Likelihood Inference for Dynamic Linear Models with Markov Switching Parameters: On the Efficiency of the Kim Filter

Kyu Ho Kang*
Young Min Kim†

February 2017

Abstract

The Kim filter (KF) approximation is commonly used for the likelihood calculation of dynamic linear models with Markov regime-switching parameters. Despite its popularity, its approximation error has not been examined in a rigorous way yet. This study investigates the reliability of the KF approximation. To measure the approximation error, we compare the outcomes of the KF method with those of the auxiliary particle filter (APF). The APF is a numerical method requiring a longer computing time, but its error can be sufficiently minimized by increasing simulation size. We conduct extensive simulation and empirical studies for comparison purposes. According to our experiments, the likelihood values obtained from the KF approximation are practically identical to those of the APF, and the KF method becomes more reliable when regimes are more persistent. Consequently, the KF approximation is a fast and efficient approach for maximum likelihood estimation of regime-switching dynamic linear models. This article contributes to the literature by providing evidence to justify the use of the KF method.

(JEL classification: C22, C63)

Keywords: State space model, auxiliary particle filter, Kalman filter, maximum likelihood estimation

*Department of Economics, Korea University, Seoul, South Korea, E-mail: kyuho@korea.ac.kr, Tel: +82-2-3290-5132

†*Corresponding author*, Department of Economics, Korea University, Seoul, South Korea, E-mail: eccentricecon@korea.ac.kr, Tel: +82-2-3290-2200

1 Introduction

Dynamic linear models with Markov regime-switching parameters are widely used in empirical macroeconomics and finance because of their flexibility. This flexibility is attributed to two types of unobserved state variables in the model: continuous latent variables following an autoregressive process and discrete latent variables governed by a first-order Markov process. This class of models includes the regime-switching dynamic common factor, time-varying parameter, and unobserved component models. These models are particularly useful for detecting drastic regime changes in business cycles. Business cycles can be empirically characterized by a dynamic common factor among various coincident macroeconomic variables. The asymmetric nature of expansions and contractions or volatility reduction in the business cycles can be modeled by allowing the common factor dynamics to be subject to regime shifts according to a Markov process.

For statistical inference, maximum likelihood estimation (MLE) is typically employed in a classical approach. However, the exact likelihood calculation of the Markov regime-switching dynamic linear models is challenging. The source of the problem is that the conditional density of observations is not expressed in a closed form because the conditional density of continuous latent variables depends on the entire history of the discrete state variables.

To resolve this problem, the Kim filter (KF) approximation was proposed in the work of Kim (1994). For convenience, we refer to continuous and discrete state variables as factors and regimes, respectively. The key idea behind the KF method is to approximate the conditional density of the factors, so that it depends only on the distribution of the most recent regimes, and not on their entire history. The KF method makes the likelihood calculation feasible, and it is easy to implement practically. Therefore, a number of studies including Kahn and Rich (2007), Timmermann (2001), Morley and Piger (2012), Kang, Kim, and Morley (2009), and Chauvet (1998) have employed this approximation method for MLE of regime-switching dynamic linear models.

This work is motivated by the fact that the approximation error has not been investigated rigorously despite the popularity of the KF method. The primary goal of this study is to test whether the KF method is efficient for the likelihood computation. To this end, we measure its approximation error by comparing outcomes from the KF method with those from the auxiliary particle filter (APF) method. The APF, introduced by Pitt and Shephard (1999), relies on a numerical integration, and its numerical error can be sufficiently minimized by increasing simulation size.[1] As the difference between the likelihood values from the KF and APF methods is smaller, the KF method is regarded as more efficient. It should be noted that using the APF is not desirable for MLE because this numerical method requires a much longer computing time than the KF method.

This idea of comparing the KF and APF methods is based on the work of Herbst and Schorfheide (2015), in which the Kalman filter and particle filtering for dynamic linear models without regime shifts are compared. We conduct extensive simulation and empirical studies for the comparison. In the simulation study, the likelihood inference of the methods of dynamic common factor, time-varying parameter, and unobserved component models with Markov switching parameters is evaluated in terms of accuracy and computing time given the simulated data. The models considered in the empirical application are Friedman's plucking model and a business cycle turning points model.

According to our experiments, the likelihood values obtained from the KF approximation are almost identical to those of the APF method, which is robust to model specification and sample size. Because the KF algorithm is simpler, and easier to implement in practice than the APF method, the KF approximation is a fast and efficient approach for MLE. Our work contributes to the literature on regime-switching dynamic linear models by providing evidence to justify the use of the KF method. However, it should be noted that the performance of the KF method tends to be less reliable as regime shifts are more frequent. By a Monte Carlo simulation, we show that the KF approximation error has a larger variance when regimes are less persistent.

The rest of this paper is organized as follows. In Section 2, we present a general

[1] Many studies on a Bayesian framework have applied the APF for the likelihood computation, given the model parameters. For instance, Chib, Nardari, and Shephard (2002) developed the APF for a stochastic volatility model that can be expressed in a state-space model with exogenous and independent switching parameters. Recently, Kang (2014) extended his APF algorithm for dynamic linear models with endogenous Markov regime-switching parameters.

specification of our dynamic linear models with Markov switching parameters considered in this paper. Section 3 describes the KF and APF methods for likelihood inference. Sections 4 and 5 provide the results of KF approximation error evaluations based on various simulation and empirical applications. The conclusion is presented in Section 6.

2 Model Framework

We consider a joint stochastic process, $(y_t,\ \beta_t,\ s_t)$ where y_t is a $q \times 1$ vector of observable dependent variables and β_t is a $k \times 1$ vector of continuous state variables at time t. s_t is a discrete latent regime-indicator taking one of the values $\{1, 2, ..., N\}$ where N is the number of regimes. The state variables, $(\beta_t,\ s_t)$, are unobserved in general. $s_t = j$ indicates that the dependent and state variables are drawn from the jth regime at time t. Assume that $\mathcal{F}_{t-1}$ denotes the history of the observation sequence up to time $t-1$ and x_t is a $h \times 1$ vector of the predetermined or strictly exogenous variables. We then assume that at each time t, s_t, β_t, and y_t are sequentially, not simultaneously, generated from the following dynamic linear model with regime-switching parameters:

$$s_t|\mathbf{P} \sim Markov(s_0, \mathbf{P}), \tag{1}$$

$$\beta_t|\beta_{t-1}, \Theta, s_t \sim \mathcal{N}(\mu_{s_t} + G_{s_t}\beta_{t-1}, Q_{s_t}), \tag{2}$$

$$y_t|\mathcal{F}_{t-1}, \Theta, \beta_t, s_t \sim \mathcal{N}(H_{s_t}\beta_t + F_{s_t}x_t, R_{s_t}) \tag{3}$$

where s_0 is the initial regime at time 0, $\mathcal{N}(\cdot, \cdot)$ denotes the multivariate normal distribution, and

$$\Theta = \{H_{s_t}, F_{s_t}, R_{s_t}, \mu_{s_t}, G_{s_t}, Q_{s_t}\}_{s_t=1}^{N}$$

is the set of model parameters in the transition equation (2) and measurement equation (3). For generality, we allow all parameters to change over time depending on the regime.

