Trend-Cycle Decompositions of Post-war Real GDP Revisited: Classical and Bayesian Perspectives

by

Chang-Jin Kim
University of Washington
and
Jaeho Kim
University of Oklahoma

This version: April, 2016

Abstract

While maximum likelihood estimation of an ARIMA model indicates that postwar U.S. real GDP is a trend stationary process (TSP), Bayesian estimation of the same model suggests that it is a difference stationary process (DSP) with a stochastic trend component explaining most of the variations in real GDP. This paper suggests that the two competing models of real GDP should be examined within the Bayesian framework, as it is relatively free from the pile-up problem.

Our empirical results based on Bayesian model comparison does not provide us with any decisive evidence in favor of either model. However, there exists convincing evidence that the cycle from the DSP model, but not from the TSP model, has out-of-sample predictive power for future output growth at short horizons and has information beyond the historical means for output growth. We argue that the highly persistent TSP cycle without any predictive power may be related to spurious periodicity discussed in Nelson and Kang (1981).


1 Chang-Jin Kim: Dept. of Economics, University of Washington, Seattle, WA (E-mail: changjin@u.washington.edu); Jaeho Kim: Department of Economics, University of Oklahoma, Norman, OK (E-mail: jaeho@ou.edu). Chang-Jin Kim acknowledges financial support from the Bryan C. Cressey Professorship at the University of Washington. Jaeho Kim acknowledges financial support from the Grover and Creta Ensley Fellowship in Economic Policy at the University of Washington. We thank Dukpa Kim, Charles R. Nelson, Richard Startz, and Eric Zivot for helpful comments.
1. Introduction

Since the seminal work of Nelson and Plosser (1982), one of the important issues in empirical macroeconomics has been to investigate the degree of persistence in real economic activity or the relative importance of the stochastic trend and the cyclical components in real GDP. However, researchers have provided conflicting evidence.

Based on estimation of ARMA models for real output growth, Nelson and Plosser (1982) and Campbell and Mankiw (1987) conclude that transitory shocks are relatively unimportant in explaining the dynamics of real output, while permanent shocks must dominate. On the contrary, within an unobserved-components model (hereafter, UC model) framework in which the permanent and transitory shocks are assumed to be uncorrelated, Clark (1987) reports evidence that a significant portion of real GDP is explained by the cyclical component. This result is then challenged by Morley et al. (2003), who show that the stochastic trend explains most of the variations in real GDP once the assumption of zero correlation between the permanent and the transitory shocks is dropped. They further show that the Beveridge-Nelson decomposition of real GDP based on an ARIMA(2,1,2) model and that based on an unobserved-components model are identical.

Recently, by allowing for a structural break in the long-run mean growth rate of real GDP in the mid-1970s, Perron and Wada (2009) show that the log of U.S. real GDP [1947:Q2-1998:Q4] follows a trend stationary process (TSP) and that variations in real GDP are ascribed mostly to the cyclical component. In particular, by estimating an ARIMA(2,1,2) model for the log of real GDP based on maximum likelihood estimation, they show that the point estimates of the moving-average coefficients sum to unity and thus, the data are over-differenced.

However, a later section will show, Bayesian estimation of the same model provides us very different results, even when we employ reasonably uninformative priors. It leads to an estimate of the moving-average root which is considerably smaller than unity, implying that real GDP follows a difference stationary process (DSP). The trend-cycle decomposition obtained from the Bayesian approach implies that most of the variations in real GDP can be explained by the stochastic trend component, consistent with the implications of Nelson and Plosser (1982) and Morley et al. (2003).
One goal of this paper is to provide a comprehensive analysis of why classical and Bayesian inferences of the same ARIMA model may result in such dramatic differences in empirical results. We first confirm existing literature which reports that the maximum likelihood approach suffers from the ‘pile-up problem,’ which suggests that there exists non-zero probability that the moving-average root may be estimated to be 1 in a finite sample even when its true value is less than 1. We then show that, while the maximum likelihood approach suffers from higher probability of the pile-up problem as the model gets more complicated, the Bayesian approach is relatively free from the pile-up problem. We also provide a discussion on the source of such differences.

Another goal of this paper is to re-investigate the trend-cycle decompositions for the log of postwar real GDP (1947:Q1-2014:Q4) within the Bayesian framework, which is relatively free from the pile-up problem. For this purpose, we estimate both the TSP and DSP models for the log of real GDP using the Markov-chain monte carlo (MCMC) methods, by incorporating one or two structural breaks in the long-run mean growth rate. As the unobserved components model that involves stochastic trend and the stationary cycle may potentially suffer from the identification problem, we estimate an ARIMA model with invertible moving roots (DSP model) and with a non-invertible moving average root (TSP model). For Bayesian model selection based on in-sample fits, we employ the Deviance Information Criterion (DIC) by Spiegelhalter et al. (2002).

Nelson (2008) argues that predictability should be an essential criterion in choosing between the two competing models of trend-cycle decomposition, as the in-sample criteria could deliver misleading messages. Following Nelson (2008), we also compare the out-of-sample predictive power of the cyclical components implied by the two competing models. In particular, we evaluate whether these cyclical components contain information about future output growth and has information beyond the historical means for output growth.

A brief summary of empirical results are as follows. Bayesian model comparison based on in-sample fits does not provide us with any decisive evidence in favor of either model. This result 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 when postwar quarter data are employed. However, we show that the cyclical component obtained
from the DSP model, but not from the TSP model, has out-of-sample predictive power for future output growth. In particular, while the DSP predictive regression beats the random walk model at short prediction horizons, the TSP model cannot beat the random walk model at any of the horizons considered. We interpret these results as evidence supporting the DSP implication for the log of real output, as in Nelson and Plosser (1982), Morley et al. (2003), and Nelson (2008).

In Section 2, we show that results from Bayesian estimation of an ARIMA(2,1,2) model for the postwar real GDP data are very different from those of maximum likelihood estimation. In Section 3, we provide answers to the question of why empirical results from the classical and the Bayesian approaches can be so different. In Section 4, we perform Bayesian model comparison between the two competing models based on in-sample fits. Section 5 compares the out-of-sample predictive power of the cyclical components obtained from the two competing models. Section 6 concludes the paper.


