bhargava (1983), Anderson and Takemura (1986), Tanaka and Satchell (1989), Davis and Dunsmuir (1996), Davis et al. (1995)

I.3. Interpretation of MA Coefficients

	Model with y_t	*Model with Δy_t*
True DGP (Pure Random Walk)	$y_t = x_t,$ $x_t = x_{t-1} + v_t$	$\Delta y_t = v_t$
Estimation Model		$\Delta y_t = e_t - \theta e_{t-1}$

- When θ is estimated to be 0, it can be interpreted as strong evidence in favor of a permanent trend component.

I.3. Interpretation of MA Coefficients

	Model with y_t	*Model with Δy_t*
True DGP (Pure Stationary Process)	$y_t = z_t,$ $z_t = \phi z_{t-1} + \varepsilon_t$	$(1-\phi L)\Delta y_t = \varepsilon_t - \varepsilon_{t-1}$
Estimation Model		$(1-\phi L)\Delta y_t = e_t - \theta e_{t-1}$

- When θ is estimated to be 1, it can be interpreted as strong evidence in favor of a transitory cyclical component.

I.4. Pile-up Problem with Nuisance Parameters

MA (1) with Intercept

$$\Delta y_t = \mu + e_t - \theta e_{t-1}, \; e_t \sim i.i.d \; N(0, \sigma^2)$$

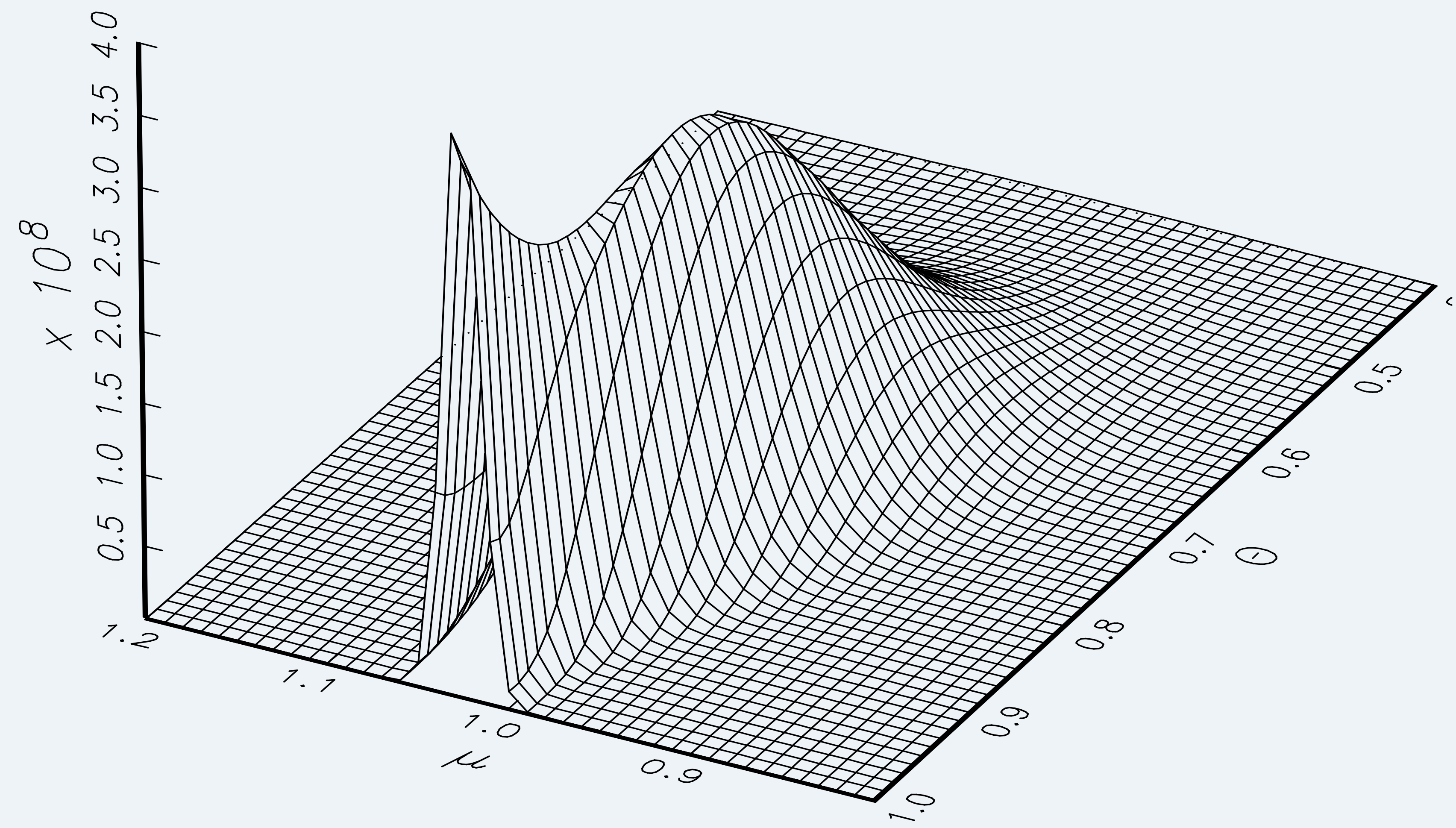
$$t = 1,2, \dots, T,$$

$$[\theta = 0.8, \sigma^2 = 1, \mu = 1]$$

- When θ is of our central interest, μ is treated as a nuisance parameter.

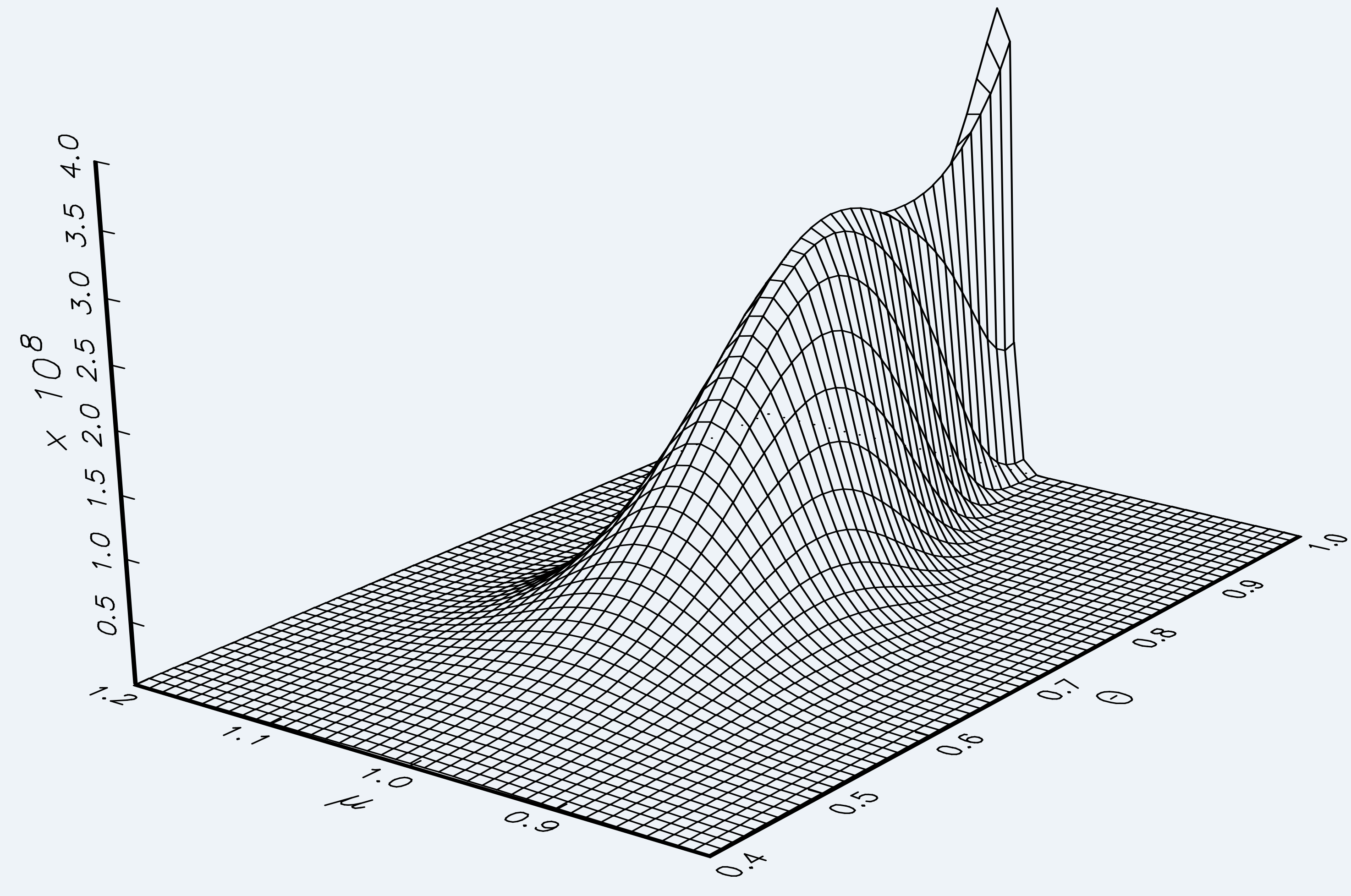
I.4. Pile-up Problem with Nuisance Parameters

<u>Likelihood Surface of a Representative Sample</u>



I.4. Pile-up Problem with Nuisance Parameters

Likelihood Surface of a Representative Sample



I.4. Pile-up Problem with Nuisance Parameters

Classical MLE (Profile Likelihood)

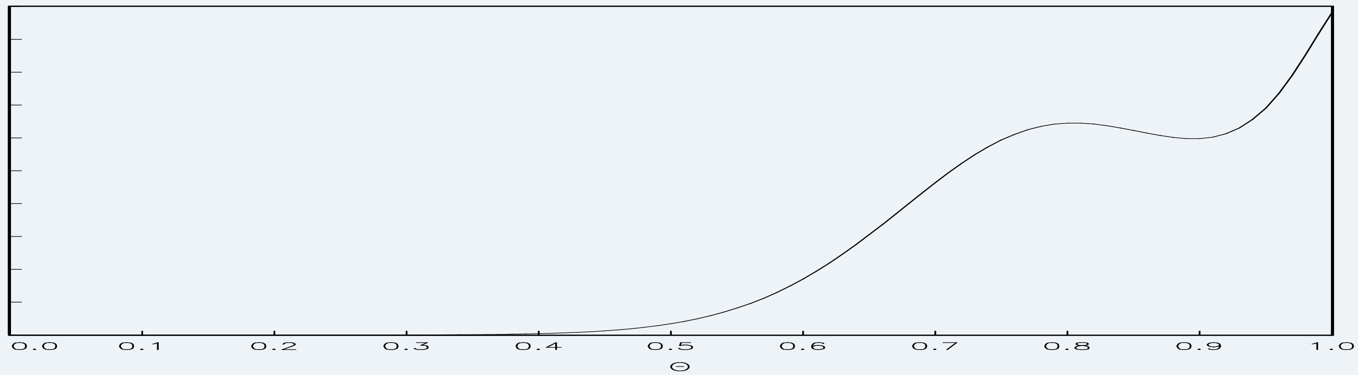
$$\hat{L}(\theta) = sup_{\mu} L(\theta, \mu)$$

Posterior Density (Integrated Likelihood)

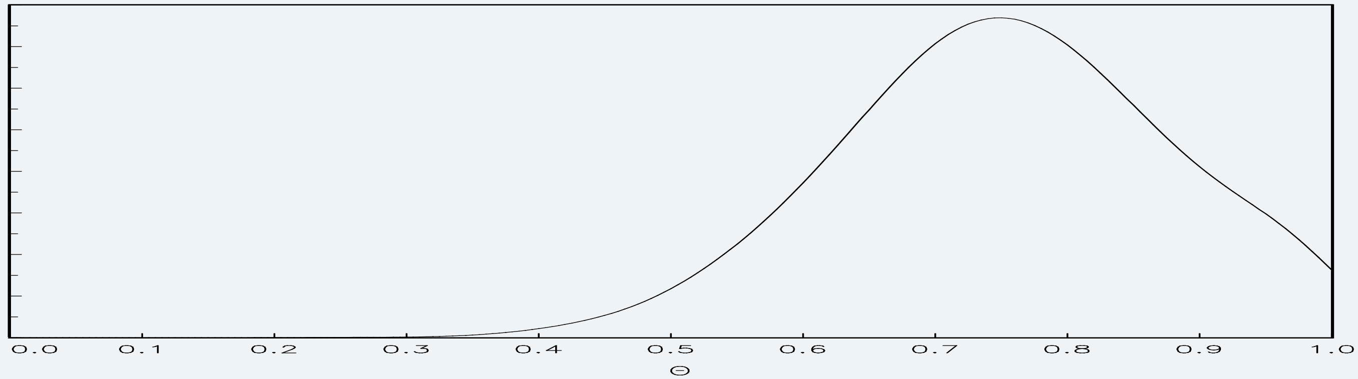
$$L(\theta) = \int L(\theta, \mu)\, d\mu$$

I.4. Pile-up Problem with Nuisance Parameters

Comparison of Profile likelihood and Posterior Distribution for MA coefficient: Representative Sample



$$\hat{L}(\theta) = sup_{\mu} L(\theta, \mu)$$



$$L(\theta) = \int L(\theta, \mu)\, d\mu$$

Purposes of this Paper

1. **We show that the conventional wisdom does not apply to the empirical model of real GDP.**

   ARIMA model of real GDP (Perron and Wada (JME, 2009))

   Result from ML Approach: "Real GDP is a Trend Stationary Process (TSP)"
   Result from Bayesian Approach: "Real GDP is a Difference Stationary Process(DSP)"

2. **We show that the Maximum Likelihood approach suffers severely from the 'pile-up' problem, while the Bayesian approach is relatively free from it. We explain why the two approaches produce different results.**
   >> Difference in how to handle the nuisance parameters

3. **We employ a new multi-move Markov-Chain Monte Carlo (MCMC) algorithm proposed by Kim and Kim (JBES, 2015) to estimate a structural break ARIMA model of real GDP.**

4. **We provide convincing evidence that the cycle measure from a DSP model has out-of-sample predictive power for future output growth at short horizons.**

II. Trend-Cycle Decomposition: Classical and Bayesian Perspectives

Unobserved Components Model of Real GDP (Structural Model)

$$y_t = x_t + z_t,$$

$$x_t = \mu_t + x_{t-1} + \upsilon_t,$$

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-1} + \varepsilon_t - \theta_0 \varepsilon_{t-1},$$

$$\begin{bmatrix} \upsilon_t \\ \varepsilon_t \end{bmatrix} \sim i.i.d.N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\upsilon^2 & \rho\sigma_\nu\sigma_\varepsilon \\ \rho\sigma_\varepsilon\sigma_\nu & \sigma_\varepsilon^2 \end{bmatrix} \right).$$

- y_t is the log of real GDP; x_t is a stochastic trend component; z_t is a cyclical component; μ_t is the time-varying long-run mean growth
- When $\sigma_\upsilon^2 = 0$, y_t is a Trend Stationary Process (TSP).
- When $\sigma_\upsilon^2 > 0, y_t$ is a Difference Stationary Process (DSP).