The regime s_t is governed by a first-order finite-state Markov process with a transition probability matrix,

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{bmatrix}, \tag{4}$$

where $p_{ij} = \Pr[s_t = j|s_{t-1} = i]$ is the transition probability from regime $s_{t-1} = i$ to regime $s_t = j$, and $\sum_{j=1}^{N} p_{ij} = 1$ for all $i = 1, 2, .., N$. Without loss of generality, we assume that the transition probabilities are constant over time.

Once regime s_t is generated given s_{t-1} independently of the history of the factors and endogenous variables, $\{\beta_i, y_i\}_{t=0}^{t-1}$, β_t follows a first-order vector-autoregressive (VAR(1)) process in equation (2). The regime-dependent intercept, VAR coefficients, and factor shock variance-covariance are denoted by $\mu_{s_t} : k \times 1$, $G_{s_t} : k \times k$, and $Q_{s_t} : k \times k$, respectively.

Conditioned on the factors and regimes at time t, the dependent variable y_t is normally distributed as in equation (3). The marginal impact of the factors on the dependent variable is captured by $H_{s_t} : q \times k$. F_{s_t} is a $q \times h$ coefficient matrix of x_t, and R_{s_t} is a $q \times q$ variance-covariance matrix of the measurement errors.

Our model specification is completed by initializing the state variable and regime, $(\beta_0,\ s_0)$ at time 0. First, given the transition probability matrix $\mathbf{P}$, the initial regime is assumed to be generated from its unconditional distribution if the Markov-regime process is ergodic. That is, $s_0 = j$ is drawn with the steady-state probability,

$$\begin{bmatrix} \Pr[s_0 = 1] \\ \Pr[s_0 = 2] \\ \vdots \\ \Pr[s_0 = N] \end{bmatrix} = \left(\bar{\mathbf{P}}'\bar{\mathbf{P}}\right)^{-1} \bar{\mathbf{P}}' \begin{bmatrix} \mathbf{0}_{N\times 1} \\ 1 \end{bmatrix} : N \times 1 \tag{5}$$

where

$$\bar{\mathbf{P}} = \begin{bmatrix} \mathbf{I}_N - \mathbf{P}' \\ \mathbf{1}_N' \end{bmatrix} : (N+1) \times N.$$

For $N = 2$, the unconditional probabilities are simply obtained as

$$\begin{aligned} \Pr[s_0 = 1] &= \frac{1 - p_{22}}{2 - p_{11} - p_{22}}, \\ \Pr[s_0 = 2] &= 1 - \Pr[s_0 = 1]. \end{aligned} \tag{6}$$

If all regimes are transient as in a changepoint process, then s_0 is fixed at one.

Finally, given the initial regime s_0, the initial factors are generated differently, depending on whether the factor process is stationary or nonstationary. When it is stationary, β_0 is assumed to be distributed as its unconditional distribution of β_t conditioned on the initial state $s_0 = j$ as

$$\beta_0|\mathbf{\Theta}, s_0 = j \sim \mathcal{N}(\bar{\beta}_0^j, \bar{P}_0^j) \tag{7}$$

where $\bar{\beta}_0^j$ is a $k \times 1$ unconditional mean vector and $\bar{P}_0^j$ is a $k \times k$ unconditional variance-covariance matrix of β_t. Suppose that I_k denotes the $k \times k$ identity matrix, $\otimes$ is the Kronecker product, and $vec(\cdot)$ is the vectorization of a matrix. $(\bar{\beta}_0^j, \bar{P}_0^j)$ can then be derived as

$$\bar{\beta}_0^j = (I_k - G_{s_0=j})^{-1}\mu_{s_0=j},$$
$$\bar{P}_0^j = (I_k - G_{s_0=j} \otimes G_{s_0=j})^{-1} vec(Q_{s_0=j}),$$

respectively. On the other hand, if β_t is nonstationary, $\beta_0 = \bar{\beta}_0^j$ is treated as an additional regime-independent parameter to be estimated. As β_0 is no longer a random variable, its variance-covariance matrix, $\bar{P}_0 = \bar{P}_0^j$ should be fixed at a $k \times k$ zero matrix.

3 Likelihood Inference

If the regimes are non-stochastic and observable, the state-space model presented in equations (2) and (3) is linear and Gaussian, and the Kalman filter is applicable for the exact likelihood calculation. The unobservability of regimes, however, makes the transition equation nonlinear and the likelihood computation analytically infeasible via the Kalman filter.

To elaborate on the source of infeasibility, let $\mathbf{Y} = \{y_t\}_{t=1}^T$ denote a sequence of observations. Conditioned on $(\mathcal{F}_{t-1}, \Theta, \mathbf{P})$, by integrating out the current continuous and discrete state variables (s_t, β_t), the log likelihood $\log f(\mathbf{Y}|\Theta, \mathbf{P})$ can be obtained by the sum of the log likelihood densities over time:

$$f(y_t|\mathcal{F}_{t-1}, \Theta, \mathbf{P}) = \int f\left(y_t|s_t, \beta_t, \mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) p\left(s_t, \beta_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) d\left(s_t, \beta_t\right), \tag{8}$$

$$\text{and } \log f(\mathbf{Y}|\Theta, \mathbf{P}) = \sum_{t=1}^{T} \log f(y_t|\mathcal{F}_{t-1}, \Theta, \mathbf{P}). \tag{9}$$

The conditional density of (s_t, β_t) is then computed by integrating over the entire history of the state variables $\{(s_i, \beta_i)\}_{i=0}^{t-1}$ because the factor process is regime-dependent and

$$\begin{aligned}
& p\left(s_t, \beta_t | \mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) \\
& = \int p\left(s_t, \beta_t | \mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}, s_{t-1}, \beta_{t-1}\right) p\left(s_{t-1}, \beta_{t-1} | \mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) d\left(s_{t-1}, \beta_{t-1}\right) \\
& = \int p\left(\beta_t | \mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}, s_t, s_{t-1}, \beta_{t-1}\right) p\left(s_t | \mathbf{P}, s_{t-1}\right) \\
& \quad \times p\left(s_{t-1}, \beta_{t-1} | \mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) d\left(s_{t-1}, \beta_{t-1}\right), \\
& p\left(s_{t-1}, \beta_{t-1} | \mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) \\
& = \int p\left(\beta_{t-1} | \mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}, s_{t-1}, s_{t-2}, \beta_{t-2}\right) f(y_t | \mathcal{F}_{t-2}, \Theta, \mathbf{P}) \\
& \quad \times p\left(s_{t-1} | \mathbf{P}, s_{t-2}\right) p\left(s_{t-2}, \beta_{t-2} | \mathcal{F}_{t-2}, \mathbf{\Theta}, \mathbf{P}\right) d\left(s_{t-2}, \beta_{t-2}\right), \\
& p\left(s_{t-2}, \beta_{t-2} | \mathcal{F}_{t-2}, \mathbf{\Theta}, \mathbf{P}\right) \\
& = \int p\left(\beta_{t-2} | \mathcal{F}_{t-2}, \mathbf{\Theta}, \mathbf{P}, s_{t-2}, s_{t-3}, \beta_{t-2}\right) f(y_{t-2} | \mathcal{F}_{t-3}, \Theta, \mathbf{P}) \\
& \quad \times p\left(s_{t-2} | \mathbf{P}, s_{t-3}\right) p\left(s_{t-3}, \beta_{t-3} | \mathcal{F}_{t-3}, \mathbf{\Theta}, \mathbf{P}\right) d\left(s_{t-3}, \beta_{t-3}\right), \\
& \qquad \vdots
\end{aligned}$$