2.1. Structural vs. Reduced-Form Models: Identification Issues

Consider the following unobserved components model for the log of real GDP:

\[ y_t = x_t + z_t, \]
\[ x_t = \mu_t + x_{t-1} + v_t \quad (1) \]
\[ z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \epsilon_t - \theta_0 \epsilon_{t-1} \]
\[ \begin{bmatrix} v_t \\ \epsilon_t \end{bmatrix} \sim i.i.d. N \begin{bmatrix} \sigma_v^2 & \rho \sigma_v \sigma_\epsilon \\ \rho \sigma_\epsilon \sigma_v & \sigma_\epsilon^2 \end{bmatrix}, \]

where \( y_t \) is the log of real GDP; \( \mu_t \) is the long-run mean growth rate; \( x_t \) is a stochastic trend component; and \( z_t \) is a cyclical component with all the roots of \((1 - \phi_1 L - \phi_2 L^2) = 0\) and \((1 - \theta_0 L) = 0\) lying outside the complex unit circle.
Under the maintained assumption that $\theta_0 = 0$, the literature suggests that different specifications for the long-run mean growth rate $\mu_t$ or a restriction on the correlation coefficient $\rho$ can lead to different trend-cycle decompositions. For example, with a zero restriction on the $\rho$ parameter and a random walk specification for $\mu_t$, Clark (1987) estimates the cyclical component ($z_t$) to be highly persistent and shows that a significant portion of real GDP is explained by this component. By assuming that $\mu_t$ is constant and allowing for a possibility that $\rho$ may be non-zero, Morley et. al (2003) estimates the cyclical component to be noisy and considerably smaller than that in Clark (1987). On the contrary, by modeling $\mu_t$ as a constant interrupted by a permanent change occurring in 1973:Q1, Perron and Wada (2009) estimate the variance of the permanent shocks $\sigma^2_v$ to be zero, suggesting that real GDP is a trend stationary process. More recently, Luo and Startz (2014) show Bayesian estimation of the above model favors the position of Perron and Wada (2009) that the variance of the trend component of GDP is relatively small once a structural break is allowed in the mean growth rate, even though the evidence is much less than decisive.

Potential drawbacks for estimating the above structural unobserved-components model are: i) the confidence interval for the $\rho$ parameter is so large that trend-cycle decomposition based on the model involves very large uncertainty; and ii) the model is not identified once we drop the maintained assumption that $\theta_0 = 0$. One way to overcome these limits is to estimate a reduced-form ARIMA model for the above structural unobserved components model and employ the Beveridge-Nelson (1981) decomposition for the trend-cycle decomposition. For example, if we assume that $\theta_0 \neq 0$ and $\mu_t$ is the long-run mean growth rate with a permanent shift, a reduced-form ARIMA model is derived as:

\[
\text{Reduced-form ARIMA(2,1,2) Model}
\]

\[
\Delta y_t = \mu_t + \Delta y^*_t,
\]

\[
\Delta y^*_t = \phi_1 \Delta y^*_{t-1} + \phi_2 \Delta y^*_{t-2} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2},
\]

\[
e_t \sim i.i.d. N(0, \sigma^2_e),
\]

where $\sigma^2_e$, $\theta_1$, and $\theta_2$ are functions of $\phi_1$, $\phi_2$, $\theta_0$, $\sigma^2_v$, $\sigma^2_e$, and $\rho$. 

5
2.2. Maximum Likelihood and Bayesian Estimation Results: Comparison

Perron and Wada (2009) provide maximum likelihood estimation results for the above reduced-form ARIMA(2,1,2) model using the sample covering the period 1947:Q1 - 1998:Q2, by specifying the long-run mean growth rate $\mu_t$ as:

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

(3)

$$
S_t = \begin{cases} 
1, & \text{if } t \geq 1973 : Q1; \\
0, & \text{otherwise}.
\end{cases}
$$

Their results are replicated in Table 1.A. Note that the maximum likelihood estimate for $\theta_1 + \theta_2$ is 1, which suggests that $\sigma_v^2 = 0$ in the structural unobserved component model in equation (1). That is, the maximum likelihood estimation results indicate that the log of real GDP is a trend stationary process (TSP), of which all the variation in the log of real GDP is explained by the cyclical component. The cyclical component implied by the maximum likelihood estimation of the model, which is very persistent and large in magnitude, is depicted in the upper panel of Figure 1.

Within the Bayesian framework in which we employ reasonably flat priors, however, we have very different results. The posterior mean of $\theta_1 + \theta_2$ is 0.311 and the posterior mode is 0.460, as reported in Table 1.B. This indicates that the log of real GDP is a difference stationary process (DSP). The cyclical component implied by this Bayesian estimation is illustrated in the lower panel of Figure 1. It is small in magnitude and noisy, indicating that the stochastic trend component explains most of the variation. The results are in line with in line with Nelson and Plosser (1982) and Morley et al. (2003).

In the case of a flat prior, a conventional wisdom is that Bayesian inference is not very different from classical inference, as the likelihood dominates the posterior density. An important question then is: “Why do classical and Bayesian approaches produce such strikingly different estimates of the ARMA parameters and trend-cycle decompositions?” The next section provides answers to this question.

3. Why Are the Results from Classical and Bayesian Approaches so Different?
The Bernstein von Mises theorem ensures that, in nicely behaved models, the log-likelihood becomes asymptotically quadratic and dominates the prior. In this case, standard maximum likelihood theory implies that Bayesian and frequentist inference coincide. In small samples or less nicely behaved models, however, such approximate equivalence does not always hold. \footnote{Well known examples include weak instruments, autoregressive roots close to unity and partially identified models.}

In what follows, we show that the dramatically different results in the previous section stem from the differences in how the nuisance parameters are handled within the classical and the Bayesian frameworks. We illustrate that, when the likelihood is not approximately jointly quadratic in a finite sample, integrating out nuisance parameters over a flat prior (the integrated likelihood) and maximizing out nuisance parameters (the profiled likelihood) are not necessarily close. \footnote{We define parameters of an ARIMA model other than the MA coefficients as nuisance parameters.} Thus, for the maximum likelihood approach, there exists non-zero probability that the moving-average root may be estimated to be 1 even when its true value is less than 1. This phenomenon is called the ‘pile-up problem’ in the literature. The Bayesian approach, however, is relatively free from this pile-up problem, as the posterior distribution is not dependent upon the nuisance parameters.

Section 3.1 graphically illustrates these points. Section 3.2 performs Monte Carlo experiments showing that, as the number of the nuisance parameters increases in a finite sample, the probability of the pile-up problem increases for the maximum likelihood approach, while the Bayesian approach is relatively free from this problem.

3.1. The Source of Differences Between Classical and Bayesian Results

Consider the following MA(1) process:

\begin{equation}
y_t = \mu + e_t - \theta e_{t-1}, \quad e_t \sim \text{i.i.d.} \mathcal{N}(0, \sigma^2), \quad t = 1, 2, \ldots, T,
\end{equation}

\[|\theta| < 1,\]

where \(\sigma^2\) is assumed to be known. Following Smith and Naylor (1987) and Berger et al. \footnote{Well known examples include weak instruments, autoregressive roots close to unity and partially identified models.}
(1999), we can explain the reason for a potential divergence between Bayesian and classical inferences by comparing the profile likelihood and the flat-prior posterior density (or integrated likelihood), given below:

\[
\text{Profile Likelihood} : \hat{L}(\theta) = \sup_{\mu} L(\theta, \mu), \quad (5)
\]

\[
\text{Posterior Density (Integrated Likelihood)} : L(\theta) = \int L(\theta, \mu) d\mu, \quad (6)
\]

where \(L(\theta, \mu)\) is the likelihood function.