II. Trend-Cycle Decomposition: Classical and Bayesian Perspectives

Unobserved Components Model of Real GDP (Structural Model)

$$y_t = x_t + z_t,$$

$$x_t = \mu_t + x_{t-1} + \upsilon_t,$$

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-1} + \varepsilon_t - \theta_0 \varepsilon_{t-1},$$

$$\begin{bmatrix} \upsilon_t \\ \varepsilon_t \end{bmatrix} \sim i.i.d.N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\upsilon^2 & \rho\sigma_\nu\sigma_\varepsilon \\ \rho\sigma_\varepsilon\sigma_\nu & \sigma_\varepsilon^2 \end{bmatrix}\right).$$

- Clark (QJE, 1987), Morley et al. (RES, 2003), Oh et al. (JE, 2008), Perron and Wada (JME, 2009), Luo and Startz (JME, 2014)
- **The structural model is not identified.**

II. Trend-Cycle Decomposition: Classical and Bayesian Perspectives

ARIMA (2,1,2) Model of Real GDP (Reduced-form Model)

$$\Delta y_t = \mu_t + \Delta y_t^*,$$
$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$
$$e_t \sim i.i.d. N\,(0, \sigma^2).$$

- When $\theta_1 + \theta_2 = 1$, y_t is a Trend Stationary Process. ($\sigma_v^2 = 0$)
- When $\theta_1 + \theta_2 < 1$, y_t is a Difference Stationary Process. ($\sigma_v^2 > 0$)
- **Bayesian and ML approaches produce very different parameter estimates.**

II. Trend-Cycle Decomposition: Classical and Bayesian Perspectives

ARIMA (2,1,2) Model of Real GDP (Perron and Wada's Model)

$$\Delta y_t = \mu_t + \Delta y_t^*,$$

$$\mu_t = \mu_0 + \mu_d S_t,$$

$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$

$$e_t \sim i.i.d. N(0, \sigma^2),$$

$$S_t = 0 \;\; for \;\; t \leq 1973{:}I, \quad S_t = 1, \; otherwise.$$

- Result from ML Approach: $\theta_1 + \theta_2 = 1$ (y_t is a Trend Stationary Process.)
- Result from Bayesian Approach: $\theta_1 + \theta_2 < 1$ (y_t is a Difference Stationary Process.)

II. Trend-Cycle Decomposition: Classical and Bayesian Perspectives

ARIMA (2,1,2) with a Known Break Point in Mean [1947:Q1-1998:Q2]

$$\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*,$$
$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$
$$e_t \sim i.i.d. N(0, \sigma^2),$$
$$S_t = 0 \ \ for \ \ t \leq 1973{:}Q1, \quad S_t = 1, \ otherwise.$$

Maximum Likelihood Estimates

Parameters	Global Maximum		Local Maximum	
	Estimates	S.E.	Estimates	S.E.
μ_0	0.951	0.021	0.979	0.116
μ_d	-0.287	0.038	-0.332	0.166
$\phi_1 + \phi_2$	0.921	0.020	0.630	0.104
ϕ_2	-0.601	0.109	-0.731	0.150
$\boldsymbol{\theta_1 + \theta_2}$	**0.999**	**0.003**	**0.546**	**0.142**
θ_2	-0.283	0.137	-0.542	0.211
σ^2	0.876	0.086	0.922	0.091
Long-run Impulse-Response	0.000	0.042	1.228	0.312
Log Likelihood	-278.930		-282.710	

II. Trend-Cycle Decomposition: Classical and Bayesian Perspectives

ARIMA (2,1,2) with a Known Break Point in Mean [1947:Q1-1998:Q2]

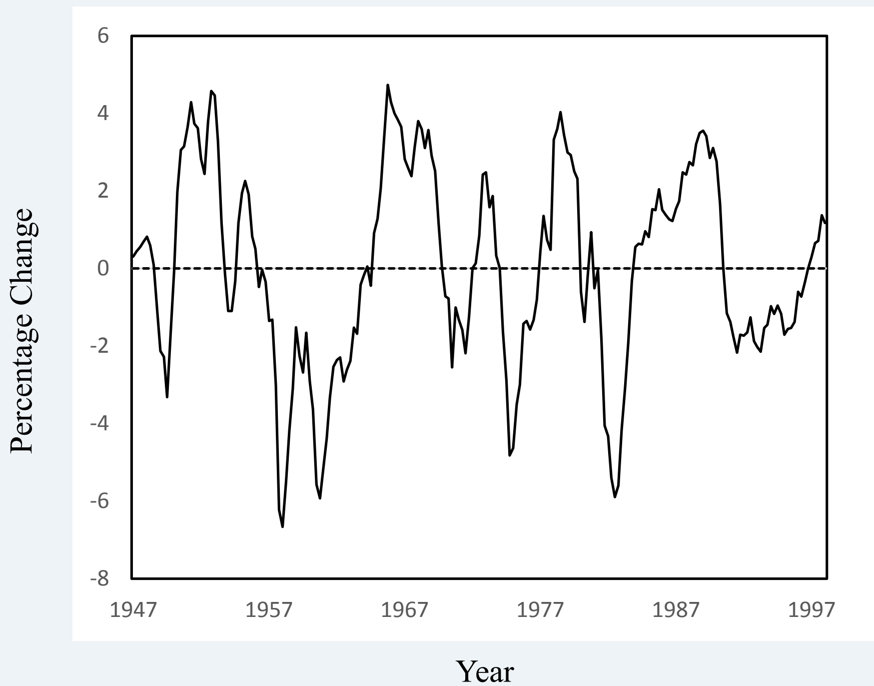
$$\Delta y_t = \mu_0 + \mu_1 S_t + \Delta y_t^*,$$
$$\Delta y_t^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},$$
$$e_t \sim i.i.d.N\,(0, \sigma^2),$$
$$S_t = 0 \;\; for \;\; t \leq 1973{:}Q1, \;\; S_t = 1, \; otherwise.$$

Bayesian Estimates

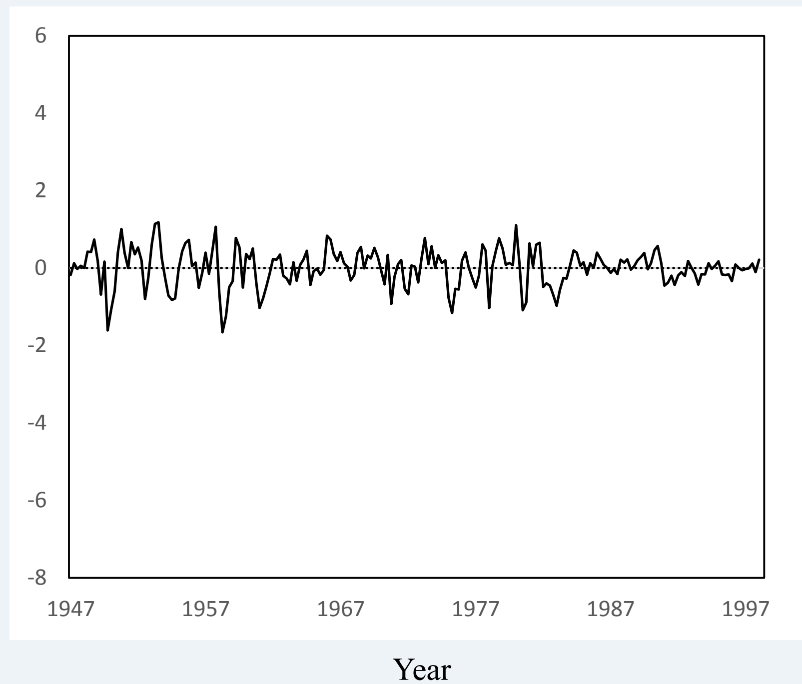
Parameters	Prior		Posterior				
	Mean	SD	Mean	Mode	SD	90 % HPDI	
μ_0	1	1	0.996	0.972	0.129	0.798	1.208
μ_d	-0.4	1	-0.363	-0.310	0.176	-0.639	-0.067
$\phi_1 + \phi_2$	0.5	1	0.521	0.595	0.245	0.165	0.939
ϕ_2	0	1	-0.422	-0.522	0.249	-0.801	-0.020
$\boldsymbol{\theta_1 + \theta_2}$	**0.5**	**1**	**0.311**	**0.460**	**0.390**	**-0.279**	**1.000**
θ_2	0	1	-0.325	-0.373	0.174	-0.596	-0.027
σ^2	1	2	0.966	0.952	0.095	0.815	1.119
Long-run Impulse-Response			1.356	1.420	0.334	0.899	1.893

II. Trend-Cycle Decomposition: Classical and Bayesian Perspectives

Classical Inference (MLE)



Bayesian Inference (Posterior Mode)



Cyclical Component

Purposes of this Paper

1. **We show that the conventional wisdom does not apply to the empirical model of real GDP.**

   ARIMA model of real GDP (Perron and Wada (JME, 2009))

   Result from ML Approach: "Real GDP is a Trend Stationary Process (TSP)"
   Result from Bayesian Approach: "Real GDP is a Difference Stationary Process(DSP)"

2. **We show that the Maximum Likelihood approach suffers severely from the 'pile-up' problem, while the Bayesian approach is relatively free from it. We explain why the two approaches produce different results.**
   >> Difference in how to handle the nuisance parameters

3. **We employ a new multi-move Markov-Chain Monte Carlo (MCMC) algorithm proposed by Kim and Kim (JBES, 2015) to estimate a structural break ARIMA model of real GDP.**

4. **We provide convincing evidence that the cycle measure from a DSP model has out-of-sample predictive power for future output growth at short horizons.**