As a result, the conditional density of the observations y_t, $f(y_t|\mathcal{F}_{t-1}, \Theta, \mathbf{P})$ should be obtained by integrating over the entire history of the state variables up to time t. Because of this high dimensionality, the exact likelihood calculation is not feasible in general.

To overcome the computational difficulty, we can consider two approaches that rely on different approximation methods for the likelihood inference. One approach is the Kim filter, which is the approximate Kalman filter algorithm. The key idea of this method is to collapse $N \times N$ possible regimes into N possible regimes at each filtering recursion by approximation. Consequently, the conditional density of the factor β_t depends on $(s_t, s_{t-1}, \beta_{t-1})$, not $(\{s_i\}_{i=0}^{t}, \{\beta_i\}_{i=0}^{t-1})$ at each iteration of the recursions. The other approach is a sequential Monte Carlo method, which is widely used for the likelihood inference of nonlinear state space models. As one of the sequential Monte Carlo algorithms, the APF is implemented recursively to update the distribution of (s_t, β_t) conditional on the current information at time t using a set of particles with appropriate weights. The following subsections provide a detailed discussion of the KF and APF methods.

3.1 Kim Filter

The KF algorithm consists of three steps: the conditional Kalman filter, Hamilton filter, and approximation. Given $(\mathbf{Y}, \Theta, \mathbf{P})$, the log likelihood $\ln f(\mathbf{Y}|\mathbf{\Theta}, \mathbf{P})$ is initialized at zero. At time $t = 0$, the initial unobserved state variables distribution, $p(s_0, \beta_0|\mathbf{\Theta}, \mathbf{P})$, is given at the unconditional probability discussed in Section 2. For $t = 1, 2, \ldots, T$, the details of the KF algorithm steps for calculating the likelihood function are presented as follows.

Algorithm 1: **Likelihood Inference from the Kim Filter**

Step 1: (*Conditional Kalman Filter*) The objective of the conditional Kalman filter step is to obtain the inference of β_t conditional on $(s_{t-1} = i, s_t = j)$ and information up to time t. By running the Kalman filters for each pair of (s_t, s_{t-1}), the necessary quantities are recursively computed as

$$\mathbb{E}[\beta_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, s_t = j, s_{t-1} = i] = \beta^{(i,j)}_{t|t-1} = \mu_j + G_j\beta^i_{t-1|t-1}, \tag{10}$$

$$Var[\beta_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, s_t = j, s_{t-1} = i] = P^{(i,j)}_{t|t-1} = G_jP^i_{t-1|t-1}G'_j + Q_j, \tag{11}$$

$$\mathbb{E}[y_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, s_t = j, s_{t-1} = i] = y^{(i,j)}_{t|t-1} = y_t - H_j\beta^{(i,j)}_{t|t-1} - F_jx_t, \tag{12}$$

$$Var[y_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, s_t = j, s_{t-1} = i] = f^{(i,j)}_{t|t-1} = H_jP^{(i,j)}_{t|t-1}H'_j + R_j, \tag{13}$$

$$\mathbb{E}[\beta_t|\mathcal{F}_t, \mathbf{\Theta}, s_t = j, s_{t-1} = i] = \beta^{(i,j)}_{t|t} = \beta^{(i,j)}_{t|t-1} + K^{(i,j)}_t(y_t - y^{(i,j)}_{t|t-1}), \tag{14}$$

$$Var[\beta_t|\mathcal{F}_t, \mathbf{\Theta}, s_t = j, s_{t-1} = i] = P^{(i,j)}_{t|t} = P^{(i,j)}_{t|t-1} - K^{(i,j)}_tH'_jP^{(i,j)}_{t|t-1} \tag{15}$$

where $\beta^i_{t-1|t-1} = \mathbb{E}[\beta_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, s_{t-1} = i]$ and $P^i_{t-1|t-1} = Var[\beta_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, s_{t-1} = i]$ are the conditional expectation and variance-covariance matrix of β_t given $(\mathcal{F}_{t-1}$, $\mathbf{\Theta}$, $s_{t-1} = i)$, respectively. $K^{(i,j)}_t$ is the Kalman gain conditioned on $s_t = j$ and $s_{t-1} = i$ for $i, j = 1, 2, \ldots, N$, which is given by

$$K^{(i,j)}_t = P^{(i,j)}_{t|t-1}H'_j(f^{(i,j)}_{t|t-1})^{-1}.$$

Step 2: (*Hamilton Filter*) In this step, we apply the Hamilton filter (Hamilton (1989)) to obtain the likelihood density (i.e., $f(y_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P})$) and the filtered probabilities (i.e., $f(s_t, s_{t-1}|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P})$ and $f(s_t|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P})$). Given $f(s_{t-1}|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P})$, the following quantities can be sequentially computed:

$$f(s_t, s_{t-1}|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}) = \Pr[s_t|s_{t-1}]f(s_{t-1}|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}), \tag{16}$$

$$f(y_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}) = \sum_{s_t}\sum_{s_{t-1}} f(y_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, s_t, s_{t-1})f(s_t, s_{t-1}|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}), \tag{17}$$

$$f(s_t, s_{t-1}|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P}) = \frac{f(y_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, s_t, s_{t-1})f(s_t, s_{t-1}|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P})}{f(y_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P})}, \tag{18}$$

$$f(s_t|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P}) = \sum_{s_{t-1}} f(s_t, s_{t-1}|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P}), \tag{19}$$

where the conditional density of y_t

$$f(y_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, s_t = j, s_{t-1} = i) = \mathcal{N}\left(y_t|y^{(i,j)}_{t|t-1}, f^{(i,j)}_{t|t-1}\right)$$

can be computed using equations (12) and (13). Equation (16) is the predictive probability of the regimes and equations (18) and (19) are the filtered probabilities.