The posterior density of \(\theta\) is not dependent upon the nuisance parameter, as it is obtained by integrating the likelihood function with respect to the nuisance parameter. However, this is not always the case for profile likelihood, as it is obtained by maximizing likelihood function with respect to the nuisance parameter. Pierce (1971) proves that, in a regression model with ARMA(1,1) disturbances, the maximum likelihood estimator of \(\theta\) and the regression coefficients (in our case, \(\mu\)) are correlated in a finite sample, even though they are asymptotically independent and jointly normal. Thus, the likelihood function may not be quadratic, making the shape of the profile likelihood different from that of the flat-prior posterior distribution in a small sample. This suggests that the posterior mode at the peak of the posterior density may be different from the maximum likelihood estimate at the peak of the profile likelihood.

In order to illustrate how the profile likelihood and the posterior density can be different in a small sample, we generate many arbitrary data sets by assigning the following parameter values for the model in equation (3):

\[
\mu = 1, \quad \theta = 0.8, \quad \sigma^2 = 1, \quad T = 50
\]

Then, for each data set generated, we apply the maximum likelihood and the Bayesian estimation procedures to obtain \(\hat{\theta}_{ML}\) and the posterior distribution of \(\theta\), by assuming \(\sigma^2\) to be known. Through this simulation experiment, we find that the posterior distributions for \(\theta\) do not always have peaks at the maximum likelihood estimate. Instead, we have three categories of shapes for the posterior distributions, which are shown in Figure 2 and described below:
Type #1: $\hat{\theta}_{ML}$ is within the invertible region, and the peak of the posterior distribution is around $\hat{\theta}_{ML}$;
Type #2: $\hat{\theta}_{ML} = 1$ and the peak of the posterior distribution is at $\theta = 1$;
Type #3: $\hat{\theta}_{ML} = 1$ but the peak of the posterior distribution is at the invertible region.

We take a representative sample for each of the above three types and draw a three-dimensional likelihood surface as a function of $\mu$ and $\theta$, by fixing $\sigma^2$ at its true value. They are drawn in Figure 2.A, 2.B, and 2.C, along with the corresponding profile likelihood and flat-prior posterior distribution for $\theta$. For Type #1 and Type #2 in Figures 2.A and 2.B, in which the posterior distributions peak at $\hat{\theta}_{ML}$, the shapes of the posterior density and the likelihood are very similar.

For Type #3, however, the posterior distribution peaks at around the true value of 0.8, while the maximum likelihood estimate of $\theta$ is 1. It shows that the most likely value of $\theta$ may not always be the ones near the maximum likelihood estimate. This explains the differences between the Bayesian and classical inferences in the empirical ARIMA (2,1,2) model of real GDP reported in Tables 1.A and 1.B. These tables suggests that the posterior distribution of $\theta_1 + \theta_2$ peaks at the local maximum of the likelihood function (an invertible region) and not at the maximum likelihood estimate of 1.

3.2. The Nature of the Pile-Up problem within the Classical and the Bayesian Frameworks

Many authors investigate the finite sample properties of the maximum likelihood estimator of the moving average parameter in an MA(1) model, especially when the moving average parameter is close to unity. Following the initial work of Kang (1975), several authors including Sargan and Bhargava (1983), Anderson and Takemura (1986) and Tanaka and Satchell (1989) show that the process can be estimated to be noninvertible with a unit root in the MA part even when the true process is invertible. This is referred to as a pile-up problem and it could occur with a high probability in a finite sample.  

\footnote{Asymptotic properties of $\hat{\theta}_{ML}$ are derived in Davis and Dunsmuir (1996), and Davis et}
However, the issue of the pile-up problem has not been investigated rigorously within the Bayesian framework. The only Bayesian paper on the pile-up problem that we know of is DeJong and Whiteman (1993), who show that the posterior distributions of $\theta$ do not pile up at unity regardless of the proximity of $\theta$ to unity under a simple MA(1) model without intercept. In this section, we carry out a simulation study to show how the probability of the pile-up problem is affected by model complexity, sample sizes, within both the classical and the Bayesian frameworks.

For this purpose, we consider the following four data generating processes:

**Model #1: MA(1) without Intercept**

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

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

**Model #2: MA(1) with Intercept**

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

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

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

$$y_t = \mu + \mu_1 S_t + e_t - \theta e_{t-1}, \quad e_t \sim i.i.d. N(0, \sigma^2),$$

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

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

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

$$y_t = \mu_0 + \mu_1 S_t + u_t,$$

al. (1995) for the case where $\theta$ is close or equal to 1. They show that the conventional central limit theorem does not work in such a case.
\[ u_t = \phi u_{t-1} + e_t - \theta e_{t-1}, \quad e_t \sim i.i.d. N(0, \sigma^2), \]

\[ S_t = 0, \quad \text{for } t \leq \frac{T}{2}; \quad S_t = 1, \quad \text{otherwise}, \]

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

For each of the above models, we generate 5,000 sets of data and apply the maximum likelihood and Bayesian estimation procedures to the generated data sets. Then, we obtain sampling distributions for the maximum likelihood estimate of \( \theta (\hat{\theta}_{ML}) \) and the posterior mode of \( \theta (\hat{\theta}_{mode}) \), which are used to calculate the probabilities of the pile-up problem.

We consider three different sample sizes \((T = 50, 100, \text{and } 200)\). Maximization of the log likelihood function is performed using the Gauss optimization package. For Bayesian inference, the model is estimated based on the MCMC algorithm proposed by Chib and Greenberg (1994). We impose the constraint that \(|\theta| < 1\) in both estimation procedures and report \( Pr[0.995 < \hat{\theta}_{ML} < 1] \) and \( Pr[0.995 < \hat{\theta}_{mode} < 1] \) as the probabilities of the pile-up problem, following DeJong and Whiteman (1993).

Results for the maximum likelihood estimation approach as reported in the second column of Table 2 show that the pile-up problem is a small sample property. The probability of the pile-up problem decreases as the sample size increases for a given model. For a given sample size, the probability of the pile-up problem increases as the model gets more complicated. For a MA(1) model with a constant intercept, the simplest model under consideration, the pile-up problem that exists when \( T = 50 \) almost disappears when \( T = 200 \). However, for an ARMA(1,1) model with a structural break in mean, the most complicated model under consideration, the probability of the pile-up problem remains as high as 23.6% even when the size is increased up to 200.

For Bayesians, there is only one realization of the data set, so contemplating the probability of the pile-up problem in repeated sampling may be conceptually irrelevant. However, in order for us to be able to evaluate and directly compare the probabilities of the pile-up problem for both the Bayesian and classical approaches, we define \( \hat{\theta}_{mode} \) as a Bayesian estimator, which is treated as a random variable in repeated samples.

Even though we do not report the results here, our simulation study also shows that the pile-up problem for ARMA models gets worse as we assign the true value of the autoregressive parameter \((\phi)\) closer to that of the moving average parameter \( \theta \) when generating data. In such cases, there exist higher probabilities for the cancellation of the estimated MA and AR roots, and this tends to make the pile-up problem worse.
In order to consider the implication of the pile-up problem for Perron and Wada’s (2009) maximum likelihood estimation of the model in equations (2) and (3), we conducted additional Monte Carlo experiments. For this purpose, we generated 5,000 sets of data according to the data generating process in (2) and (3), by assuming that the posterior modes reported in Table 1.B are the true parameter values. The sample size is set to be the same as the number of observations (\(T = 205\)) in the sample employed by Perron and Wada (2009). We then apply the maximum likelihood estimation procedure to the generated data sets. The estimator for \(\theta_1 + \theta_2\) piles up at unity, and the probability of the pile-up problem turns to be almost 0.4.