III. Pile-up Problem with Nuisance Parameters

Model #1: MA (1) without Intercept

$$y_t = e_t - \theta e_{t-1},\ e_t \sim i.i.d\ N\ (0, \sigma^2)$$

$$t = 1,2,\ldots,T,$$

$$[\theta = 0.8, \sigma^2 = 1]$$

Model #2: MA (1) with Intercept

$$y_t = \mu + e_t - \theta e_{t-1},\ e_t \sim i.i.d\ N\ (0, \sigma^2)$$

$$t = 1,2,\ldots,T,$$

$$[\theta = 0.8, \sigma^2 = 1, \mu = 1]$$

Model #3: MA (1) with a Structure Break in Intercept

$$y_t = \mu + \mu_d S_t + e_t - \theta e_{t-1},$$

$$e_t \sim i.i.d\ N\ (0, \sigma^2)$$

$$S_t = 0, for\ t \leq \frac{T}{2};\ S_t = 1, otherwise,$$

$$t = 1,2,\ldots,T,$$

$$[\theta = 0.8, \sigma^2 = 1, \mu = 1, \mu_1 = -0.3]$$

Model #4: ARMA (1,1) with a Structure Break in Intercept

$$y_t = \mu + \mu_d S_t + u_t,$$

$$u_t = \phi u_{t-1} + e_t - \theta e_{t-1},$$

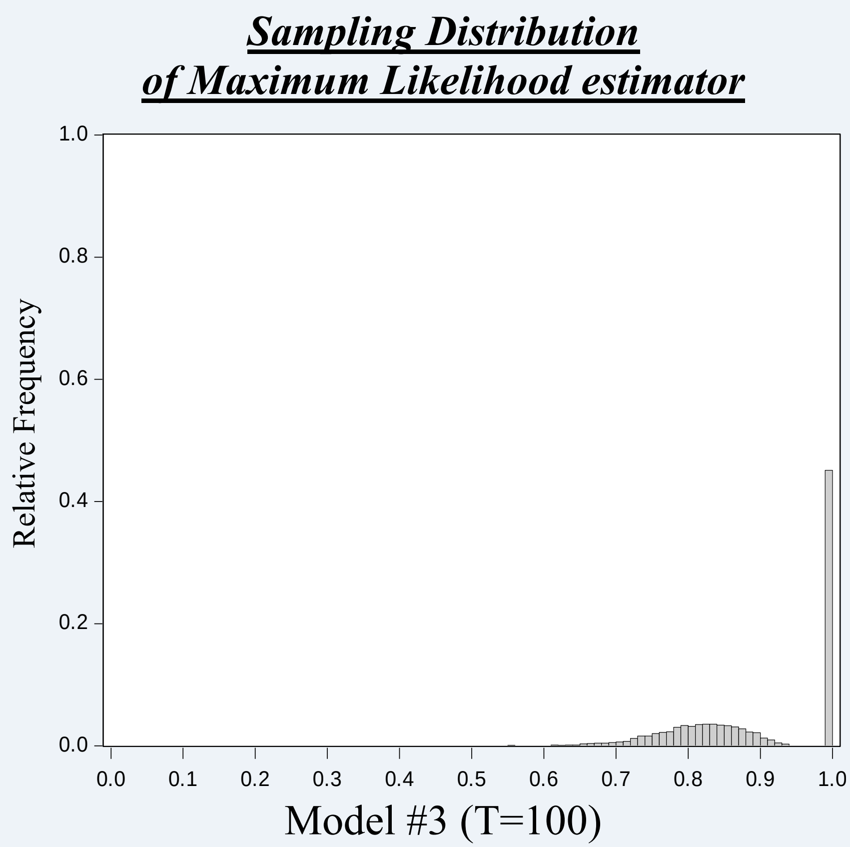
$$e_t \sim i.i.d\ N\ (0, \sigma^2)$$

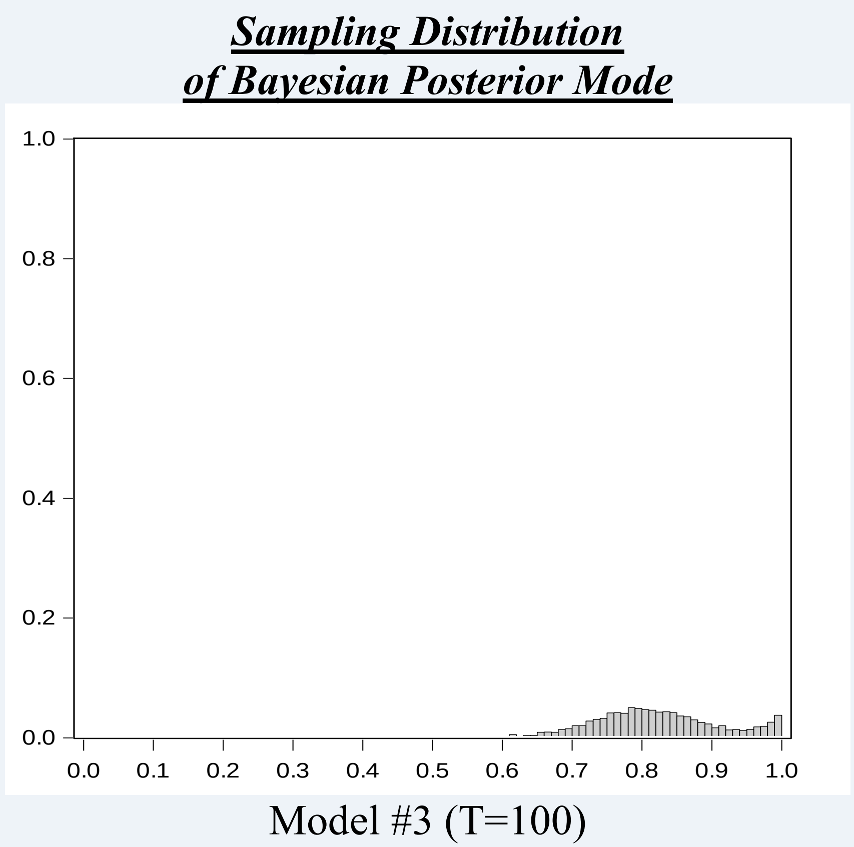
$$S_t = 0, for\ t \leq \frac{T}{2};\ S_t = 1, otherwise,$$

$$t = 1,2,\ldots,T,$$

$$[\mu = 1, \theta = 0.8, \sigma^2 = 1, \mu = 1, \mu_1 = -0.3, \phi = 0.3]$$

III. Pile-up Problem with Nuisance Parameters

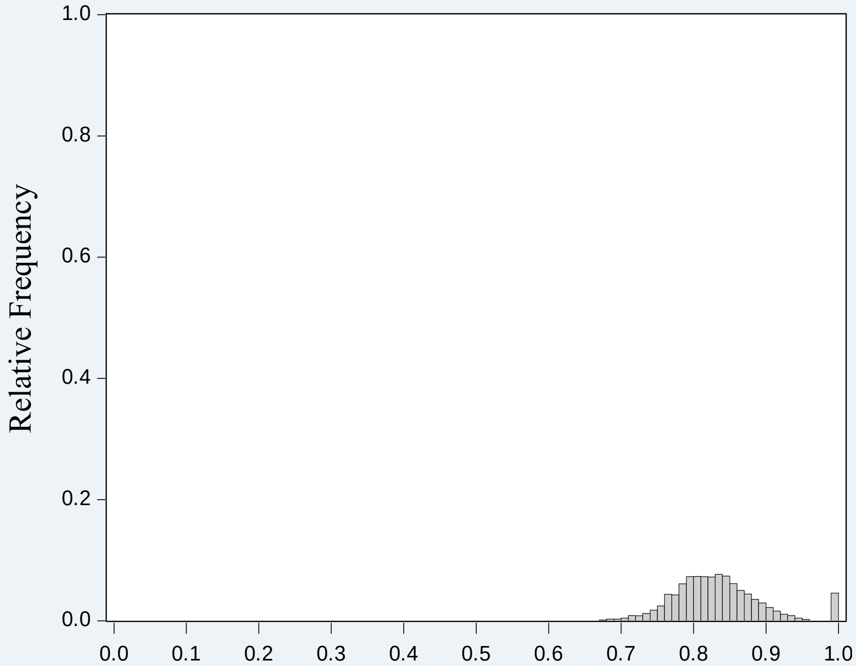




- For the models and sample sizes considered, the probabilities of the pile-up problem are '**considerably**' smaller in the Bayesian approach than in the maximum likelihood approach.

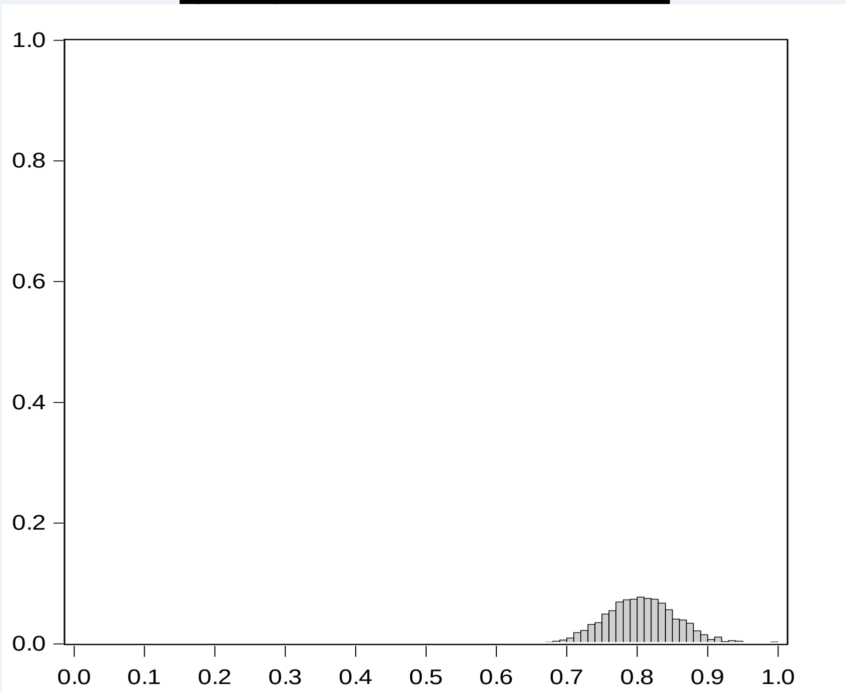
III. Pile-up Problem with Nuisance Parameters

Sampling Distribution of Maximum Likelihood estimator



Model #3 (T=200)

Sampling Distribution of Bayesian Posterior Mode

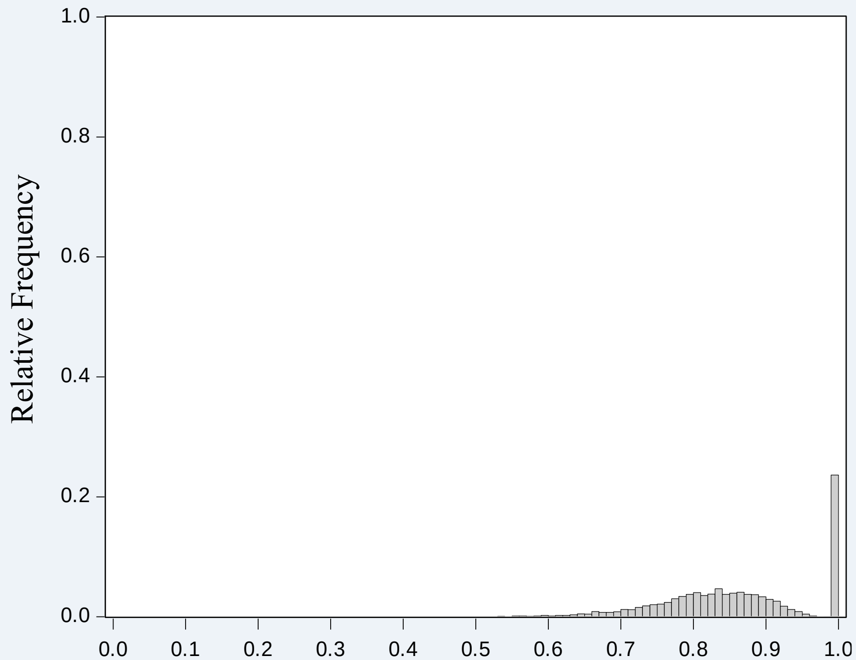


Model #3 (T=200)

- For the models and sample sizes considered, the probabilities of the pile-up problem are '**considerably**' smaller in the Bayesian approach than in the maximum likelihood approach.

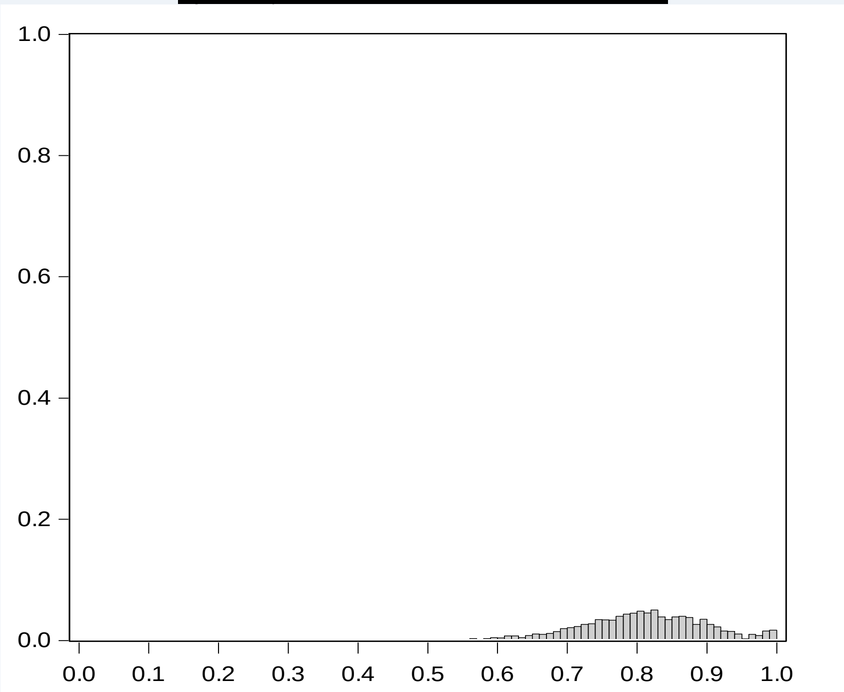
III. Pile-up Problem with Nuisance Parameters

Sampling Distribution of Maximum Likelihood estimator



Model #4 (T=200)