Step 3: (*Numerical Error*) The filtered probabilities are used in the next step when integrating out the lagged regimes to obtain the filtered factors conditioned on the current regime only. If the filtered probabilities are too close to zero, these values could be taken to be zero and the KF does not work because the filtered probabilities are used as denominators in the next step. To avoid this numerical problem, we replace the filtered probabilities by 10^{-4} whenever they are less than this value.

Step 4: (*Approximation*) In this step we derive the filtered distribution of the factor β_t conditioned on the current regime s_t,

$$\beta_t|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P}, s_t.$$

Given the normality assumption and the filtered probabilities, the KF algorithm provides the conditional expectation and variance-covariance of the factors as follows:[2]

$$\begin{aligned}\beta^j_{t|t} &= \mathbb{E}[\beta_t|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P}, s_t = j] \\ &= \frac{\sum_{i=1}^M f(s_t = j, s_{t-1} = i|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P})\beta^{(i,j)}_{t|t}}{f(s_t = j|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P})}, \qquad (20) \\ P^j_{t|t} &= Var[\beta_t|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P}, s_t = j] \\ &= \frac{\sum_{i=1}^M f(s_t = j, s_{t-1} = i|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P})[P^{(i,j)}_{t|t} + (\beta^j_{t|t} - \beta^{(i,j)}_{t|t})(\beta^j_{t|t} - \beta^{(i,j)}_{t|t})']}{f(s_t = j|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P})}. \qquad (21)\end{aligned}$$

Note that the filtered density $p(\beta_t|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P}, s_t = j)$ is a mixture of $N_t = N \times N_{t-1}$ states where N_{t-1} is the total number of regime sequences up to time $t-1$. This step is referred to as the approximation stage because $(\beta^j_{t|t}, P^j_{t|t})$ are calculated by integrating over s_{t-1}, not $\{s_i\}_{i=0}^{t-1}$. These filtered values are necessary when running the conditional kalman filter step the following time.

Step 5: If $t < T$, set $t = t + 1$ and go to Step 1; otherwise go to Step 6.

Step 6: *(Log likelihood)* Return the log likelihood

$$\ln f(\mathbf{Y}|\mathbf{\Theta}, \mathbf{P}) = \sum_{t=1}^T \ln f(y_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}) \qquad (22)$$

where the conditional density $f(y_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P})$ is calculated in equation (17).

3.2 Auxiliary Particle Filter

The essential step in the likelihood inference is updating the conditional distribution of the regimes and factors given the current information at time t. The APF is a class of simulation-based methods that recursively update the information contained in the current observations y_t to the predictive distribution, $(s_t, \beta_t)\,|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}$. This filtering procedure uses the sampling importance resampling method, which consists of two stages. The first stage is to propose candidate values of the state variables. The second stage is to reweigh them to produce draws from the target distribution, $p(s_t, \beta_t|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P})$. Per the Bayes theorem, the target density is proportional to the

[2]For the details of the derivation, refer to Kim and Nelson (1999b).

product of the conditional density of y_t and the predictive density of (s_t, β_t) .:

$$p\left(s_t, \beta_t|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P}\right) \propto f\left(y_t|s_t, \beta_t, \mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) p\left(s_t, \beta_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right).$$

The key step in the APF method is to introduce an auxiliary variable, denoted by g, where g indicates the conditional distribution of (s_t, β_t) given the gth samples $(s_{t-1}^{(g)}, \beta_{t-1}^{(g)})$ and $(\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P})$. The APF algorithm enables us to sample from

$$s_t, \beta_t, g|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P},$$

and return the filtered distribution, $s_t, \beta_t|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P}$.

Supose that the M samples $\{s_{t-1}^{(g)}, \beta_{t-1}^{(g)}\}_{g=1}^{M}$ drawn from

$$s_{t-1}, \beta_{t-1}|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}$$

are given, and that $(\hat{s}_t^{(g)}, \hat{\beta}_t^{(g)})$ comprise the mode of the transition distribution of (s_t, β_t),

$$s_t, \beta_t|s_{t-1}^{(g)}, \beta_{t-1}^{(g)}, \mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}$$

conditioned on $(s_{t-1}^{(g)},\ \beta_{t-1}^{(g)},\ \mathcal{F}_{t-1},\ \mathbf{\Theta},\ \mathbf{P})$ for each $g = 1, 2, .., M$.

The first stage is to propose R values for g generated by a proposal distribution whose mass function is given by

$$q\left(g|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P}\right) \propto f\left(y_t|\hat{s}_t^{(g)}, \hat{\beta}_t^{(g)}, \mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right).$$

In the second stage, we draw $\{s_t^{*(r)}, \beta_t^{*(r)}\}_{r=1}^{R}$ from the predictive distribution,

$$s_t, \beta_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}.$$

Then, for each $r = 1, 2, .., R$, we calculate the second importance weights,

$$w_r \propto \frac{f\left(y_t|s_t^{*(r)}, \beta_t^{*(r)}, \mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right)}{f\left(y_t|\hat{s}_t^{(g_r)}, \hat{\beta}_t^{(g_r)}, \mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right)} \tag{23}$$

where $\{g_r\}_{r=1}^{R}$ is a set of indices sampled from the M draws of g. Finally, we resample M draws for (s_t, β_t) with the second importance weights from $\{s_t^{*(r)}, \beta_t^{*(r)}\}_{r=1}^{R}$. The resampled M draws are taken as the samples from the filtered distribution, $s_t, \beta_t|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P}$.

Given $(\mathbf{Y}, \mathbf{\Theta}, \mathbf{P})$ and (M, R), the initial particles $\{s_0^{(g)}, \beta_0^{(g)}\}_{g=1}^M$ are drawn from $s_0, \beta_0|\mathbf{\Theta}, \mathbf{P}$, which is the unconditional probability as we discussed in Section 2. For each time point t, the steps of the APF algorithm can be summarized as follows.

Algorithm 2: **Auxiliary Particle Filter**

Step 1: *Obtain the mode of the proposal distribution* ($\hat{s}_t$, $\hat{\beta}_t$)*, compute the first stage weights, and sample auxiliary variables.*

Step 1-(a): Given $\mathbf{P}$, sample $\{s_{t-1}^{(g)}, \beta_{t-1}^{(g)}\}_{g=1}^M$ from $s_{t-1}, \beta_{t-1}|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}$, and $\{\hat{u}_t^{(g)}\}_{g=1}^M \sim Unif(0,1)$. Then, determine $\hat{s}_t^{(g)}$ as

$$\begin{aligned} \hat{s}_t^{(g)} &= 1 \text{ if } \hat{u}_t^{(g)} < p_1, \\ \hat{s}_t^{(g)} &= 2 \text{ if } \hat{u}_t^{(g)} \geq p_1 \end{aligned} \tag{24}$$

where $p_1 = p_{11}$ if $s_{t-1}^{(g)} = 1$, and $p_1 = p_{21}$ if $s_{t-1}^{(g)} = 2$.