The last column of Table 2 reports the results for the Bayesian approach. When the sample size is as small as 50, the probability of the pile-up problem is non-negligible for all the models even though they are much smaller than the case of maximum likelihood estimation. When the sample size is increased to 200, however, the pile-up problem almost disappears for all the models. That is, the Bayesian approach is relatively free from the pile-up problem when the sample size is reasonably large.

Monte Carlo experiment results in this section suggest that, when the classical and the Bayesian approaches lead to conflicting results such as those in Section 2, we may confer more credibility to the Bayesian inference. Alternatively, the two competing implications (the TSP and the DSP implications) of the log of real GDP should be examined within the Bayesian framework. The next session deals with this issue.


4.1. Model Specifications

In this section, we estimate two competing models of TSP and DSP for the log of real GDP within the Bayesian framework. We employ the Bayesian approach because it is

---

7 In order to consider the pile-up problem, Perron and Wada (2009) apply the median-unbiased estimation procedure of Stock and Watson (1998) to the detrended real GDP series. However, as Nelson and Kang (1981, 1984) present, in the case where the data generating process is a difference stationary process, the dynamic properties of the detrended series are artifactual.
relatively free from the pile-up problem, as the previous section shows. The two models are
given below:

**DSP Model: ARMA with Invertible MA Roots**

\[
\Delta y_t = \mu_t + \Delta y_t^* \\
(1 - \phi_1 L - \phi_2 L^2)\Delta y_t^* = (1 - \theta_1 L - \theta_2 L^2)e_t, \quad e_t \sim N(0, \sigma_t^2),
\] (7)

**TSP Model: ARMA with a Non-invertible MA Root**

\[
\Delta y_t = \mu_t + \Delta y_t^* \\
(1 - \phi_1 L - \phi_2 L^2)\Delta y_t^* = (1 - \theta_0 L)(1 - L)e_t,
\] (8)

where \( \Delta y_t \) is the log difference of real GDP; all the roots of \( 1 - \theta_1 L - \theta_2 L^2 = 0 \) lie outside
the complex unit circle and \( |\theta_0| < 1 \). One can easily show that the TSP model above can be
derived when we have \( \sigma_t^2 = 0 \) for the structural model in equation (1).

To account for the changing nature of the volatility for real GDP growth, we assume the
following stochastic volatility process for \( e_t \):

\[
\ln(\sigma_t^2) = \ln(\sigma_{t-1}^2) + \epsilon_t, \quad \epsilon_t \sim i.i.d N(0, \sigma^2) 
\] (9)

In order to account for the changing nature for the long-run growth rate of real GDP, we
allow for one or two structural breaks with unknown break points in \( \mu_t \) parameter. This is
done by employing the following specifications suggested by Chib (1998):

**One Break in Mean**

\[
\mu_t = \mu_0 + \mu_d I_{S_t=1}, \quad S_t = 0, 1, 
\] (10)

\[
I_{S_t=1} = \begin{cases} 
1, & \text{if } S_t = 1; \\
0, & \text{otherwise.} 
\end{cases}
\]

\[
0 < p_{00} < 1 \quad p_{11} = 1
\]
Two Breaks in Mean

\[ \mu_t = \mu_0 + \mu_d I_{S_t=1,2} + \mu_dd I_{S_t=2}, \quad S_t = 0, 1, 2, \]

\[ I_{S_t=1,2} = \begin{cases} 
1, & \text{if } S_t = 1 \text{ or } S_t = 2; \\
0, & \text{otherwise.}
\end{cases} \]

\[ I_{S_t=2} = \begin{cases} 
1, & \text{if } S_t = 2; \\
0, & \text{otherwise.}
\end{cases} \]

\[ 0 < p_{00} < 1, \quad p_{02} = 0; \quad p_{10} = 0, \quad 0 < p_{11} < 1; \quad p_{22} = 1, \]

where \( p_{ij} = Pr[S_t = j|S_{t-1} = i] \).

To estimate the above models, we employ an MCMC algorithm developed by Kim and Kim (2015). For Bayesian model comparison, we employ the deviance information criterion (DIC) developed by Spiegelhalter et al. (2002). The DIC consists of two components: a measure for in-sample model fit and a term for model complexity penalty such as AIC (Akaike information criterion) and BIC (Bayesian information criterion). It is well suited in comparing complex hierarchical models like our reduced-form ARIMA models with structural break in mean and stochastic volatility. A model with smaller DIC is preferred to a model with larger DIC. A more detailed exposition of the DIC is provided in appendix.

4.2. Empirical Results

The posterior moments of the parameters for the two competing models are reported in Tables 3.A and 3.B. The sample covers the period of 1947:Q2-2014:Q4. The posterior mean of \( \phi_1 + \phi_2 \) from the TSP model with one break in long run growth is 0.982 and is 0.946 for a model with two breaks. These indicate highly persistent cycles, as shown in the left panels of Figure 3.A. On the contrary, the posterior mean of \( \theta_1 + \theta_2 \) from a DSP model with one break in long run growth is 0.384 and is -0.023 for a model with two breaks. These indicate that the largest MA root is estimated to be far less than unity when there is constraint imposed on the MA root. The cyclical components of real GDP depicted on the right panels of Figure 3.A show that they are small in magnitude and noisy.

---

8 Luo and Startz (2014) consider the possibility of two potential breaks in the mean growth rate of postwar real GDP.
Inferences on the nature changes in the long-run growth rate of real GDP seem to be robust for the two competing models, as shown in Figure 3.B. For both TSP and DSP models, and regardless of the number of breaks incorporated, declines in the long-run mean growth between 1947 and late 1990s seem to be gradual. However, the declines that happened between the early and mid 2000’s seem to be more abrupt and sharp.

Based on the Deviance information criterion (DIC), a TSP model with one structural break in long run growth does not seem to be preferred to the rest of the models. The DIC measure for this model ($DIC = 661.606$) is considerably larger than those for the other models. However, the DIC measures for the other three models seem to be very similar. Even though a DSP model with one break ($DIC = 654.973$) is slightly preferred to a DSP model with two breaks ($DIC = 655.428$), it is almost indistinguishable from a TSP model with two breaks ($DIC = 654.918$). When we estimated the two competing models with one structural break in long run growth for the sample covering 1947:Q2-1998:Q4 (Perron and Wada (2009) sample), we could not distinguish between the two competing models based on the deviance information criterion, either.

Empirical results in this section lead us to conclude that Bayesian model comparison does not provide us with any decisive evidence in favor of either TSP or DSP model. 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 when postwar quarter data are employed.

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

5.1. Motivation and the Design for Out-of-Sample Prediction

Recently, based on a Bayesian model averaging approach to trend-cycle decomposition, Luo and Startz (2014) provides empirical evidence which favors the position of Perron and Wada (2009) that the variance of the trend component of GDP is relatively small once a structural break is allowed in the long run growth. They mention, however, that the evidence is much less than decisive. Our results based on Bayesian model comparison does not provide any decisive evidence in favor of either the TSP or DSP model. Note that both Luo and
Startz’s (2014) results and ours are based on in-sample fitting of the models.