Sampling Distribution of Bayesian Posterior Mode

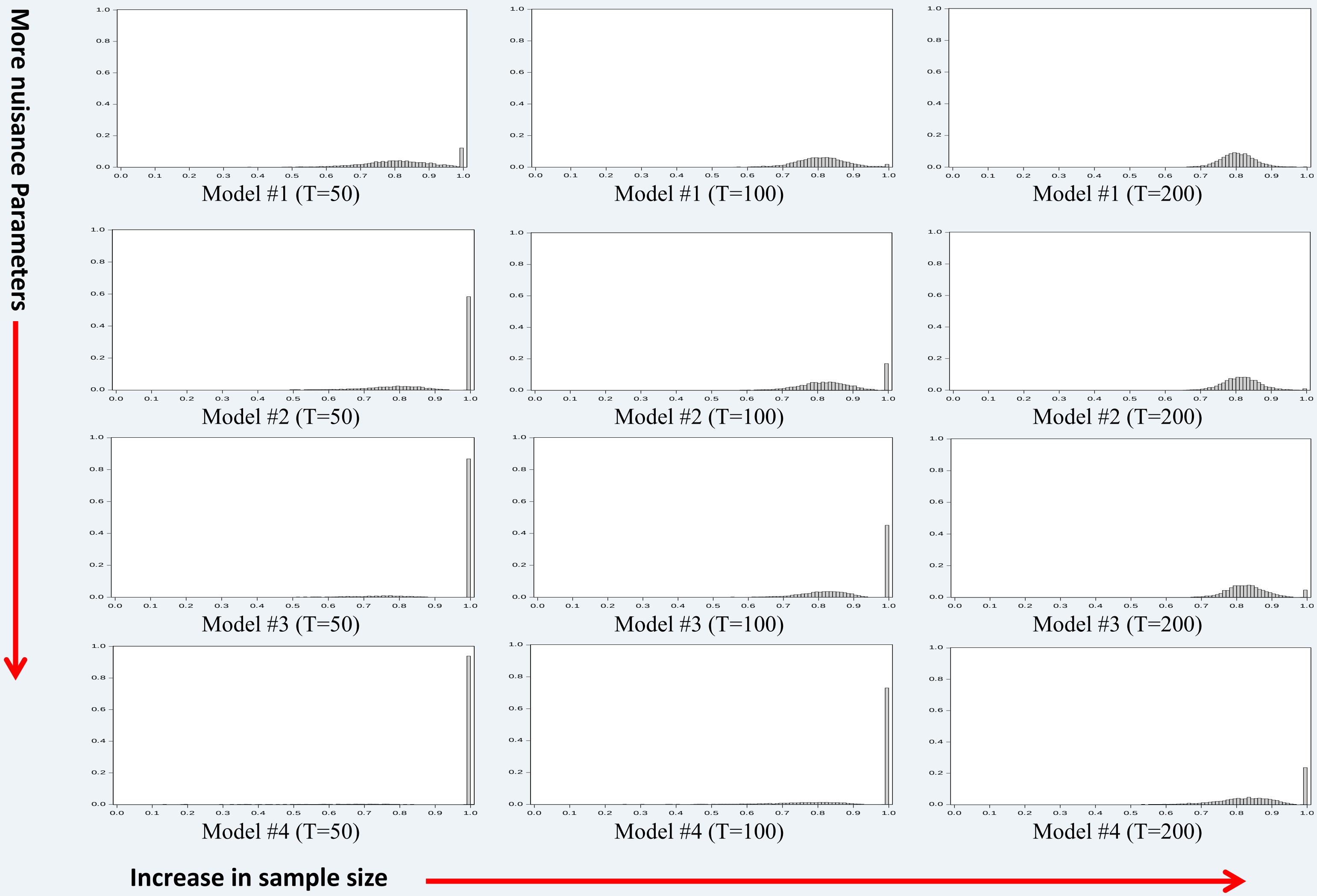


Model #4 (T=200)

- As the model gets more complicated with additional nuisance parameters (i.e. the parameters other than θ), the pile-up problem gets worse.

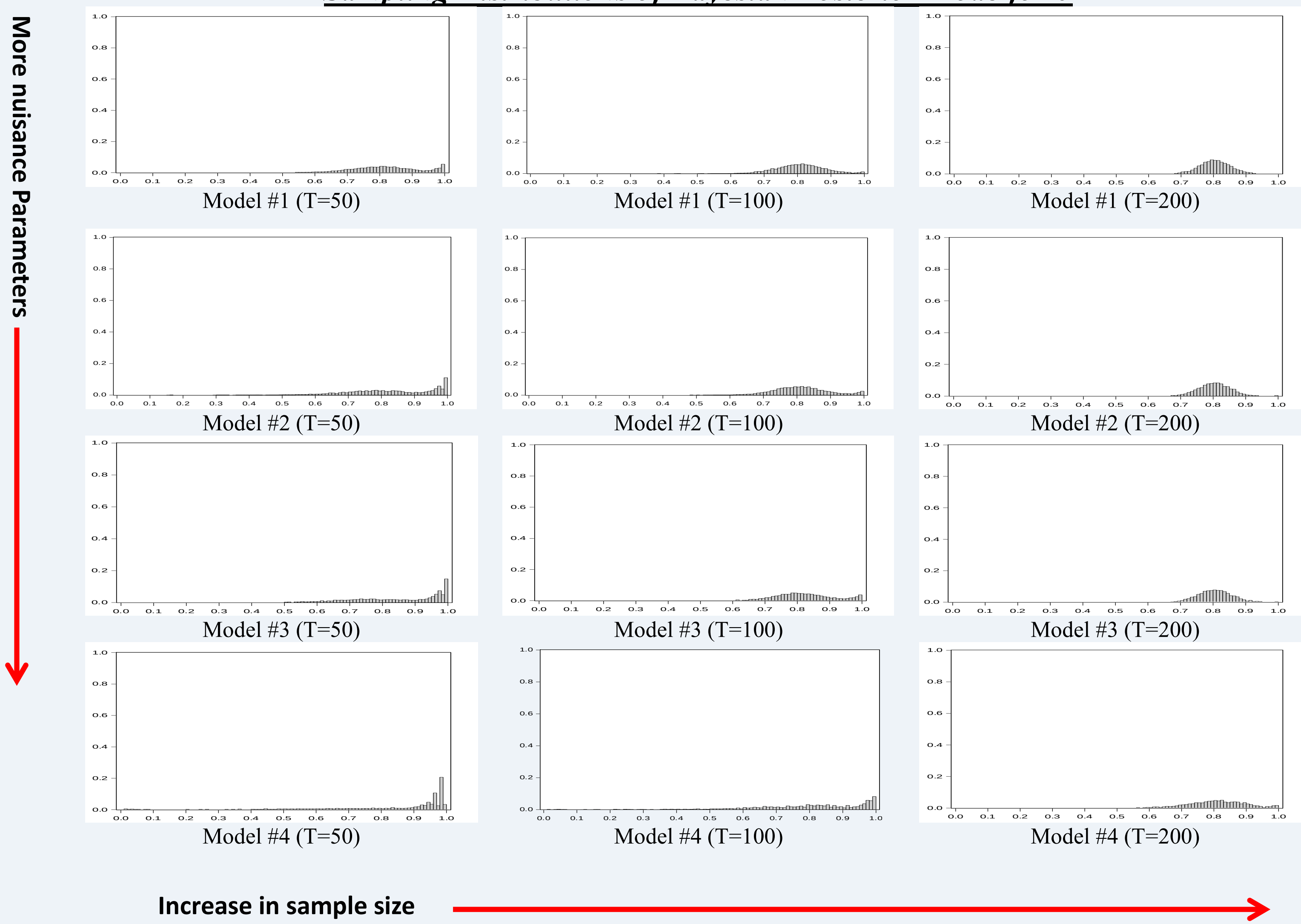
III. Pile-up Problem with Nuisance Parameters

Sampling Distributions of Maximum Likelihood estimators for θ



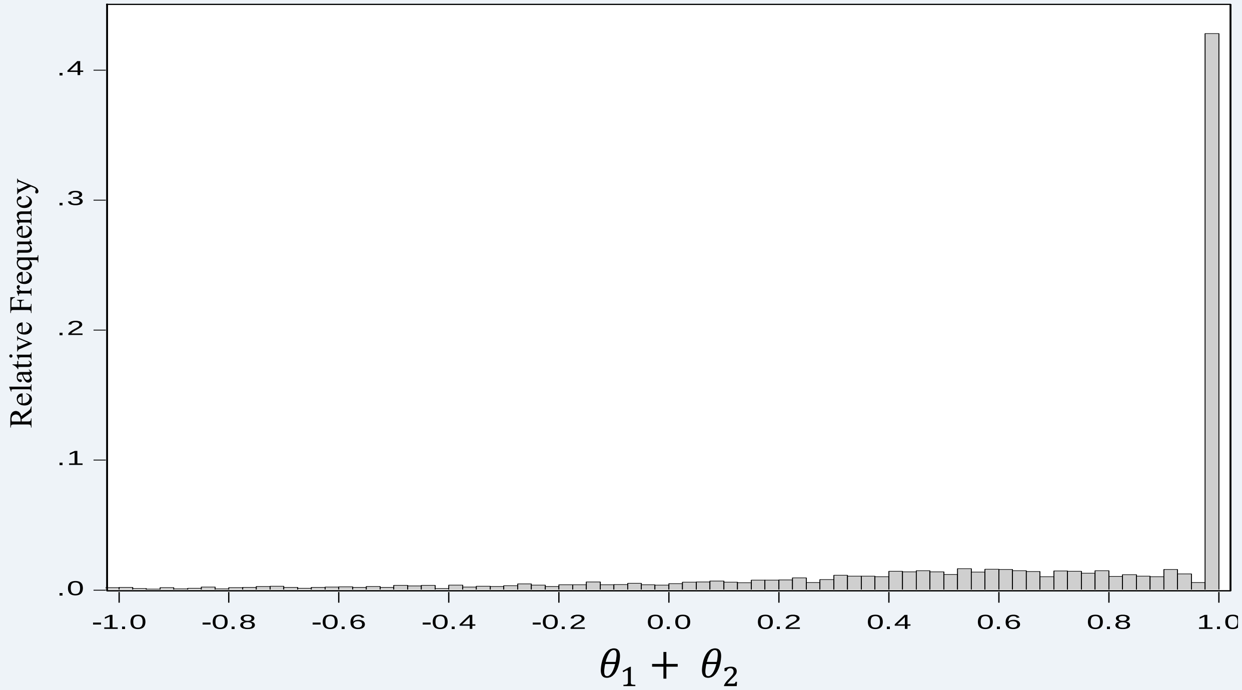
III. Pile-up Problem with Nuisance Parameters

Sampling Distributions of Bayesian Posterior Mode for θ



III. Pile-up Problem with Nuisance Parameters

Sampling Distribution of the Sum of MA Parameters from Monte Carlo Experiment [Perron and Wada's (2009) Model]



- The ML estimator piles up at unity, and the probability of the pile-up problem is calculated to be as high as 42.7%!

Purposes of this Paper

1. **We show that the conventional wisdom does not apply to the empirical model of real GDP.**

ARIMA model of real GDP (Perron and Wada (JME, 2009))

Result from ML Approach: "Real GDP is a Trend Stationary Process (TSP)"
Result from Bayesian Approach: "Real GDP is a Difference Stationary Process(DSP)"

2. **We show that the Maximum Likelihood approach suffers severely from the 'pile-up' problem, while the Bayesian approach is relatively free from it. We explain why the two approaches produce different results.**
   >> Difference in how to handle the nuisance parameters

3. **We employ a new multi-move Markov-Chain Monte Carlo (MCMC) algorithm proposed by Kim and Kim (JBES, 2015) to estimate a structural break ARIMA model of real GDP and perform Bayesian model comparisons.**

4. **We provide convincing evidence that the cycle measure from a DSP model has out-of-sample predictive power for future output growth at short horizons.**

IV. Bayesian Model Comparison: TSP vs. DSP Implications

$$\Delta y_t = \mu_t + \Delta y_t^*,$$

TSP Model: a Non-invertible MA Root

$$(1 - \phi_1 L - \phi_2 L^2)\Delta y_t^* = (1 - \theta_0 L)(1 - L)e_t, \;\; e_t \sim N(0, \sigma_t^2),$$

DSP Model: Invertible MA Roots

$$(1 - \phi_1 L - \phi_2 L^2)\Delta y_t^* = (1 - \theta_1 L - \theta_2 L^2)e_t, \;\; e_t \sim N(0, \sigma_t^2),$$

$$ln(\sigma_t^2) = ln(\sigma_{t-1}^2) + \varepsilon_t, \;\; \varepsilon_t \sim i.i.d. N(0, \sigma_\varepsilon^2).$$

- TSP Model is the corresponding reduced form model to the structural UC model when $\sigma_v^2 = 0$.
- In DSP Model, all the roots of $(1 - \theta_1 L - \theta_2 L^2)$ lie outside the complex unit circle, which means that the DSP model is the corresponding reduced form model to the structural UC model when $\sigma_v^2 > 0$.

IV. Bayesian Model Comparison: TSP vs. DSP Implications

$$\Delta y_t = \mu_t + \Delta y_t^*,$$

<u>TSP Model: a Non-invertible MA Root</u>

$$(1-\phi_1 L - \phi_2 L^2)\Delta y_t^* = (1-\theta_0 L)(1-L)e_t, \;\; e_t \sim N(0,\sigma_t^2),$$

<u>DSP Model: Invertible MA Roots</u>

$$(1-\phi_1 L - \phi_2 L^2)\Delta y_t^* = (1-\theta_1 L - \theta_2 L^2)e_t, \;\; e_t \sim N(0,\sigma_t^2),$$

$$ln(\sigma_t^2) = ln(\sigma_{t-1}^2) + \varepsilon_t, \;\; \varepsilon_t \sim i.i.d.N(0,\sigma_\varepsilon^2).$$

- A reduction in the variance of the shocks to real GDP, namely **the Great Moderation** (Kim and Nelson (1999) and McConnell and Perez-Quiros (2000)), is also incorporated by considering a stochastic volatility.