Step 1-(b) Given $\mathbf{\Theta}$, $\{\hat{s}_t^{(g)}, \beta_{t-1}^{(g)}\}_{g=1}^M$, and (y_t, x_t), calculate $\hat{\beta}_t^{(g)}$ as

$$\hat{\beta}_t^{(g)} = \mathbb{E}\left[\beta_t|\beta_{t-1}^{(g)}, \hat{s}_t^{(g)}, \mathbf{\Theta}\right] = \mu_{\hat{s}_t^{(g)}} + G_{\hat{s}_t^{(g)}}\beta_{t-1}^{(g)}, \tag{25}$$

Step 1-(c) Compute the first stage weights,

$$w_g \propto \mathcal{N}_q\left(y_t|H_{\hat{s}_t^{(g)}}\hat{\beta}_t^{(g)} + F_{\hat{s}_t^{(g)}}x_t, R_{\hat{s}_t^{(g)}}\right) \tag{26}$$

Step 1-(d): Draw $\{g_r\}_{r=1}^R$ from the indices $g = 1, ..., M$ with the probability mass proportional to

$$w_g^* = \frac{w_g}{\sum_{g=1}^M w_g}, \tag{27}$$

and associate sampled index with $\{s_{t-1}^{(g_r)}, \beta_{t-1}^{(g_r)}\}_{r=1}^R$ and $\{\hat{s}_t^{(g_r)}, \hat{\beta}_t^{(g_r)}\}_{r=1}^R$.

Step 2: *Generate* s_t^*, β_t^* *from the transition densities, compute the second stage weights, and resample* (s_t, β_t)*.*

Step 2-(a): Given $\{s_{t-1}^{(g_r)}\}_{r=1}^{R}$, $\mathbf{P}$, and $\{u_t^{(r)}\}_{r=1}^{R} \sim unif(0,1)$, sample

$$s_t^{*(r)} = 1 \text{ if } u_t^{(r)} < p_1, \tag{28}$$
$$s_t^{*(r)} = 2 \text{ if } u_t^{(r)} \geq p_1$$

where $p_1 = p_{11}$ if $s_{t-1}^{(g_r)} = 1$, and $p_1 = p_{21}$ if $s_{t-1}^{(g_r)} = 2$.

Step 2-(b): Given $\{s_t^{*(r)}, \beta_{t-1}^{(g_r)}\}_{r=1}^{R}$ and $\mathbf{\Theta}$, simulate

$$\beta_t^{*(r)} \sim \mathcal{N}_k\left(\mu_{s_t^{*(r)}} + G_{s_t^{*(r)}}\beta_{t-1}^{(g_r)}, Q_{s_t^{*(r)}}\right). \tag{29}$$

Step 2-(c): Compute the second importance weights,

$$w_r \propto \frac{\mathcal{N}_q\left(y_t|H_{s_t^{*(r)}}\beta_t^{*(r)} + F_{s_t^{*(r)}}x_t, R_{s_t^{*(r)}}\right)}{\mathcal{N}_q\left(y_t|H_{\hat{s}_t^{(g_r)}}\hat{\beta}_t^{(g_r)} + F_{\hat{s}_t^{(g_r)}}x_t, R_{\hat{s}_t^{(g_r)}}\right)}. \tag{30}$$

Step 2-(d): Resample M times from the R draws with the probability mass,

$$w_r^* = \frac{w_r}{\sum_{r=1}^{R} w_r}. \tag{31}$$

The resampled draws are the filtered samples $\{s_t^{(g)}, \beta_t^{(g)}\}_{g=1}^{M}$ from $s_t, \beta_t|\mathcal{F}_t, \mathbf{\Theta}, \mathbf{P}$.

Given the filtered samples $\{s_{t-1}^{(g)}, \beta_{t-1}^{(g)}\}_{g=1}^{M}$ from

$$s_{t-1}, \beta_{t-1}|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P},$$

it is straightforward to sample the predictive draws, $\{s_t^{(g)}, \beta_t^{(g)}\}_{g=1}^{M}$ from $s_t, \beta_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}$. Once the predictive draws for (s_t, β_t) are simulated, we can compute the one-step-ahead conditional density of y_t,

$$f\left(y_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) = \int f\left(y_t|s_t, \beta_t, \mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) p\left(s_t, \beta_t|\mathcal{F}_{t-1}, \mathbf{\Theta}, \mathbf{P}\right) d\left(s_t, \beta_t\right) \tag{32}$$

by using the simple Monte Carlo averaging of $\mathcal{N}_q\left(y_t|H_{s_t}\beta_t + F_{s_t}x_t, R_{s_t}\right)$ over the predictive draws.

Given $(\mathbf{Y}, \Theta, \mathbf{P})$, the log likelihood $\ln f(\mathbf{Y}|\boldsymbol{\Theta}, \mathbf{P})$ is initialized at zero. At time $t = 0$, initial particles $\{s_0^{(g)}, \beta_0^{(g)}\}_{g=1}^M$ are drawn from the unconditional probability as discussed in Section 2. For $t = 1, 2, .., T$, the APF algorithm steps are summarized as follows.

Algorithm 3: **Likelihood Inference using the Auxiliary Particle Filter**

Step 1: Simulate the predictive density of (s_t, β_t) and calculate the likelihood density.

Step 1-(a): Given $\{s_{t-1}^{(g)}\}_{g=1}^M$, $\{u_t^{(g)}\}_{g=1}^M \sim Unif(0,1)$, and $\mathbf{P}$, predictive values of s_t are drawn from

$$s_t^{(g)} = 1 \text{ if } u_t^{(g)} < p_1, \qquad (33)$$
$$s_t^{(g)} = 2 \text{ if } u_t^{(g)} \geq p_1$$

where $p_1 = p_{11}$ if $s_{t-1}^{(g)} = 1$, and $p_1 = p_{21}$ if $s_{t-1}^{(g)} = 2$.

Step 1-(b): Conditioned on $\{s_t^{(g)}, \beta_{t-1}^{(g)}\}_{g=1}^M$ and $\boldsymbol{\Theta}$, draw predictive values of β_t from

$$\beta_t^{(g)} \sim \mathcal{N}_k\left(\mu_{s_t^{(g)}} + G_{s_t^{(g)}}\beta_{t-1}^{(g)}, Q_{s_t^{(g)}}\right) \qquad (34)$$

Step 1-(c): Compute the one-step-ahead predictive density of y_t by a numerical integration as

$$f(y_t|\mathcal{F}_{t-1}, \boldsymbol{\Theta}, \mathbf{P}) \simeq \frac{1}{M}\sum_{g=1}^{M}\mathcal{N}_q\left(y_t|H_{s_t^{(g)}}\beta_t^{(g)} + F_{s_t^{(g)}}x_t, R_{s_t^{(g)}}\right). \qquad (35)$$

Step 2: Apply Algorithm 2 and sample $\{s_t^{(g)}, \beta_t^{(g)}\}_{g=1}^M$ from $s_t, \beta_t|\mathcal{F}_t, \boldsymbol{\Theta}, \mathbf{P}$. If $t < T$, set $t = t + 1$ and go to Step 1; otherwise go to Step 3.