As Nelson (2008) argues, if the measured cycle component is temporary then it should predict future growth rates of the opposite sign. He further states that “Predictability of the cycle implies a metric then for measuring the effectiveness of alternative decompositions: how well do they predict future growth ...?” [page 203]. Trend-cycle decomposition based on a better in-sample fit does not necessarily result in a cycle component that has higher predictive power for future growth rates. For example, Nelson and Kang (1981) show that when a linear trend model is estimated with the generated data from a pure random walk process, much variations in the generated series are attributed to the persistent cycle component to maximize an in-sample model fit, even though this estimated cycle does not have any predictive power. They name the high persistence in the estimated cycle as ‘spurious periodicity’ that results from inappropriately detrended time series. What follows describes a procedure for evaluating out-of-sample predictive power of cycle measures.

First, we obtain posterior means for the cyclical components implied by the two competing models in equations (7) and (8), using 40 year rolling windows. We allow for one structural break in the long run growth rate in each rolling sample. 9 Second, in order to obtain out-of-sample $h$-horizon real GDP growth, we estimate the following predictive regressions by OLS and obtain out-of-sample prediction for each rolling sample:

**Random Walk Model**

\[
\Delta y_{t-h:t} = \alpha_0 + u_{t-h:t}, \quad t = \tau_j + 1, \tau_j + 2, \ldots, \tau_j + \tau, \tag{12}
\]

**Prediction**: \( E[\Delta y_{\tau_j + \tau + h|I_{\tau_j + \tau}}] = \alpha_0 \),

**TSP Predictive Regression**

\[
\Delta y_{t-h:t} = \alpha_0 + \alpha_1 c_{t-h}^{TSP} + u_{t-h:t}, \quad t = \tau_j + 1, \tau_j + 2, \ldots, \tau_j + \tau, \tag{13}
\]

**Prediction**: \( E[\Delta y_{\tau_j + \tau + h|I_{\tau_j + \tau}}] = \alpha_0 + \alpha_1 c_{\tau_j + \tau}^{TSP} \).

---

9 When estimating the models for each rolling sample, we employ the following priors: \( \phi_1 + \phi_2 \sim N(0.5, 0.25^2) \) for both the TSP and DSP models; \( \theta_1 + \theta_2 \sim N(0.5, 0.25^2) \) for the DSP model; and \( \theta_0 \sim N(0, 0.25^2) \) for the TSP model. We note that empirical results are pretty robust with respect to priors employed, even though non-informative priors may lead to meaningless results for Bayesian model comparisons.
\[ \Delta y_{t-h:t} = \alpha_0 + \alpha_2 c^{DSP}_{t-h} + u_{t-h:t}, \quad t = \tau_j + 1, \tau_j + 2, \ldots, \tau_j + \tau, \quad (14) \]

\textit{Prediction:} \quad E[\Delta y_{\tau_j+\tau+h}|I_{\tau_j+\tau}] = \alpha_0 + \alpha_1 c^{DSP}_{\tau_j+\tau},

where \( \Delta y_{t-h:t} = y_t - y_{t-h} \) is the real GDP growth rate between \( t - h \) and \( t \) and \( h \) refers to the prediction horizon; \( c^{TSP}_t \) and \( c^{DSP}_t \) refer to the measures of cycles obtained for the TSP model and the DSP model, respectively (hereafter, TSP cycle and DSP cycle, respectively). Here, \( \tau_j + 1 \) refers to the beginning date of the rolling sample and \( \tau \) is the window size.

\section*{5.2. Empirical Results}

Table 4 reports the average of the \( R^2 \) values obtained by the OLS regressions of equations (13) and (14) across rolling samples, for given prediction horizon \( h \) (\( h = 1, 2, \ldots, 10 \)). The same information is illustrated in Figure 4. The average \( R^2 \) value for the TSP predictive regression is 0.022 for \( h = 1 \), and it increases with \( h \), reaching as high as 0.400 when \( h = 10 \). On the contrary, the average \( R^2 \) value for the DSP predictive regression is 0.101 for \( h = 1 \), and it decreases with \( h \), reaching as low as 0.002 when \( h = 10 \). Thus, while the DSP cycle provides better in-sample fits at short horizons (\( h = 1, 2 \)), the TSP cycle provides better in-sample fits at longer horizons (\( h \geq 3 \)).

Results for out-of-sample predictive performance are different from those for in-sample predictive performance. Table 5.A reports out-of-sample mean squared prediction errors (MSEs) based on equations (13) and (14), along with the random walk model in equation (12). Figure 5 provides a graphical illustration of the same information. The TSP predictive regression in equation (13) has the worst out-of-sample predictive performance, in spite of its high \( R^2 \) values and superior in-sample fits at long prediction horizons. Then comes the random walk model in equation (12), and the DSP predictive regression in equation (14) performs the best. These results hold regardless of the prediction horizons considered.

To formally test the null of equal predictive ability of equation (13) or (14) and a random walk model in equation (12), we apply the Clark and West (2007) test. Table 5.B presents
test results. When the null and alternative models are equations (12) and (13), respectively, we do not reject the null model at any prediction horizons, suggesting that TSP cycle does not have any predictive power beyond what is contained in historical means for output growth. However, when the random walk null model is tested against equation (14), we reject the null model at the 5% significance level, for prediction horizons 1 through 7. That is, there exists convincing evidence that the DSP cycle, unlike the TSP cycle, has predictive power for future real GDP growth. We interpret these results as evidence supporting the DSP implication for the postwar log of real output, as in Nelson and Plosser (1982), Morley et al. (2003), and Nelson (2008).

In the meantime, the absence of predictive power for the TSP cycle leads us to conclude that large and highly persistent cyclical component implied by the TSP model may be related to ‘spurious periodicity’ discussed in Nelson and Kang (1981). Furthermore, in-sample $R^2$ value that increase with the prediction horizon for the TSP predictive regression may be spurious. When the prediction horizon $h$ increases, the persistence of the dependent variable in equation (13) also increases. This, combined with the persistent TSP cycle, results in spurious regression, leading to $R^2$ values that increase with the prediction horizon. Thus, large and persistent cycle from the TSP model, which seemingly has predictive power in retrospect, may only be statistical artifacts in the terminology of Nelson (2008).

6. Summary and Conclusion

Univariate trend-cycle decompositions of postwar real GDP based on the maximum likelihood estimation and the Bayesian approaches lead to conflicting results. We show that such conflicting results stem from the differences in how the nuisance parameters are handled within the classical and Bayesian frameworks. When the likelihood is not approximately jointly quadratic in a finite sample, integrating out nuisance parameters over a flat prior (the integrated likelihood) and maximizing out nuisance parameters (the profiled likelihood) are not necessarily close.

Even though the random walk model provides better out-of-sample performance than the TSP predictive regression model, the difference is statistically insignificant.
This paper suggests that the two competing models (the TSP and the DSP model) 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 model comparison does not provide us with any decisive evidence in favor of either model.