IV. Bayesian Model Comparison: TSP vs. DSP Implications

$$\Delta y_t = \mu_t + \Delta y_t^*,$$

One Break in Mean

$$\mu_t = \mu_0 + \mu_d I_{S_t=1},$$

$$S_t = 0{,}1,$$

$$I_{S_t=1} = 1 \; if \; S_t = 1,$$

$$I_{S_t=1} = 0 \; otherwise,$$

$$0 < P_{00} < 1, P_{11} = 1$$

Two Breaks in Mean

$$\mu_t = \mu_0 + \mu_d I_{S_t=1,2} + \mu_{dd} I_{S_t=2},$$

$$S_t = 0{,}1{,}2$$

$$I_{S_t=1,2} = 1 \; if \; S_t = 1 \; or \; 2,$$

$$I_{S_t=1,2} = 0 \; otherwise,$$

$$I_{S_t=2} = 1 \; if \; S_t = 2,$$

$$I_{S_t=2} = 0 \; otherwise,$$

$$0 < P_{00} < 1, P_{02} = 0, P_{10} = 0,$$

$$0 < P_{11} < 1, P_{22} = 1$$

IV. Bayesian Model Comparison: TSP vs. DSP Implications

- $\Psi = [\mu, \phi, \theta, \sigma^2]$, $\mathrm{P} = [p_{00}, p_{11}]$,
- $\tilde{S} = [S_1, S_2, \dots, S_T]$

- Markov Chain Monte Carlo Algorithm

Step1. Draw $\tilde{S}$ from $f[\tilde{S} \mid P, \psi, \tilde{Y}]$.
(Kim and Kim (JBES, 2015))

Step3. Draw ψ from $f[\psi \mid P, \tilde{S}, \tilde{Y}]$.
(Chib and Greenberg (JE, 1994), Nakatuma (JE, 2000))

Step4. Draw P from $f[P \mid \psi, \tilde{S}, \widetilde{D}, \tilde{Y}] = f[P \mid \tilde{S}, \widetilde{D}]$.
(Beta Distribution)

IV. Bayesian Model Comparison: TSP vs. DSP Implications

Target Density: $\Pr[\tilde{S} | \tilde{Y}]$

Single-move sampler:

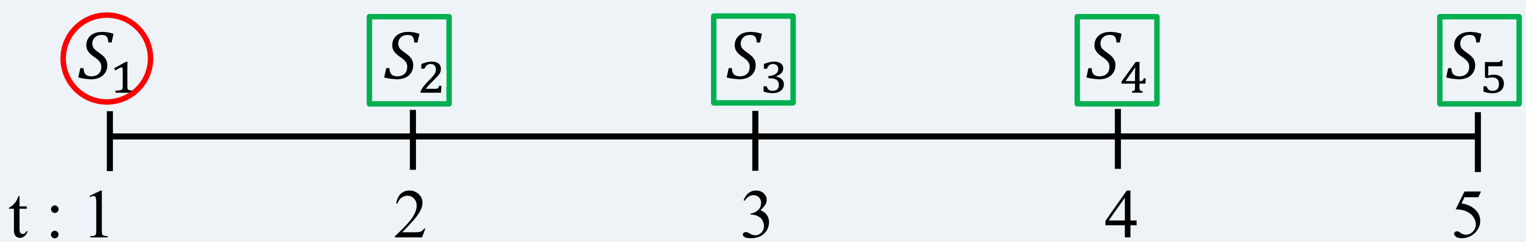
$$\Pr[S_t | \tilde{S}_{\neq t}, \tilde{Y}], \quad t = 1, 2, \ldots, T.$$

IV. Bayesian Model Comparison: TSP vs. DSP Implications

Target Density: $\Pr[S_1, S_2, S_3, S_4, S_5 | \tilde{Y}]$

Single-move sampler:

$$\Pr[S_1 | \tilde{S}_{\neq 1}, \tilde{Y}], \quad t = 1,2,3,4,5.$$

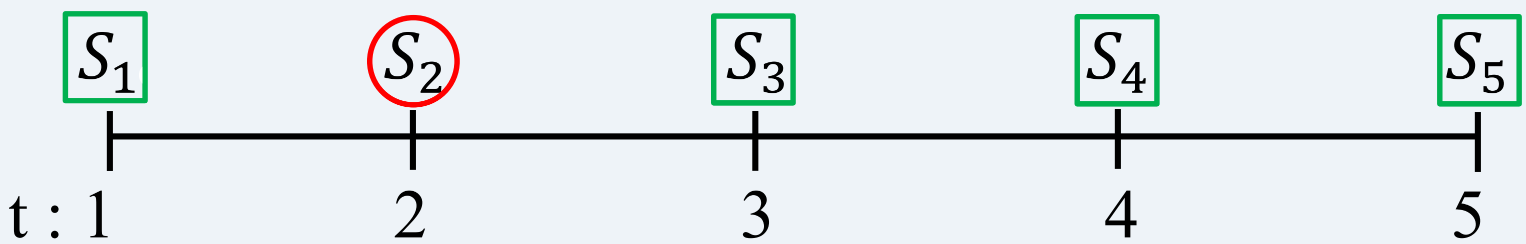


IV. Bayesian Model Comparison: TSP vs. DSP Implications

Target Density: $\Pr[S_1, S_2, S_3, S_4, S_5 | \tilde{Y}]$

Single-move sampler:

$$\Pr[S_2 | \tilde{S}_{\neq 2}, \tilde{Y}], \quad t = 1,2,3,4,5.$$

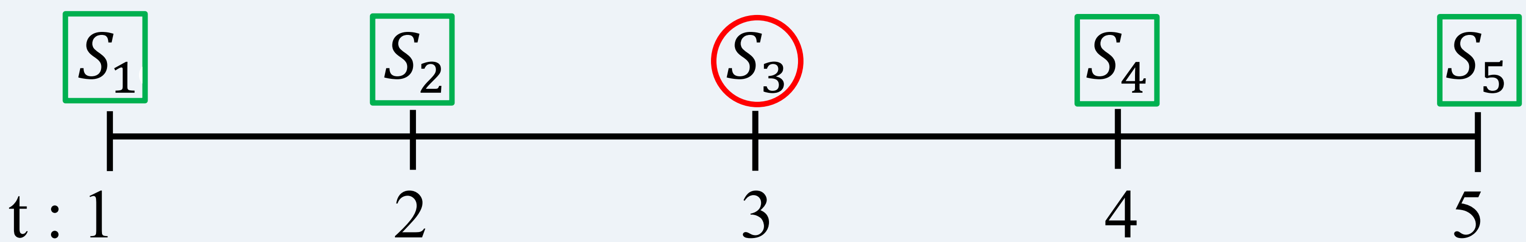


IV. Bayesian Model Comparison: TSP vs. DSP Implications

Target Density: $\Pr[S_1, S_2, S_3, S_4, S_5 | \tilde{Y}]$

Single-move sampler:

$$\Pr[S_3 | \tilde{S}_{\neq 3}, \tilde{Y}], \quad t = 1,2,3,4,5.$$

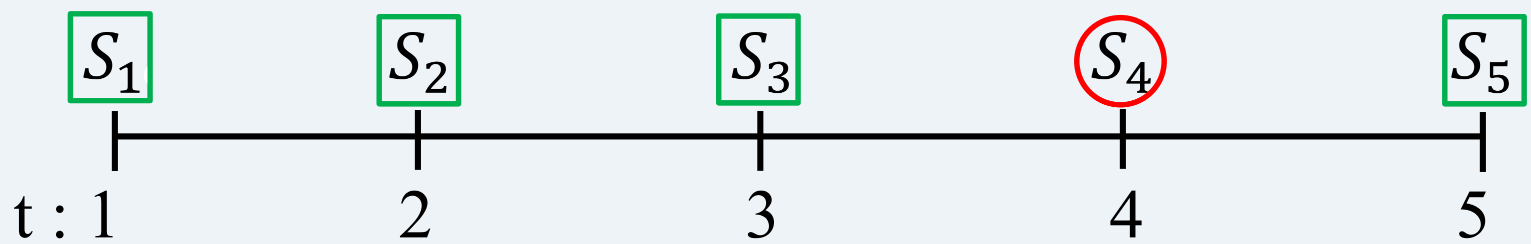


IV. Bayesian Model Comparison: TSP vs. DSP Implications

Target Density: $\Pr[S_1, S_2, S_3, S_4, S_5 | \tilde{Y}]$

Single-move sampler:

$$\Pr[S_4 | \tilde{S}_{\neq 4}, \tilde{Y}], \quad t = 1,2,3,4,5.$$

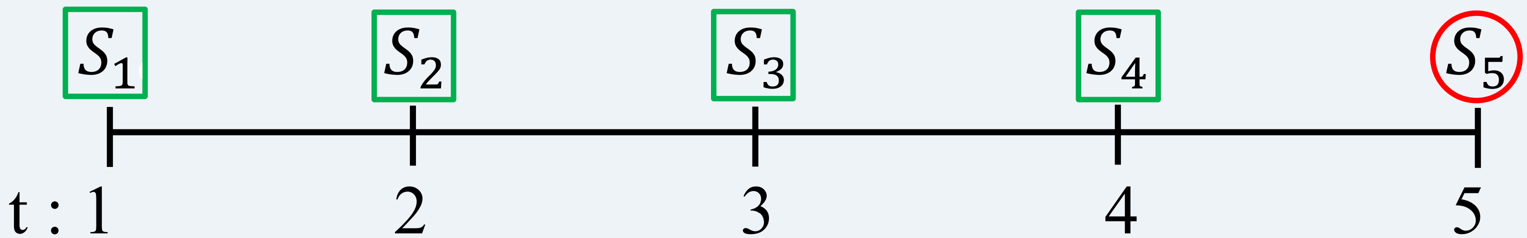


IV. Bayesian Model Comparison: TSP vs. DSP Implications

Target Density: $\Pr[S_1, S_2, S_3, S_4, S_5 | \tilde{Y}]$

Single-move sampler:

$$\Pr[S_5 | \tilde{S}_{\neq 5}, \tilde{Y}], \quad t = 1,2,3,4,5.$$

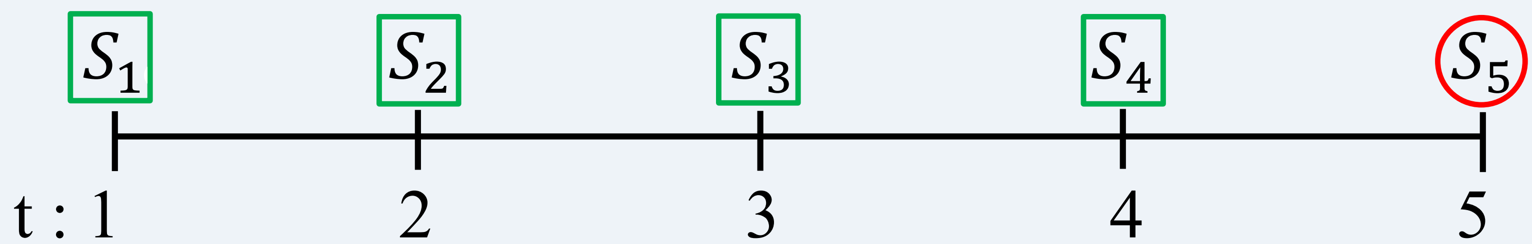


IV. Bayesian Model Comparison: TSP vs. DSP Implications

Target Density: $\Pr[S_1, S_2, S_3, S_4, S_5 | \tilde{Y}]$

Single-move sampler:

$$\Pr[S_5 | \tilde{S}_{\neq 5}, \tilde{Y}], \quad t = 1,2,3,4,5.$$



- Problems

1. Huge Computation
2. High Autocorrelations across MCMC samples

IV. Bayesian Model Comparison: TSP vs. DSP Implications

Target Density: $\Pr[\tilde{S} \,|\, \tilde{Y}]$

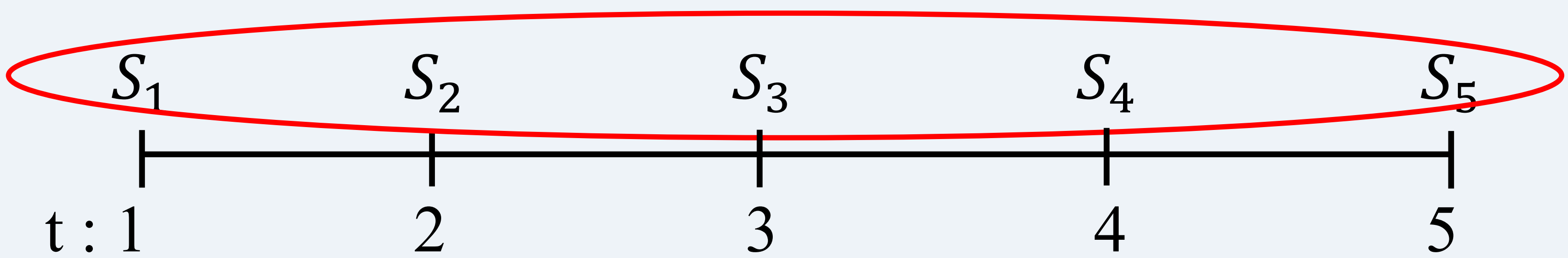
Multi-move sampler: $\Pr[\tilde{S} \,|\, \tilde{Y}]$