Step 3: Return the log likelihood,

$$\ln f(\mathbf{Y}|\boldsymbol{\Theta}, \mathbf{P}) = \sum_{t=1}^{T}\ln f(y_t|\mathcal{F}_{t-1}, \boldsymbol{\Theta}, \mathbf{P}). \qquad (36)$$

Note that given the parameters we can obtain the filtered probabilities of the regimes and filtered factors over time by averaging $\{s_t^{(g)}, \beta_t^{(g)}\}_{g=1}^M$ drawn from $s_t, \beta_t|\mathcal{F}_t, \boldsymbol{\Theta}, \mathbf{P}$ in Step 2.

4 Simulation Study

The KF approximation method is evaluated based on simulation studies and empirical applications in comparison with the APF method. We begin by measuring the approximation error of the KF method using simulated data. We consider three representative examples of dynamic linear models with Markov switching parameters, which can be expressed as special cases of the general model specification in Section 2.

4.1 Models

4.1.1 Dynamic Common Factor Model with Markov Switching Parameters

The first specification is a dynamic common factor model with Markov switching parameters. This model specification is widely used to empirically characterize business cycles by a dynamic factor β_t among multiple observed variables y_t. The dynamic factor is subject to regime shifts modeled by a Markov process s_t. At time t, (s_t, β_t, y_t) are sequentially, not simultaneously, generated from the following generic dynamic linear model with regime-switching parameters:

$$\begin{aligned} s_t|\mathbf{P} &\sim Markov(s_0, \mathbf{P}), \\ \beta_t|\beta_{t-1}, \Theta, s_t &\sim \mathcal{N}(G_{s_t}\beta_{t-1}, Q_{s_t}), \\ y_t|\mathcal{F}_{t-1}, \Theta, \beta_t, s_t &\sim \mathcal{N}(H_{s_t}\beta_t, R_{s_t}). \end{aligned} \tag{37}$$

The transition matrix for regime s_t is given by

$$\mathbf{P} = \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{bmatrix}.$$

Throughout Sections 4 and 5, we assume that the regime s_t is governed by a two-state first-order Markov process, in which all regimes are recurrent.[3]

For data simulation, bivariate observations y_t are generated by one common factor

[3] In the estimation of regime-switching models, regime identification is an important issue. However, the focus of this paper is the likelihood inference given the parameters. For this reason we do not discuss the label switching problem.

(i.e., $k = 1$). Our choice of parametrization is given by

	$s_t = 1$	$s_t = 2$
G_{s_t}	0.5	0.9
Q_{s_t}	1	3
H_{s_t}	$\begin{pmatrix} 1 & -0.5 \end{pmatrix}'$	$\begin{pmatrix} 1 & 0.5 \end{pmatrix}'$
R_{s_t}	$1 \times I_2$	$4 \times I_2$

In this parametrization, the common factor process is regime-dependent. Particularly, the factor is more persistent and volatile in regime 2. Further, the factor loadings and measurement shock variance differ in different regimes.

4.1.2 Time-Varying Parameter Model with Markov Switching Parameters

Consider the case in which there is a possibility of gradual changes in linear regression coefficients over time along with regime shifts in the coefficient process or volatility. This can be easily modeled by a time-varying parameter model with Markov switching parameters as follows:

$$\beta_t|\beta_{t-1}, \Theta, s_t \sim \mathcal{N}(\beta_{t-1}, Q_{s_t}), \tag{38}$$
$$y_t|\mathcal{F}_{t-1}, \Theta, \beta_t, s_t \sim \mathcal{N}(H_t\beta_t, R_{s_t})$$

where $H_t \sim Unif(0, 2)$ is a covariate uncorrelated with the measurement errors. We fix the measurement error variance and parameter shock variance such that $Q_{s_t=1} = R_{s_t=1} = 1$, $Q_{s_t=2} = 5$, and $R_{s_t=2} = 3$. As the variances in regime 2 are larger than in regime 1, the simulated data is supposed to reveal more drastic coefficient changes and higher volatility during the period corresponding to the second regime.

4.1.3 Unobserved Component Model with Markov Switching Parameters

In an unobserved component model the observed time series data is decomposed into the stochastic trend and cyclical components. The trend component is typically assumed to be a random walk process, and the cyclical component follows a stationary vector autoregressive process. We consider a simple specification:

$$\beta_t|\beta_{t-1}, \Theta, s_t \sim \mathcal{N}(\mu_{s_t} + G_{s_t}\beta_{t-1}, Q_{s_t}),$$
$$y_t|\mathcal{F}_{t-1}, \Theta, \beta_t, s_t \sim \mathcal{N}(\beta_t, R_{s_t})$$

where y_t is univariate and assumed to be the sum of a first-order autoregressive process and a white noise component. These unobserved components are identified as the difference in persistence. We set the parameters as presented below, so that the dependent variable is more persistent and volatile under regime 2.

	$s_t = 1$	$s_t = 2$
μ_{s_t}	2	1
G_{s_t}	0.5	0.9
Q_{s_t}	1	4
R_{s_t}	1	2

4.2 Likelihood Comparison

For each model, we simulate the time series of dependent variables and factors given the true parameters. We consider five different sample sizes ($T = 80, 100, 200, 400$, and 800) to see whether the size of the KF approximation error is robust to sample size. The first and third 25% of samples are drawn under regime 1 and the others are drawn under regime 2. For instance, in the case of $T = 80$, the first and third 20 observations and factors are simulated by the parameters under regime 1, and the others under regime 2. Using the simulated data, we run the KF and APF algorithms once to compute and compare the likelihood values at the true parameters. For the APF, resampling and particle sizes are given by $R = 50,000$ and $M = 50,000$, respectively. The likelihood values obtained from the APF with such large simulation sizes are accurate enough to be regarded as the true likelihood values.

Table 1 presents the results of the likelihood computations. There are three important findings in this table. First, the log likelihoods approximated by the KF method are almost identical to those of the APF. The average approximation error is almost zero and the maximum absolute error is 0.47 in this experiment. The KF method does not seem to lead to a misleading likelihood ratio test. Second, the performance of the KF method

Table 1: **Log Likelihood (ln L) and Computing Time (Time) Comparison Between the KF and APF methods** DCF, TVP, and UC indicate the dynamic common factor model, time-varying parameter model, and unobserved component model, respectively. The computing time is in seconds.