We evaluate out-of-sample predictive power of the cycles following Nelson (2008), who argues that predictability should be an essential criterion in choosing between the two competing models of the trend-cycle decomposition. There exists convincing evidence that the DSP cycle, but not the TSP cycle, 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). We also argue that the highly persistent TSP cycle without any predictive power may be related to spurious periodicity discussed in Nelson and Kang (1981).
Appendix. Deviance Information Criterion (DIC)

For Bayesian model comparisons in this paper, we employ the following Deviance Information Criterion (DIC) developed by Spiegelhalter et al. (2002):

\[
DIC = E_{\zeta|\Delta Y}[-2lnf(\Delta Y|\zeta)] + 2\{lnf(\Delta Y|\bar{\zeta}) - E_{\zeta|\Delta Y}[lnf(\Delta Y|\zeta)]\}
\]

where \(\Delta Y = \{\Delta y_1, \Delta y_2, ..., \Delta y_T\}\); \(\zeta\) represents all model parameters and latent stochastic volatility process; \(\bar{\zeta}\) is the posterior mean of \(\zeta\); \(E_{\zeta|\Delta Y}[\cdot]\) is the posterior expectation with respect to \(\zeta\) conditional on \(\Delta Y\); \(lnf(.|\zeta)\) is the log likelihood evaluated at \(\zeta\). The term \(D(\zeta) = -2lnf(\Delta Y|\zeta)\), called a deviance, is often used to measure a model’s goodness of fit in classical statistics. Likewise, the posterior expectation of the deviance denoted by \(E_{\zeta|\Delta Y}[-2lnf(\Delta Y|\zeta)]\) captures a Bayesian model fit. Because the log likelihood is multiplied by \(-2\), the smaller the posterior expectation of the deviance is, the better the Bayesian in-sample model fit is. The DIC imposes a penalty for model complexity defined as \(2\{lnf(\Delta Y|\bar{\zeta}) - E_{\zeta|\Delta Y}[lnf(\Delta Y|\zeta)]\}\) through the effective number of parameters. The idea is that as a model gets more complicated, the likelihood surface spreads out and covers more of the unlikely parameter space, reducing \(E_{\zeta|\Delta Y}[lnf(\Delta Y|\zeta)]\). Conversely, the log likelihood evaluated at \(\bar{\zeta}\) increases due to higher flexibility in the model. Therefore, the difference between \(lnf(\Delta Y|\bar{\zeta})\) and \(E_{\zeta|\Delta Y}[lnf(\Delta Y|\zeta)]\) increases, making the DIC to impose a heavier penalty on a more complex model. Therefore, a model with smaller DIC is preferred to a model with larger DIC.
References


21


Table 1.A. Maximum Likelihood Estimation of ARIMA(2,1,2) Model: a Known Break Point in Mean [Sample: 1947:Q1~1998: Q2]

\[ \Delta y_t = \mu_0 + \mu_d S_t + \Delta y_{t-1}^*, \]
\[ \Delta y_{t-1}^* = \phi_1 \Delta y_{t-1}^* + \phi_2 \Delta y_{t-2}^* + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}, \]
\[ \epsilon_t \sim i.i.d N(0, \sigma^2), \]

\[ S_t = 0 \text{ for } t \leq 1973: Q1, \quad S_t = 1 \text{ for } t > 1973: Q1. \]

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Global Maximum</th>
<th>Local Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimates</td>
<td>S.E.</td>
</tr>
<tr>
<td>( \mu_0 )</td>
<td>0.951</td>
<td>0.021</td>
</tr>
<tr>
<td>( \mu_d )</td>
<td>-0.287</td>
<td>0.038</td>
</tr>
<tr>
<td>( \phi_1 + \phi_2 )</td>
<td>0.921</td>
<td>0.020</td>
</tr>
<tr>
<td>( \phi_2 )</td>
<td>-0.601</td>
<td>0.109</td>
</tr>
<tr>
<td>( \theta_1 + \theta_2 )</td>
<td>0.999</td>
<td>0.003</td>
</tr>
<tr>
<td>( \theta_2 )</td>
<td>-0.283</td>
<td>0.137</td>
</tr>
<tr>
<td>( \sigma^2 )</td>
<td>0.876</td>
<td>0.086</td>
</tr>
</tbody>
</table>

| Long-run Impulse-Response | 0.000 | 0.042 | 1.228 | 0.312 |
| Log Likelihood | -278.930 | -282.710 |

Note:  
1. The data set used is same as in Morley et al. (2003) and Perron and Wada (2009).  
2. S.E. refers to the standard error.  
3. S.E. of the long-run impulse response is reported using the delta method.
Table 1.B. Bayesian Estimation of ARIMA(2,1,2) Model: a Known Break Point in Mean [Sample: 1947:Q1–1998:Q2]

\[
\Delta y_t = \mu_0 + \mu_t \Delta t + \Delta y_t^*,
\]

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

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

\[S_t = 0 \text{ for } t \leq 1973:Q1, \quad S_t = 1 \text{ for } t > 1973:Q1.\]

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Prior</th>
<th>Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>(\mu_0)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(\mu_t)</td>
<td>-0.4</td>
<td>1</td>
</tr>
<tr>
<td>(\phi_1 + \phi_2)</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>(\phi_2)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(\theta_1 + \theta_2)</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>(\theta_2)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(\sigma^2)</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Long-run Impulse-Response

|          | 1.356 | 1.420 | 0.334 | 0.899 | 1.893 |

Note: 1. The data set used is same as in Morley et al. (2003) and Perron and Wada (2009).
2. Burn-in / Total iterations = 5,000 / 55,000
3. S.D. refers to a standard deviation. HPDI refers to a highest posterior density interval.
Table 2. Probability of Pile-up Problem: Classical and Bayesian Approaches

\[ y_t - \mu_t = \phi y_{t-1} - \mu_{t-1} + \epsilon_t - \theta \epsilon_{t-1}, \quad \epsilon_t \sim i.i.d \text{ } N(0, \sigma^2) \]
\[ \mu_t = \mu_0 + \mu_a \delta_t \]

<table>
<thead>
<tr>
<th>MA(1) without Intercept</th>
<th>[ \theta = 0.8, \sigma^2 = 1, \phi = 0, \mu_0 = 0, \mu_a = 0 ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T = 50 )</td>
<td>0.119</td>
</tr>
<tr>
<td>( T = 100 )</td>
<td>0.016</td>
</tr>
<tr>
<td>( T = 200 )</td>
<td>0.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MA(1) with Intercept</th>
<th>[ \theta = 0.8, \sigma^2 = 1, \phi = 0, \mu_0 = 1, \mu_a = 0 ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T = 50 )</td>
<td>0.582</td>
</tr>
<tr>
<td>( T = 100 )</td>
<td>0.169</td>
</tr>
<tr>
<td>( T = 200 )</td>
<td>0.008</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MA(1) a Structural Break in Intercept</th>
<th>[ \theta = 0.8, \sigma^2 = 1, \phi = 0, \mu_0 = 1, \mu_a = -0.3 ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T = 50 )</td>
<td>0.867</td>
</tr>
<tr>
<td>( T = 100 )</td>
<td>0.447</td>
</tr>
<tr>
<td>( T = 200 )</td>
<td>0.044</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ARMA(1,1) with a Structural Break in Intercept</th>
<th>[ \theta = 0.8, \sigma^2 = 1, \phi = 0.3, \mu_0 = 1, \mu_a = -0.3 ]</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T = 50 )</td>
<td>0.939</td>
</tr>
<tr>
<td>( T = 100 )</td>
<td>0.730</td>
</tr>
<tr>
<td>( T = 200 )</td>
<td>0.236</td>
</tr>
</tbody>
</table>