IV. Bayesian Model Comparison: TSP vs. DSP Implications

Target Density: $\Pr[\, S_1, S_2, S_3, S_4, S_5 \mid \tilde{Y}]$

Multi-move sampler:

$$\Pr[S_1, S_2, S_3, S_4, S_5 \mid \tilde{Y}]$$



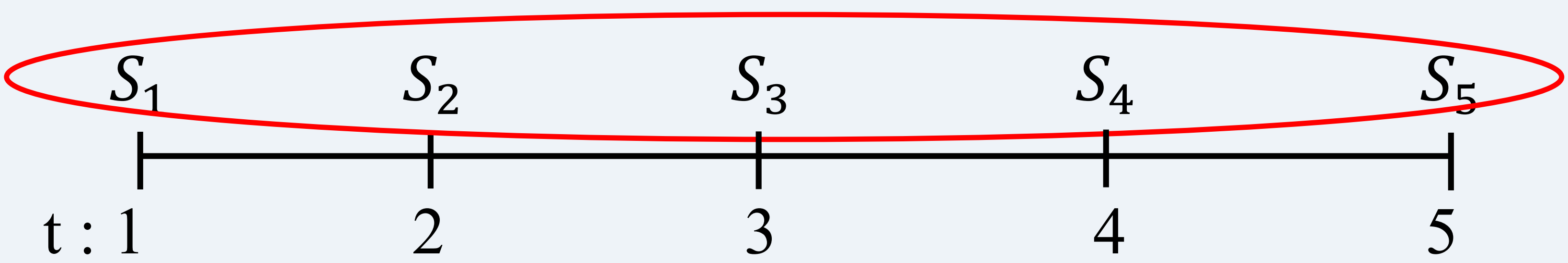
IV. Bayesian Model Comparison: TSP vs. DSP Implications

Target Density: $\Pr[\, S_1, S_2, S_3, S_4, S_5 \mid \tilde{Y}]$

Multi-move sampler:

$$\Pr[S_1, S_2, S_3, S_4, S_5 \mid \tilde{Y}]$$

$$= \Pr[S_5 \mid \tilde{Y}]\, \Pr[S_4 \mid \tilde{Y}, S_5] \; \dots \; \Pr[S_1 \mid \tilde{Y}, S_2, S_3, S_4, S_5\,]$$



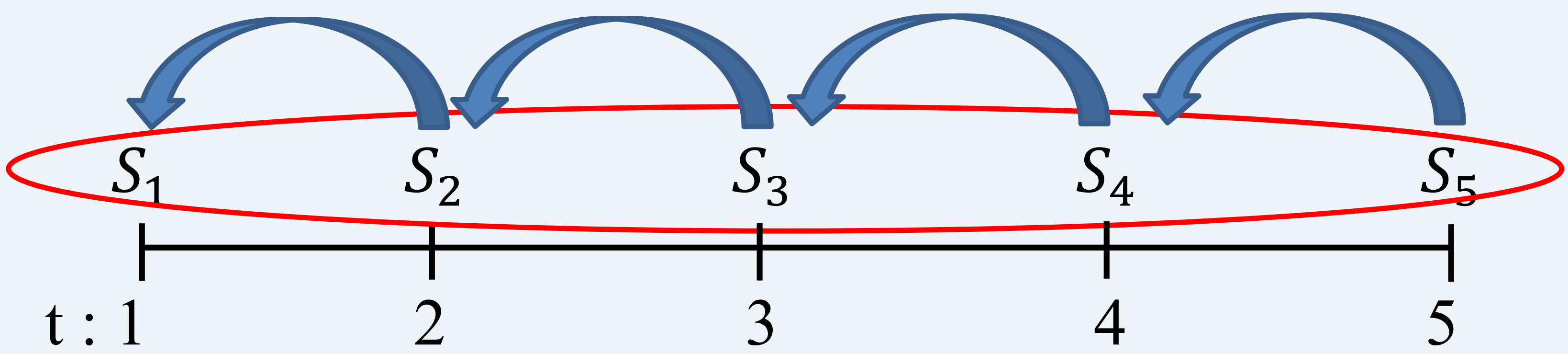
IV. Bayesian Model Comparison: TSP vs. DSP Implications

Target Density: $\Pr[\, S_1, S_2, S_3, S_4, S_5 \mid \tilde{Y}]$

Multi-move sampler:

$$\Pr[S_1, S_2, S_3, S_4, S_5 \mid \tilde{Y}]$$

$$= \Pr[S_5 \mid \tilde{Y}]\, \Pr[S_4 \mid \tilde{Y}, S_5]\ \ldots\ \Pr[S_1 \mid \tilde{Y}, S_2, S_3, S_4, S_5\,]$$



IV. Bayesian Model Comparison: TSP vs. DSP Implications

Multi-move sampler: $\Pr[\,\tilde{S} \mid \tilde{Y}\,]$

- Computational Efficiency
- Fast Convergence
- Accurate Bayesian Inference
- Theoretical justification:
  (Liu et al. (1994, 1995), and Scott (2002))

IV. Bayesian Model Comparison: TSP vs. DSP Implications

- Multi-move Sampling Algorithm with MH:

$$f(\tilde{S}\,|\tilde{Y}\,) \;= f(S_T\,|\,\tilde{Y})\,f(S_{T-1}\,|\,S_T,\tilde{Y})\;\ldots f(S_1|S_2,S_3,\ldots,S_T,\tilde{Y})$$

$$=\;\sum_{t=1}^{T} f\,(S_t\,|S_{t+1},\ldots,S_T,\tilde{Y}),$$

where

$$f(S_t|S_{t+1},\ldots,S_T,\tilde{Y})\;\propto f(S_{t+1}|S_t)f(S_t\,|\,\tilde{Y}_t)\prod_{k=t+1}^{T} f(y_k|\tilde{Y}_{k-1},\tilde{S}_k)$$

$$\approx\; f(S_{t+1}|S_t)f(S_t|\,\tilde{Y}_t),$$

$f(S_{t+1}|S_t)$ is a transition probability and $f(S_t|\,\tilde{Y}_t)$ is a filtered probability.

IV. Bayesian Model Comparison: TSP vs. DSP Implications

- Multi-move Sampling Algorithm with MH:

- $f(S_t|\tilde{Y}_t) \approx h(S_t|\tilde{Y}_t)$ is the approximated filtered probability from the approximate Kalman filter developed by Kim (1994).

-State Space Representation of Switching ARMA (p,q) Model

$$y_t = \mu_{S_t} + H\alpha_t,$$

$$\alpha_t = F_{S_t}\alpha_{t-1} + Ge_t, e_t \sim i.i.d\ N(0, \sigma_{S_t}^2)$$

IV. Bayesian Model Comparison: TSP vs. DSP Implications

- Multi-move Sampling Algorithm with MH:

-State Space Representation of Switching ARMA (p,q) Model

$$y_t = \mu_{S_t} + H\,\alpha_t,$$

$$\alpha_t = F_{S_t}\alpha_{t-1} + Ge_t, e_t \sim i.i.d\; N(0, \sigma^2_{S_t})$$

-Conditional Kalman Filter ($S_t = j$ and $S_{t-1} = i$)

$$\alpha_{t|t-1}{}^{(i,j)} = F_j \alpha_{t-1|t-1}{}^{(i)}, \qquad P_{t|t-1}{}^{(i,j)} = F_j P_{t-1|t-1}{}^{(i)} F_j' + GG'\sigma_j^2,$$

$$\eta_{t|t-1}{}^{(i,j)} = y_t - \mu_{S_t} - H\,\alpha_{t|t-1}{}^{(i,j)}, \qquad f_{t|t-1}{}^{(i,j)} = H\,P_{t|t-1}{}^{(i,j)} H',$$

$$\alpha_{t|t}{}^{(i,j)} = \alpha_{t|t-1}{}^{(i,j)} + P_{t|t-1}{}^{(i,j)} H' \left[f_{t|t-1}{}^{(i,j)}\right]^{-1} \eta_{t|t-1}{}^{(i,j)},$$

$$P_{t|t}{}^{(i,j)} = \left(I - P_{t|t-1}{}^{(i,j)} H' \left[f_{t|t-1}{}^{(i,j)}\right]^{-1} H\right) P_{t|t-1}{}^{(i,j)}.$$

IV. Bayesian Model Comparison: TSP vs. DSP Implications

- Multi-move Sampling Algorithm with MH:

-Hamilton Filter ($S_t = j$ and $S_{t-1} = i$)

$$\eta_{t|t-1}{}^{(i,j)}, f_{t|t-1}{}^{(i,j)} \textit{from conditional Kalman filter}$$

$$f(S_t, S_{t-1}|\tilde{Y}_{t-1}) = f(S_t|S_{t-1})f\left(S_{t-1}|\tilde{Y}_{t-1}\right),$$

$$f\left(y_t|\tilde{Y}_{t-1}\right) = \sum_{S_t}\sum_{S_{t-1}} f(y_t|S_t, S_{t-1}, \tilde{Y}_{t-1})f(S_t, S_{t-1}|\tilde{Y}_{t-1}),$$

$$f(S_t, S_{t-1}|\tilde{Y}_t) = \frac{f(y_t, S_t, S_{t-1}|\tilde{Y}_{t-1})}{f\left(y_t|\tilde{Y}_{t-1}\right)} = \frac{f(y_t|S_t, S_{t-1}, \tilde{Y}_{t-1})f(S_t, S_{t-1}|\tilde{Y}_{t-1})}{f\left(y_t|\tilde{Y}_{t-1}\right)},$$

$$\Pr[S_t = j \mid \tilde{Y}_t] = \sum_{S_{t-1}} f(S_t = j, S_{t-1}|\tilde{Y}_t).$$

IV. Bayesian Model Comparison: TSP vs. DSP Implications

- Multi-move Sampling Algorithm with MH:

-Approximate Kalman Filter by Kim (1994) ($S_t = j$ and $S_{t-1} = i$)

$$\alpha_{t|t}{}^{(i,j)}, P_{t|t}{}^{(i,j)} \mathit{from\ conditional\ Kalman\ filter}$$

$$f(S_t = j, S_{t-1}|\tilde{Y}_t), \Pr[S_t = j\big|\tilde{Y}_t]\ \mathit{from\ Hamilton\ filter}$$

$$\alpha_{t|t}{}^{(j)} \approx \frac{\sum_{S_{t-1}} f(S_t = j, S_{t-1}|\tilde{Y}_t)}{\Pr[S_t = j\big|\tilde{Y}_t]}\ \alpha_{t|t}{}^{(i,j)},$$

$$P_{t|t}{}^{(j)} \approx \frac{\sum_{S_{t-1}} f(S_t = j, S_{t-1}|\tilde{Y}_t)}{\Pr[S_t = j\big|\tilde{Y}_t]}\ [P_{t|t}{}^{(i,j)} + \left(\alpha_{t|t}{}^{(j)} - \alpha_{t|t}{}^{(i,j)}\right)\left(\alpha_{t|t}{}^{(j)} - \alpha_{t|t}{}^{(i,j)}\right)']$$

IV. Bayesian Model Comparison: TSP vs. DSP Implications

- Multi-move Sampling Algorithm with MH:

- $f(S_t|\,\tilde{Y}_t) \approx h(S_t|\,\tilde{Y}_t)$ is the approximated filtered probability from the approximate Kalman filter developed by Kim (1994).

-The candidate generating density for S_t is

$$g(S_t|S_{t+1},\dots,S_T,\tilde{Y}) \propto f(S_{t+1}|S_t)h(S_t|\,\tilde{Y}_t)$$

-The candidate generating density for $\tilde{S}$ is

$$G(\tilde{S}|\,\tilde{Y}) \propto \prod_{t=1}^{T} g(S_t|S_{t+1},\dots,S_T,\tilde{Y})$$

-The target density for $\tilde{S}$ is

$$F(\tilde{S}|\,\tilde{Y}) \propto \prod_{t=1}^{T} f(S_{t+1}|S_t)f(y_t|\,\tilde{S}_t\,,\tilde{Y}_{t-1})$$

IV. Bayesian Model Comparison: TSP vs. DSP Implications

- Multi-move Sampling Algorithm with MH:

-The acceptance probability is

$$\alpha(\tilde{S}^J, \tilde{S}^{J-1} | \tilde{Y})$$

$$= \min\left(\frac{F(\tilde{S}^J | \tilde{Y})}{F(\tilde{S}^{J-1} | \tilde{Y})} \frac{G(\tilde{S}^{J-1} | \tilde{Y})}{G(\tilde{S}^J | \tilde{Y})}, 1\right),$$

$$= \min\left(\prod_{t=1}^{T} \frac{f\left(y_t \middle| \tilde{S}_t^J, \tilde{Y}_{t-1}\right)}{f\left(y_t \middle| \tilde{S}_t^{J-1}, \tilde{Y}_{t-1}\right)} \prod_{t=1}^{T} \frac{h\left(\tilde{S}_t^{J-1} \middle|, \tilde{Y}_t\right)}{h\left(\tilde{S}_t^J \middle| \tilde{Y}_t\right)} \prod_{t=0}^{T-1} \frac{h\left(\tilde{S}_{t+1}^J \middle|, \tilde{Y}_t\right)}{h\left(\tilde{S}_{t+1}^{J-1} \middle| \tilde{Y}_t\right)}, 1\right),$$

where $\tilde{S}^J$ and $\tilde{S}^{J-1}$ are the sequences of the regime indicate variables generated at the current and previous iterations of the MCMC algorithm.