(a) DCF

	APF		KF	
	$\ln L$	Time	$\ln L$	Time
$T = 80$	-328.31	615.99	-328.19	0.06
$T = 100$	-405.52	765.89	-405.34	0.04
$T = 200$	-803.75	1546.29	-803.65	0.07
$T = 400$	-1633.13	3098.10	-1633.50	0.13
$T = 800$	-3213.81	6187.61	-3214.15	0.25

(b) TVP

	APF		KF	
	$\ln L$	Time	$\ln L$	Time
$T = 80$	-184.34	585.00	-184.46	0.03
$T = 100$	-242.13	724.21	-241.60	0.02
$T = 200$	-469.72	1447.01	-469.94	0.03
$T = 400$	-934.77	2921.31	-934.61	0.06
$T = 800$	-1805.31	5780.46	-1804.92	0.11

(c) UC

	APF		KF	
	$\ln L$	Time	$\ln L$	Time
$T = 80$	-172.92	585.17	-172.97	0.01
$T = 100$	-218.31	730.79	-218.49	0.01
$T = 200$	-439.57	1441.91	-439.62	0.03
$T = 400$	-883.84	2910.35	-883.65	0.05
$T = 800$	-1713.41	5760.40	-1713.73	0.11

is robust to the model specifications and sample sizes as shown in the table. Third, the computing speed of the KF method is incomparable. The APF takes hundreds of seconds whereas the KF completes within one second. In addition, Figure 1 plots the log conditional densities of the observations over time, and clearly indicates that the log likelihood densities calculated by the KF method are indistinguishable from those of the APF method. Consequently, those simulation studies strongly demonstrate that the KF method is fast and efficient enough to be used for likelihood inference. The KF method is a more desirable approach for likelihood function optimization than the APF method.

Another important element of approximation quality is the accuracy of the filtered states and filtered probabilities of regimes. We report the quantities of the filtered states and regime probabilities obtained from a single run of the KF and APF in Figures 2 and 3. These figures depict the sequence of the filtered factors and filtered probabilities of the regimes along with the corresponding true values in the case of T=200, as an example. There is little difference between the filtered values produced by the KF and APF, which is a finding robust to the sample sizes.

4.3 Regime Persistence and Approximation Error

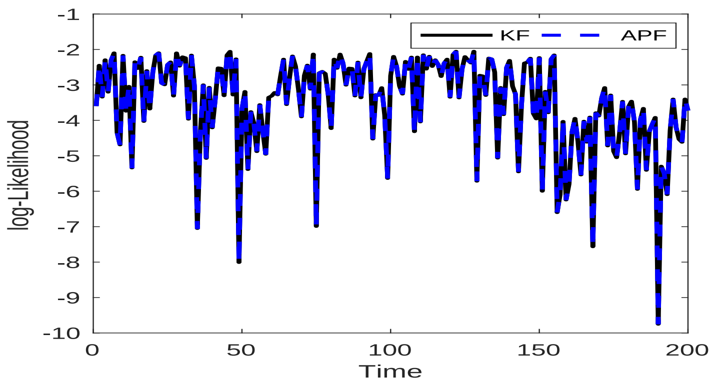
We now examine the sensitivity of the KF approximation error to regime persistence. It is important to investigate whether the KF method successfully approximates the historical path of the regimes regardless of regime shift frequency. To do this, we repeat the simulation in Section 4.2 500 times with a sample size of 80. For each simulation, the regime process is generated according to a Markov process with the following transition matrices:

$$\mathbf{P}_H = \begin{bmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{bmatrix} \text{ and } \mathbf{P}_L = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$$

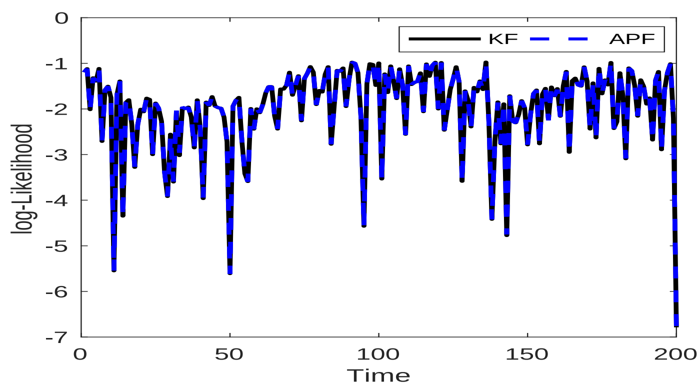
$\mathbf{P}_H$ and $\mathbf{P}_L$ indicate high and low regime persistence, respectively. As a result, the regime shifts generated by $\mathbf{P}_L$ occur more frequently than those generated by $\mathbf{P}_H$. For each of the 500 simulated data sets, we run the KF and APF and store the approximation error,

$$\ln f\left(\mathbf{Y}|\boldsymbol{\Theta},\mathbf{P}\right)_{KF} - \ln f\left(\mathbf{Y}|\boldsymbol{\Theta},\mathbf{P}\right)_{APF},\ \mathbf{P} \in \{\mathbf{P_H},\mathbf{P_L}\}.$$

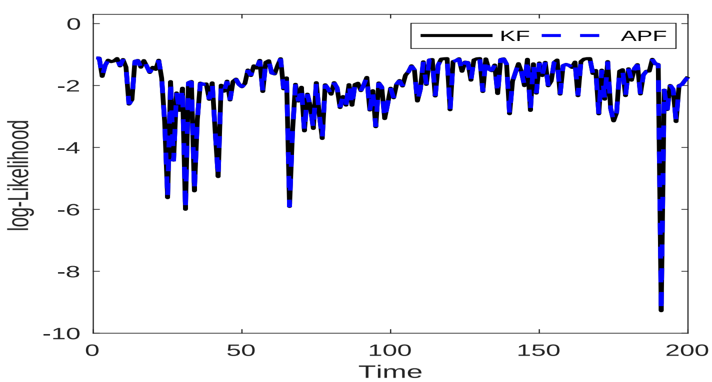
Figure 1: **Log Likelihood Densities** This figure plots the log likelihood densities over time. DCF, TVP, and UC indicate the dynamic common factor model, time-varying parameter model, and unobserved component model, respectively. The conditional densities from the KF and APF methods are the black dashed line and blue dotted line, respectively.