Note:
1. For Bayesian estimates, we employ posterior modes.
2. Probability of pile-up problem is approximated by \( \text{Pr}[0.995 \leq \theta \leq 1] \) as in DeJong and Whiteman (1993).
Table 3.A: Bayesian Estimation of ARIMA(2,1,2) Models: One Unknown Break Point in Mean

\[
\Delta y_t = \mu_t + \Delta y^*_t, \\
\mu_t = \mu_0 + \mu_d I(S_t = 1),
\]

**TSP Model: a Non-invertible MA Root**

\[
1 - \phi_1 L - \phi_2 L^2 \Delta y^*_t = 1 - \phi_0 L - 1 - L \epsilon_t, \epsilon_t \sim N(0, \sigma^2_t),
\]

**DSP Model: Invertible MA Roots**

\[
1 - \phi_1 L - \phi_2 L^2 \Delta y^*_t = (1 - \phi_1 L - \phi_2 L^2) \epsilon_t, \epsilon_t \sim N(0, \sigma^2_t),
\]

\[
\ln \sigma^2_t = \ln \sigma^2_{t-1} + \epsilon_t, \epsilon_t \sim i.i.d. N(0, \sigma^2_t).
\]

<table>
<thead>
<tr>
<th>Prior</th>
<th>Posterior</th>
<th>Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
</tr>
<tr>
<td>(p_0)</td>
<td>0.99</td>
<td>0.01</td>
</tr>
<tr>
<td>(\mu_0)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>(\mu_d)</td>
<td>-0.4</td>
<td>1</td>
</tr>
<tr>
<td>(\phi_1 + \phi_2)</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>(\phi_2)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(\theta_1 + \theta_2)</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>(\theta_2)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>(\sigma^2_\epsilon)</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>(\sigma^2_\eta)</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

90% HPDI of Brake Dates

<table>
<thead>
<tr>
<th>Prior</th>
<th>Posterior</th>
<th>Posterior</th>
</tr>
</thead>
</table>

DIC

<table>
<thead>
<tr>
<th>Prior</th>
<th>Posterior</th>
<th>Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>661.606</td>
<td>654.973</td>
<td></td>
</tr>
</tbody>
</table>

Note: 1. Burn-in / Total iterations = 5,000 / 55,000
2. S.D. refers to a standard deviation. HPDI refers to a highest posterior density.
3. Range of break dates is based on 50% HPDI.
Table 3.B. Bayesian Estimation of ARIMA(2,1,2) Models: Two Unknown Break Points in Mean  

\[
\Delta y_t = \mu_t + \Delta y_{t-1}, \\
\mu_t = \mu_0 + \mu_d l S_t = 1, 2 + \mu_{dd} l S_t = 2 , \\
TSP Model: a Non-invertible MA Root
\]

\[
1 - \phi_1 L - \phi_2 L^2 \Delta y_{t-1} = \beta_0 L (1 - L) \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2), \\
DSP Model: Invertible MA Roots
\]

\[
1 - \phi_1 L - \phi_2 L^2 \Delta y_{t-1} = (1 - \beta_1 L - \beta_2 L^2) \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2), \\
\ln \sigma^2_t = \ln \sigma^2_{t-1} + \epsilon_t, \quad \epsilon_t \sim i.i.d. N(0, \sigma^2). 
\]

<table>
<thead>
<tr>
<th>Prior</th>
<th>TSP Model</th>
<th>DSP Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>$\varphi_{00}$</td>
<td>0.99</td>
<td>0.01</td>
</tr>
<tr>
<td>$\varphi_{11}$</td>
<td>0.98</td>
<td>0.01</td>
</tr>
<tr>
<td>$\mu_0$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$\mu_d$</td>
<td>-0.4</td>
<td>1</td>
</tr>
<tr>
<td>$\mu_{dd}$</td>
<td>-0.4</td>
<td>1</td>
</tr>
<tr>
<td>$\phi_1 + \phi_2$</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$\beta_1 + \beta_2$</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$\sigma^2_t$</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>$\sigma^2_0$</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

|--------------------------|--------------------------|--------------------------|

DIC | 654.918 | 655.428 |

Note: 1. Burn-in / Total iterations = 5,000 / 55,000  
2. S.D. refers to a standard deviation. HPDI refers to a highest posterior density.  
3. Ranges of break dates are based on 50% HPDI.
Table 4. In-sample Fitting: Average $R^2$’s from Rolling Predictive Regressions

\[ TSP \text{ Predictive Regression: } \Delta y_{t:t+h} = \alpha_0 + \alpha_1 c_t^{TSP} + u_{t+h} \]
\[ DSP \text{ Predictive Regression: } \Delta y_{t:t+h} = \alpha_0 + \alpha_2 c_t^{DSP} + u_{t+h} \]

<table>
<thead>
<tr>
<th>h</th>
<th>Cycle from TSP Model</th>
<th>Cycle from DSP Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.022</td>
<td>0.101</td>
</tr>
<tr>
<td>2</td>
<td>0.059</td>
<td>0.103</td>
</tr>
<tr>
<td>3</td>
<td>0.107</td>
<td>0.079</td>
</tr>
<tr>
<td>4</td>
<td>0.158</td>
<td>0.059</td>
</tr>
<tr>
<td>5</td>
<td>0.211</td>
<td>0.033</td>
</tr>
<tr>
<td>6</td>
<td>0.260</td>
<td>0.018</td>
</tr>
<tr>
<td>7</td>
<td>0.306</td>
<td>0.006</td>
</tr>
<tr>
<td>8</td>
<td>0.344</td>
<td>0.001</td>
</tr>
<tr>
<td>9</td>
<td>0.372</td>
<td>0.001</td>
</tr>
<tr>
<td>10</td>
<td>0.400</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Note: 1. Rolling window size is 40 years.
Table 5.A. Out-of-sample Predictive Performances: Mean Squared Errors from Rolling Predictive Regressions

\[
\text{TSP Predictive Regression: } \Delta y_{t:t+h} = \alpha_0 + \alpha_1 c^{\text{TSP}}_t + u_{t+h} \\
\text{DSP Predictive Regression: } \Delta y_{t:t+h} = \alpha_0 + \alpha_2 c^{\text{DSP}}_t + u_{t+h} \\
\text{Random Walk Model: } \Delta y_{t:t+h} = \alpha_0 + u_{t+h}
\]