IV. Bayesian Model Comparison: TSP vs. DSP Implications

- Multi-move Sampling Algorithm with MH:

Step1. Cast the Markov-switching ARMA model into in a state-space form, conditional on all the parameters.

Step2. Run the approximate Kalman filter by Kim (1994) and save $h(S_t | \tilde{Y}_t)$.

Step3. Generate S_t sequentially for $t = T, T-1, \ldots, 1$ based on the single candidate density.

Step4. Accept or reject the candidate $\tilde{S}^J$ globally according to the acceptance probability.

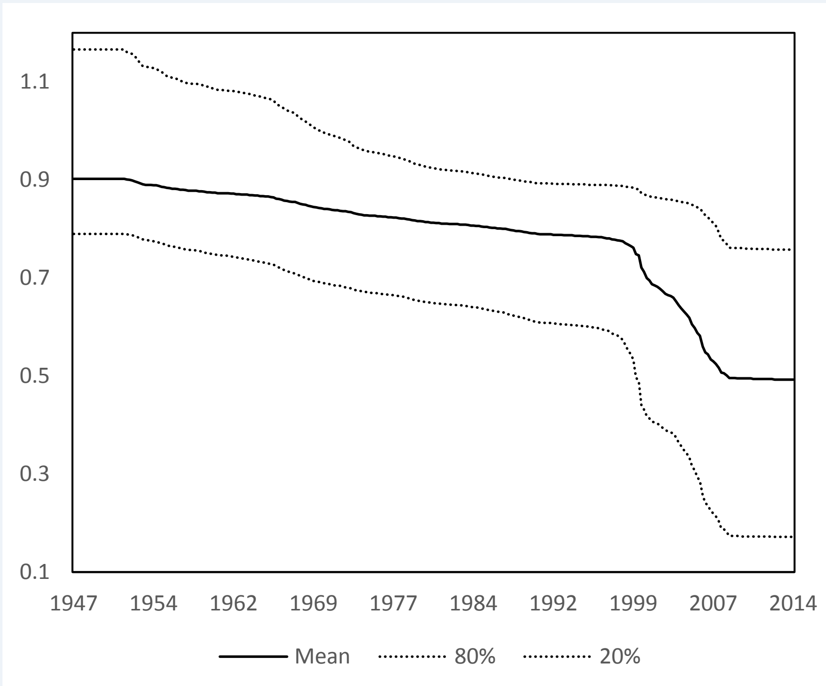
IV. Bayesian Model Comparison: TSP vs. DSP Implications

One Break in Mean

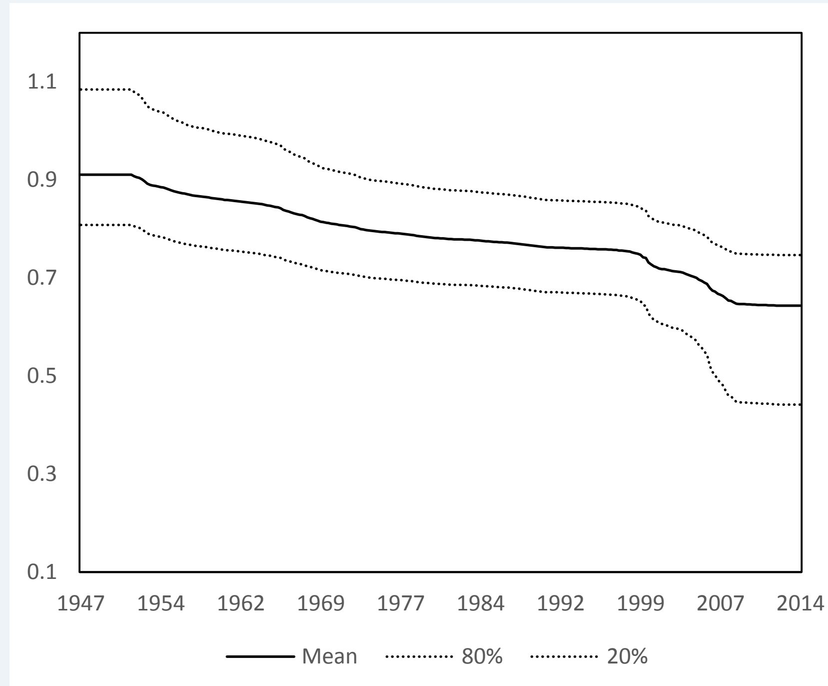
$$\Delta y_t = \mu_t + \Delta y_t^*$$

	Prior		*TSP Model* Posterior			*DSP Model* Posterior		
	Mean	SD	Mean	SD	90 % HPDI	Mean	SD	90 % HPDI
p_{00}	0.99	0.01	0.992	0.006	[0.985, 0.999]	0.991	0.008	[0.981, 0.999]
μ_0	1	1	0.902	0.132	[0.764, 1.083]	0.955	0.190	[0.703, 1.246]
μ_d	-0.4	1	-0.409	0.167	[-0.673, -0.135]	-0.359	0.201	[-0.614, 0.000]
$\phi_1 + \phi_2$	0.5	1	0.982	0.013	[0.964, 0.999]	0.384	0.240	[-0.008, 0.767]
ϕ_2	0	1	-0.533	0.116	[-0.711, -0.330]	-0.106	0.287	[-0.551, 0.356]
$\theta_1 + \theta_2$	**0.5**	**1**	**-**	**-**	**-**	**-0.023**	**0.385**	**[-0.668, 0.590]**
θ_2	0	1	-	-	-	-0.200	0.159	[-0.428, 0.072]
θ_0	0	1	-0.204	0.126	[-0.399, 0.017]	-	-	-
σ_ε^2	0.01	0.03	0.005	0.003	[0.002, 0.009]	0.008	0.005	[0.002, 0.014]
σ_0^2	1	2	1.351	0.358	[0.835, 1.823]	1.345	0.368	[0.840, 1.963]
50% HPDI of Brake Dates			[2000:Q1~2008:Q3]			[1952:Q4~2008:Q3]		
DIC			**661.606**			**654.973**		

IV. Bayesian Model Comparison: TSP vs. DSP Implications



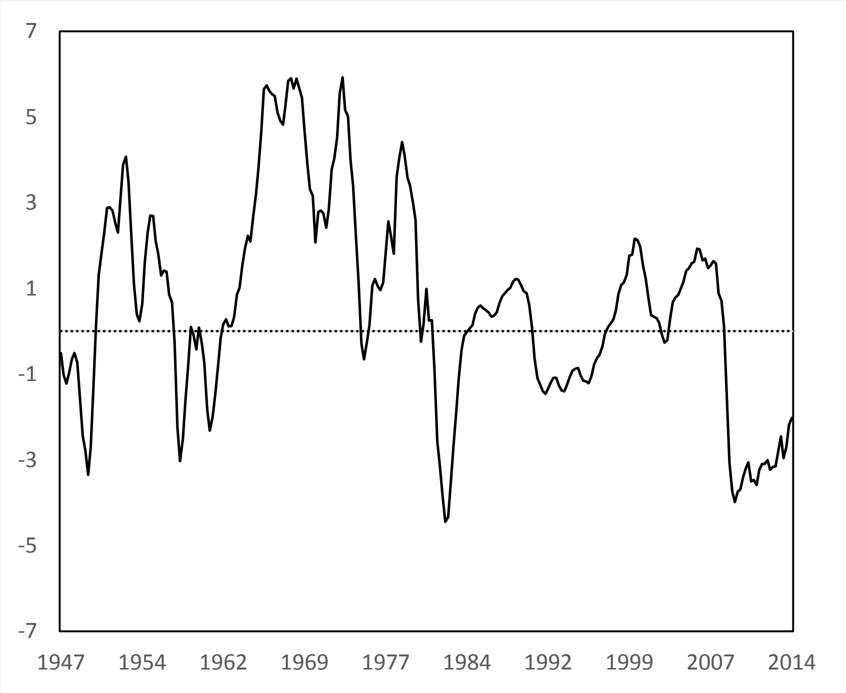
TSP Model with One Break in Mean



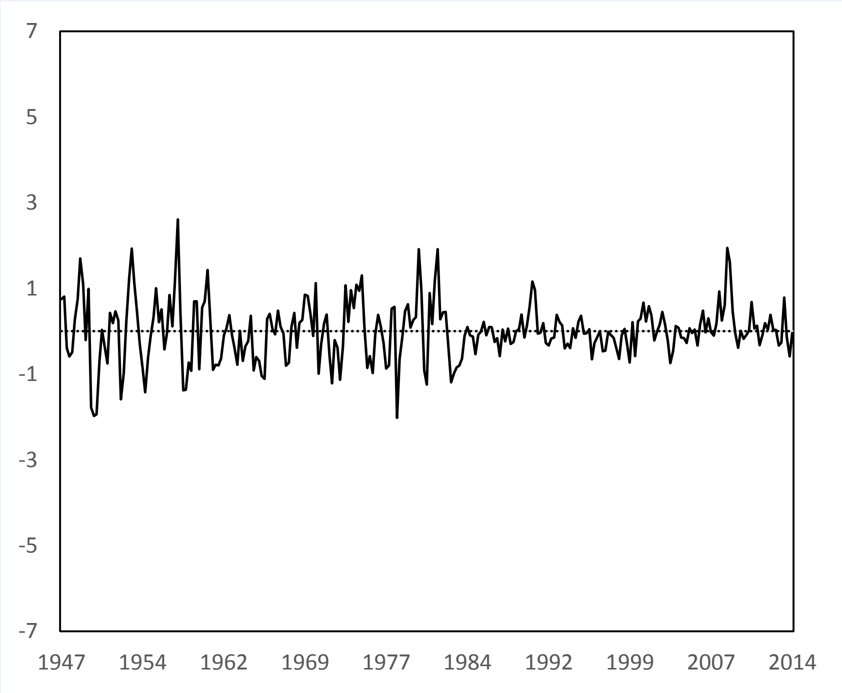
DSP Model with One Break in Mean

- The structural break in the mean growth rate is very sharp, beginning around late 1990's.
- However, there exists a gradual decline in the long-run growth rate before 2000's.

IV. Bayesian Model Comparison: TSP vs. DSP Implications



TSP Model with One Break in Mean



DSP Model with One Break in Mean

- The extracted cycle from the TSP model is persistent and large in magnitude while the cycle measure from the DSP model is small and noisy.

IV. Bayesian Model Comparison: TSP vs. DSP Implications

Two Breaks in Mean

$$\Delta y_t = \mu_t + \Delta y_t^*$$

	Prior		TSP Model Posterior			DSP Model Posterior		
	Mean	SD	Mean	SD	90 % HPDI	Mean	SD	90 % HPDI
μ_0	1	1	0.984	0.139	[0.802, 1.169]	1.044	0.214	[0.740, 1.370]
μ_d	-0.4	1	-0.216	0.132	[-0.384, 0.000]	-0.247	0.186	[-0.518, 0.000]
μ_{dd}	-0.4	1	-0.389	0.156	[-0.639,-0.096]	-0.296	0.187	[-0.567, 0.000]
$\phi_1 + \phi_2$	0.5	1	0.964	0.021	[0.937, 1.000]	0.399	0.246	[-0.026, 0.785]
ϕ_2	0	1	-0.549	0.115	[-0.732,-0.358]	-0.141	0.291	[-0.588, 0.334]
$\boldsymbol{\theta_1 + \theta_2}$	**0.5**	**1**	**-**	**-**	**-**	**0.014**	**0.395**	**[-0.640, 0.630]**
θ_2	0	1	-	-	-	-0.215	0.163	[-0.445, 0.086]
θ_0	0	1	-0.230	0.131	[-0.436,-0.003]	-	-	-
σ_ε^2	0.01	0.03	0.006	0.004	[0.002, 0.010]	0.007	0.005	[0.002, 0.013]
σ_0^2	1	2	1.503	0.536	[0.695, 2.291]	1.384	0.412	[0.812, 2.017]
50 % HPDI of Brake Dates			Break 1: [1952:Q4~1974:Q4] Break 2: [2005:Q2~2007:Q4]			Break 1: [1952:Q4~2006:Q2] Break 2: [2000:Q1~2009:Q1]		
DIC			**654.918**			**655.428**		

IV. Bayesian Model Comparison: TSP vs. DSP Implications

Deviance Information Criterion (DIC)

	TSP Model	*DSP Model*
One Break	**661.606**	**654.973**
Two Breaks	**654.918**	**655.428**

- Under the model specifications with one break in mean, the DSP model is strongly preferred. However, under the model specifications with two breaks in mean, the TSP model is slightly preferred.

- The difference in DIC's for the two preferred models is negligible.

- This conclusion is broadly in line with Cheung and Chinn (1997), who show that unit root tests or stationarity tests do not provide a definite conclusion within the classical framework.

Purposes of this Paper

1. **We show that the conventional wisdom does not apply to the empirical model of real GDP.**

ARIMA model of real GDP (Perron and Wada (JME, 2009))

Result from ML Approach: "Real GDP is a Trend Stationary Process (TSP)"
Result from Bayesian Approach: "Real GDP is a Difference Stationary Process(DSP)"

2. **We show that the Maximum Likelihood approach suffers severely from the 'pile-up' problem, while the Bayesian approach is relatively free from it. We explain why the two approaches produce different results.**

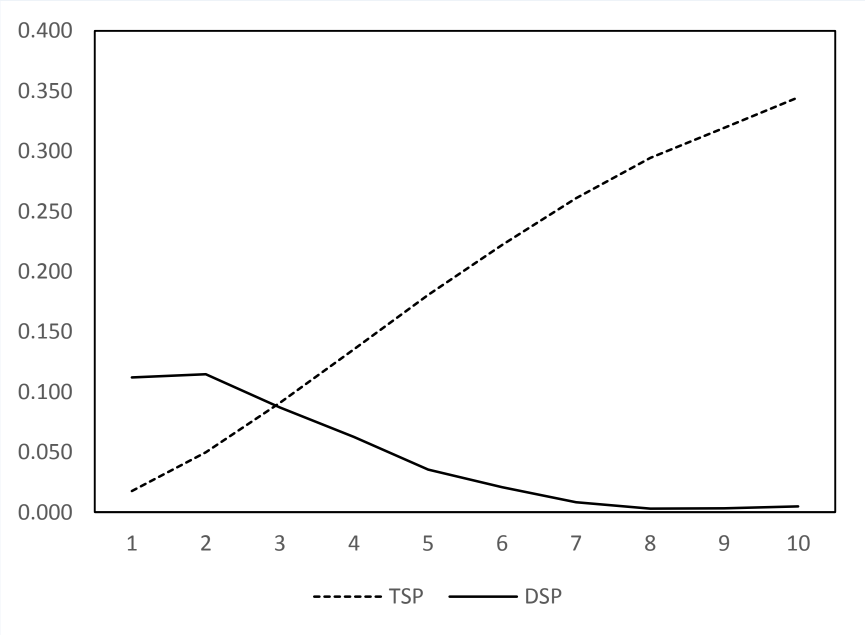
3. **We employ a new multi-move Markov-Chain Monte Carlo (MCMC) algorithm proposed by Kim and Kim (JBES, 2015) to estimate a structural break ARIMA model of real GDP and perform Bayesian model comparisons.**

4. **We provide convincing evidence that the cycle measure from a DSP model has out-of-sample predictive power for future output growth at short horizons.**

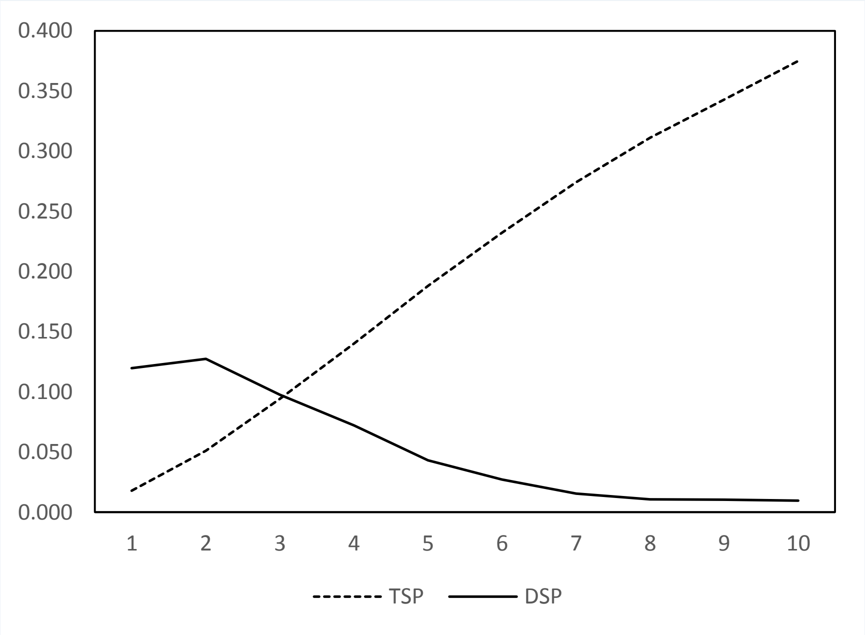
IV. Out-of-Sample Predictive Power of the Cyclical Components

In-sample Fitting: Average R^2's from Rolling Regressions

$$TSP\ Predictive\ Regression:\quad \Delta y_{t-h:t} = \alpha_0 + \alpha_1\, c_{t-h}^{TSP} + u_{t-h:t}$$
$$DSP\ Predictive\ Regression:\quad \Delta y_{t-h:t} = \alpha_0 + \alpha_2\, c_{t-h}^{DSP} + u_{t-h:t}$$



(a) *Window Size: 40 Years*



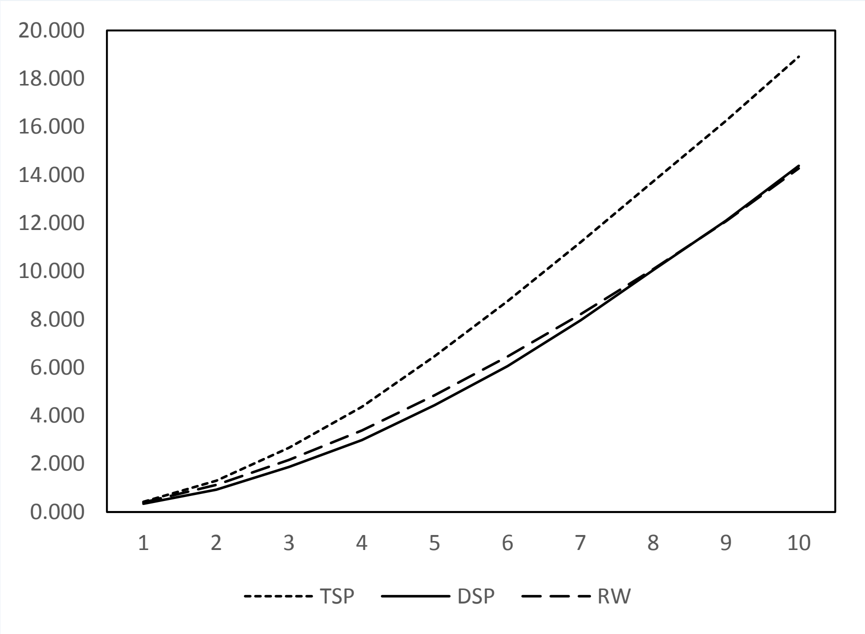
Window Size: 25 Years

Note: 1. The horizontal axis represents the prediction horizon.

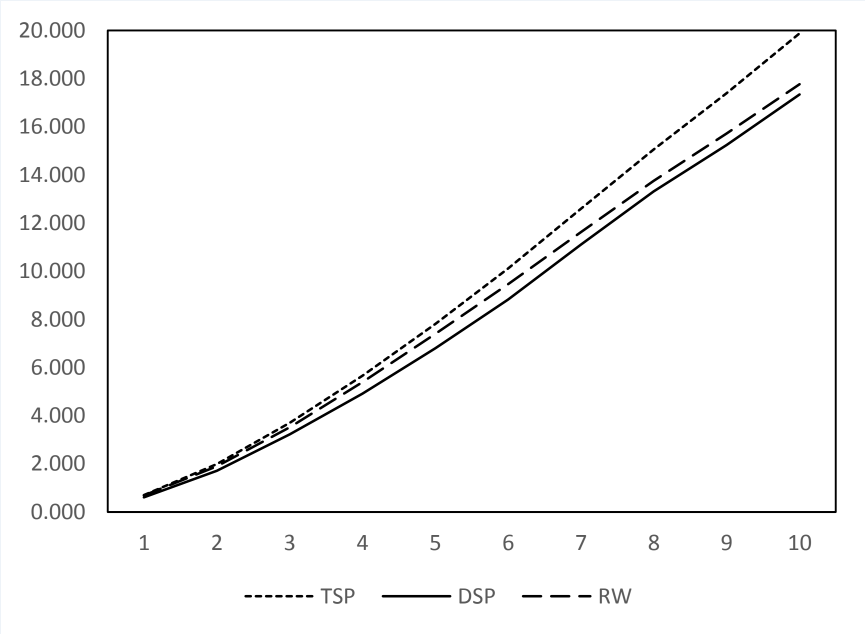
IV. Out-of-Sample Predictive Power of the Cyclical Components

Out-of-sample Predictive Performances: Mean Square Errors from Rolling Regressions

$$TSP\ Predictive\ Regression:\ \ \Delta y_{t-h:t} = \alpha_0 + \alpha_1\, c^{TSP}_{t-h} + u_{t-h:t}$$
$$DSP\ Predictive\ Regression:\ \ \Delta y_{t-h:t} = \alpha_0 + \alpha_2\, c^{DSP}_{t-h} + u_{t-h:t}$$
$$Random\ Walk\ Model:\ \ \Delta y_{t-h:t} = \alpha_0 + u_{t-h:t}$$



(a) *Window Size: 40 Years*



(b) *Window Size: 25 Years*

1. The horizontal axis represents the prediction horizon.

IV. Out-of-Sample Predictive Power of the Cyclical Components

Out-of-sample Predictive Performances: Mean Square Errors from Rolling Regressions

$$TSP\ Predictive\ Regression:\ \ \Delta y_{t-h:t} = \alpha_0 + \alpha_1\, c_{t-h}^{TSP} + u_{t-h:t}$$
$$DSP\ Predictive\ Regression:\ \ \Delta y_{t-h:t} = \alpha_0 + \alpha_2\, c_{t-h}^{DSP} + u_{t-h:t}$$
$$Random\ Walk\ Model:\ \ \Delta y_{t-h:t} = \alpha_0 + u_{t-h:t}$$

	Window Size: 40 Years			*Window Size: 25 Years*		
h	*TSP Regression*	*DSP Regression*	*Random Walk Model*	*TSP Regression*	*DSP Regression*	*Random Walk Model*
1	0.425	0.336	0.392	0.700	0.610	0.682
2	1.298	0.927	1.118	1.986	1.706	1.900
3	2.663	1.868	2.161	3.698	3.218	3.523
4	4.373	2.979	3.390	5.661	4.917	5.393
5	6.466	4.429	4.853	7.816	6.794	7.403
6	8.753	6.060	6.456	10.110	8.820	9.460
7	11.207	7.956	8.205	12.596	11.109	11.631
8	13.733	10.046	10.083	15.052	13.314	13.755
9	16.252	12.112	12.078	17.397	15.241	15.715
10	18.911	14.379	14.267	19.887	17.345	17.773

Note: 1. The MSEs are computed from 1987:Q2 for a window size of 40 years and from 1972:Q2 for a window size of 25 years.

IV. Out-of-Sample Predictive Power of the Cyclical Components

Clack and West (2007) Test

$$TSP\ Predictive\ Regression:\ \Delta y_{t-h:t} = \alpha_0 + \alpha_1\, c_{t-h}^{TSP} + u_{t-h:t}$$
$$DSP\ Predictive\ Regression:\ \Delta y_{t-h:t} = \alpha_0 + \alpha_2\, c_{t-h}^{DSP} + u_{t-h:t}$$
$$Random\ Walk\ Model:\ \Delta y_{t-h:t} = \alpha_0 + u_{t-h:t}$$

h	*Window Size: 40 Years*		*Window Size: 25 Years*	
	TSP Regression vs. Random Walk	*DSP Regression vs. Random Walk*	*TSP Regression vs. Random Walk*	*DSP Regression vs. Random Walk*
	$H_0: MSE^{RW} = MSE^{TSP}$ $H_a: MSE^{RW} > MSE^{TSP}$	$H_0: MSE^{RW} = MSE^{DSP}$ $H_a: MSE^{RW} > MSE^{DSP}$	$H_0: MSE^{RW} = MSE^{TSP}$ $H_a: MSE^{RW} > MSE^{TSP}$	$H_0: MSE^{RW} = MSE^{DSP}$ $H_a: MSE^{RW} > MSE^{DSP}$
1	-1.327	1.613*	-0.472	2.647**
2	-1.234	1.799**	0.004	2.372**
3	-1.223	1.712**	0.488	2.140**
4	-1.168	1.836**	0.746	2.146**
5	-1.112	1.884**	0.855	2.126**
6	-1.036	1.977**	0.879	1.860**
7	-0.963	2.210**	0.888	1.419*
8	-0.886	0.967	0.895	1.203*
9	-0.819	-0.216	0.872	1.246*
10	-0.741	-0.488	0.871	1.344*

Note: 1. The critical values of the CW test statistic are 1.28 and 1.64 for 10% and 5% significance levels, respectively. Two asterisks represent statistical significance at 5 % and one asterisk represents significance at 10 %.

Summary and Contributions

1. **ML Approach is severely subject to the 'pile-up' problem and the probability of the pile-up problem is higher as the model gets more complicated. However, Bayesian Approach is much more robust to the pile-up problem.**

2. **This paper suggests that the two competing models (the TSP and the DSP models) for the postwar log of real GDP should be examined within the Bayesian framework, as it is relatively free from the pile up problem. However, Bayesian in-sample model comparison does not provide us with any decisive evidence in favor of either model.**

3. **There exists convincing evidence that the DSP cycle, which is small in magnitude and noisy, has out-of-sample predictive power for future output growth and has information beyond the historical means for output growth. We interpret this result as evidence supporting the DSP implication for the log of postwar real GDP, as in Nelson and Plosser (1982), Morley et al. (2003), and Nelson (2008).**