<table>
<thead>
<tr>
<th>h</th>
<th>Cycle from TSP Model</th>
<th>Cycle from DSP Model</th>
<th>Random Walk Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.46</td>
<td>0.37</td>
<td>0.43</td>
</tr>
<tr>
<td>2</td>
<td>1.35</td>
<td>1.03</td>
<td>1.23</td>
</tr>
<tr>
<td>3</td>
<td>2.69</td>
<td>2.05</td>
<td>2.37</td>
</tr>
<tr>
<td>4</td>
<td>4.35</td>
<td>3.31</td>
<td>3.76</td>
</tr>
<tr>
<td>5</td>
<td>6.37</td>
<td>4.98</td>
<td>5.44</td>
</tr>
<tr>
<td>6</td>
<td>8.58</td>
<td>6.89</td>
<td>7.31</td>
</tr>
<tr>
<td>7</td>
<td>10.91</td>
<td>9.19</td>
<td>9.40</td>
</tr>
<tr>
<td>8</td>
<td>13.34</td>
<td>11.79</td>
<td>11.70</td>
</tr>
<tr>
<td>9</td>
<td>15.79</td>
<td>14.40</td>
<td>14.18</td>
</tr>
<tr>
<td>10</td>
<td>18.38</td>
<td>17.28</td>
<td>16.91</td>
</tr>
</tbody>
</table>

Note: 1. The MSEs are computed from 1987:Q2.
2. Rolling window size is 40 years.
Table 5.B. Out-of-sample Predictive Performances: Clack and West (2007) Test Based on Mean Squared Errors from Rolling Predictive Regressions

\[ TSP \text{ Predictive Regression: } \Delta y_{t:t+h} = \alpha_0 + \alpha_1 c^TSP_t + u_{t+h} \]
\[ DSP \text{ Predictive Regression: } \Delta y_{t:t+h} = \alpha_0 + \alpha_2 c^{DSP}_t + u_{t+h} \]
\[ \text{Random Walk Model: } \Delta y_{t:t+h} = \alpha_0 + u_{t+h} \]

<table>
<thead>
<tr>
<th>h</th>
<th>TSP Model (v.s.) Random Walk (H_0: MSE^{TSP} = MSE^{RW})</th>
<th>DSP Model (v.s.) Random Walk (H_0: MSE^{DSP} = MSE^{RW})</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.89</td>
<td>1.32*</td>
</tr>
<tr>
<td>2</td>
<td>-0.61</td>
<td>1.48*</td>
</tr>
<tr>
<td>3</td>
<td>-0.43</td>
<td>1.51*</td>
</tr>
<tr>
<td>4</td>
<td>-0.25</td>
<td>1.69**</td>
</tr>
<tr>
<td>5</td>
<td>-0.12</td>
<td>1.79**</td>
</tr>
<tr>
<td>6</td>
<td>0.00</td>
<td>1.89**</td>
</tr>
<tr>
<td>7</td>
<td>0.15</td>
<td>2.01**</td>
</tr>
<tr>
<td>8</td>
<td>0.28</td>
<td>-1.06</td>
</tr>
<tr>
<td>9</td>
<td>0.39</td>
<td>-1.28</td>
</tr>
<tr>
<td>10</td>
<td>0.51</td>
<td>-1.22</td>
</tr>
</tbody>
</table>

Note: 1. Rolling window size is 40 years.
2. The mean squared errors (MSEs) are computed from 1987:Q2 up to 2014:Q4.
3. The critical values of the CW test statistic are 1.28 and 1.64 for 10% and 5% significance levels, respectively.
4. Two asterisks represent statistical significance at 5% and one asterisk represents significance at 10%. 

Figure 1. Classical and Bayesian Inferences for Cycle Component: ARIMA(2,1,2) Model with a Known Break Point in Mean [Sample: 1947:Q1~1998:Q2]

Note:
1. The data set used is same as in Morley et al. (2003) and Perron and Wada (2009).
2. Burn-in / Total iterations = 5,000 / 55,000
Figure 2.A. Likelihood Surface of a Representative Sample for Type #1

\[ y_t = \mu + e_t - \theta e_{t-1}, \quad e_t \sim i.i.d \text{ } N(0, \sigma^2), \quad t = 1, 2, \ldots, T, \]

\[ \mu = 1, \theta = 0.8, \sigma^2 = 1, T = 50. \]
Figure 2.B. Likelihood Surface of a Representative Sample for Type #2

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

\[ \mu = 1, \theta = 0.8, \sigma^2 = 1, T = 50. \]
Figure 2.C. Likelihood Surface of a Representative Sample for Type #3

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

\[ \mu = 1, \theta = 0.8, \sigma^2 = 1, T = 50. \]
Figure 3.A. Bayesian Inference for Cyclical Component: ARIMA(2,1,2) Model with Unknown Break Points in Mean [Sample: 1947:Q1~2014:Q4]

- **TSP Model with One Break in Mean**
- **DSP Model with One Break in Mean**

Posterior Mean of Cyclical Component

- **TSP Model with Two Breaks in Mean**
- **DSP Model with Two Breaks in Mean**

Posterior Mean of Cyclical Component
Figure 3.B. The Nature of Structural Breaks in the Mean Growth Rate: ARIMA(2,1,2) Model with Unknown Break Points in Mean [Sample: 1947:Q1~2014:Q4]

*TSP Model* with One Break in Mean

*DSP Model* with One Break in Mean

60% HPDI and Median of Long-run Growth Mean

*TSP Model* with Two Breaks in Mean

*DSP Model* with Two Breaks in Mean

60% HPDI and Median of Long-run Growth Mean

Note: 1. HPDI refers to a highest posterior density.
Figure 4. In-sample Fitting: Average $R^2$'s from Rolling Regressions

\[ TSP \text{ Predictive Regression:} \ \Delta y_{t:t+h} = \alpha_0 + \alpha_1 c_t^{TSP} + u_{t+h} \]
\[ DSP \text{ Predictive Regression:} \ \Delta y_{t:t+h} = \alpha_0 + \alpha_2 c_t^{DSP} + u_{t+h} \]

Note: 1. The horizontal axis represents prediction horizon.
3. The bold lines represent the mean of $R^2$'s using the cycle from the DSP model. The dotted lines represent the mean of $R^2$'s using the cycle from the TSP model.
Figure 5. Out-of-sample Predictive Performances: Mean Square Errors from Rolling Regressions

\[ TSP \text{ Predictive Regression: } \Delta y_{t:t+h} = \alpha_0 + \alpha_1 c_t^{TSP} + u_{t+h} \]

\[ DSP \text{ Predictive Regression: } \Delta y_{t:t+h} = \alpha_0 + \alpha_2 c_t^{DSP} + u_{t+h} \]

\[ \text{Random Walk Model: } \Delta y_{t:t+h} = \alpha_0 + u_{t+h} \]

Note: 1. The MSEs are computed from 1987:Q2 up to 2014:Q4.
2. One break in the mean is allowed in extracting the cycle components.
3. Rolling window size is 40 years.
4. The horizontal axis represents prediction horizon.
5. The bold lines represent the root mean square errors obtained from the DSP Predictive Regression.
6. The dotted lines represent the root mean square errors obtained from the TSP Predictive Regression.
7. The dashed lines represent the root mean square errors obtained from the random walk model.