(a) DCF

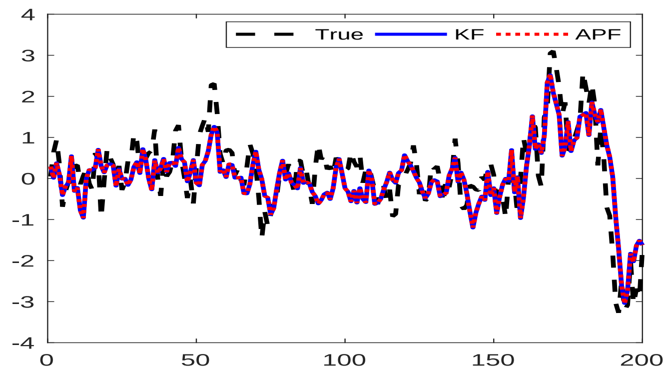


(b) TVP

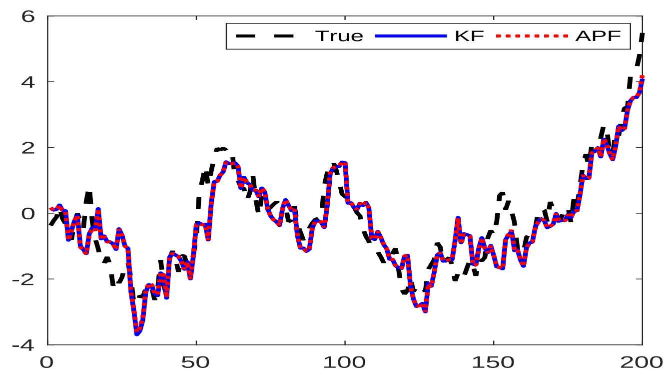


(c) UC

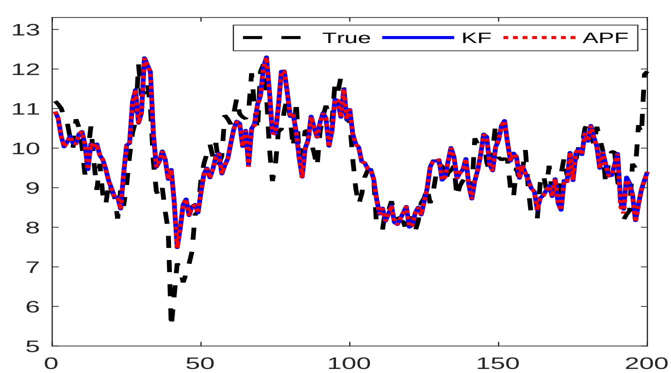
Figure 2: **Filtered Factors** This figure plots the filtered factors over time for the DCF, TVP, and UC models. The true factors, factors filtered by the KF, and factors filtered by the APF are the black dashed line, blue solid line, and red dotted line, respectively.



(a) DCF

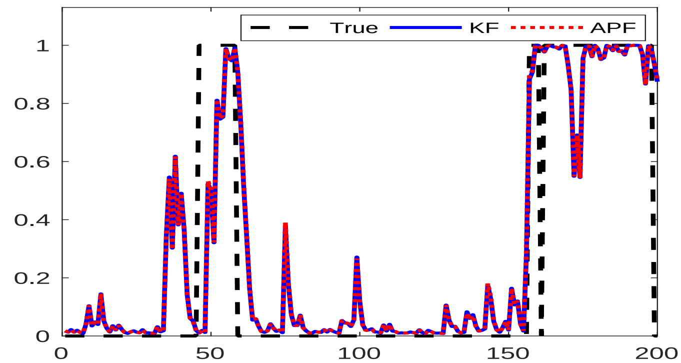


(b) TVP

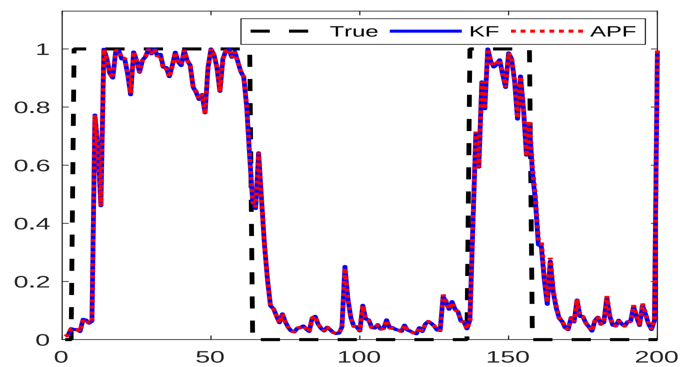


(c) UC

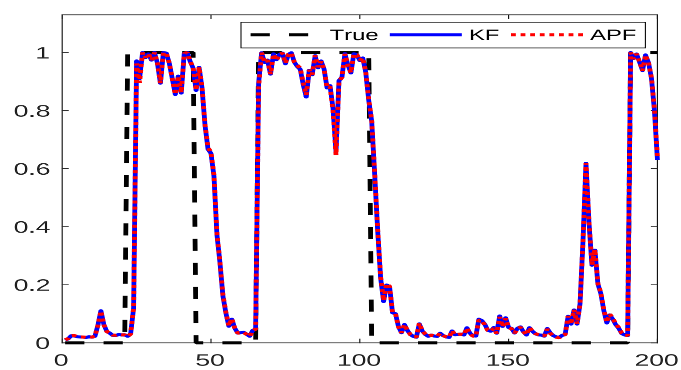
Figure 3: **Filtered Regimes** This figure plots the filtered regimes over time for the DCF, TVP, and UC models. The true regimes, regime probabilities filtered by the KF, and regime probabilities filtered by the APF are the black dashed line, blue solid line, and red dotted line, respectively.



(a) DCF



(b) TVP



(c) UC

Figure 4 plots the empirical kernel densities of the 500 approximation errors, and Table 2 reports its sample mean and standard deviation. For all models, empirical densities are centered around zero regardless of the regime persistence. However, the KF method seems to be more reliable when the regime is more persistent, because the error density with $\mathbf{P}_H$ is much sharper than that with $\mathbf{P}_L$. This is attributed to the fact that approximating the entire regime path by the recent regimes is difficult when the regime shifts are frequent and the set of potential regime paths with a high mass is large. In short, the performance of the KF method becomes uncertain when regime persistence is low, although it is robust to the model specification and sample size.

Table 2: **Summary Statistics for Approximation Error** This table presents the mean and standard deviation (s.d.) of likelihood discrepancy between the KF and APF.

	$\mathbf{P}_H$		$\mathbf{P}_L$	
	mean	s.d.	mean	s.d.
DCF	0.0007	0.0945	0.0266	0.1797
TVP	0.0149	0.1109	0.0895	0.2738
UC	0.0080	0.0147	0.0147	0.2082

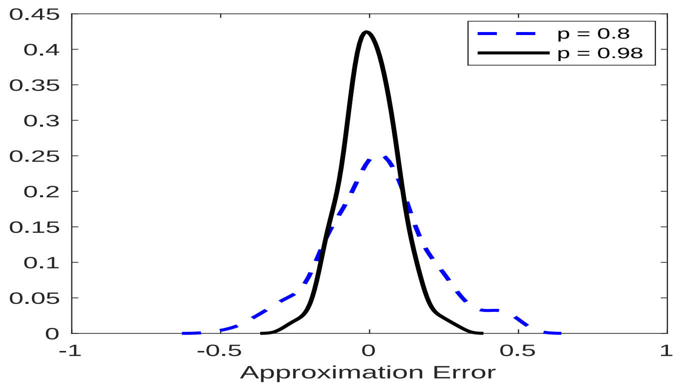
5 Empirical Applications

Regarding empirical studies, we consider two popular regime-switching dynamic linear models for business cycle analysis: Friedman's plucking model (Kim and Nelson (1999a)) and the business cycle turning points model (Kim and Nelson (1998)).

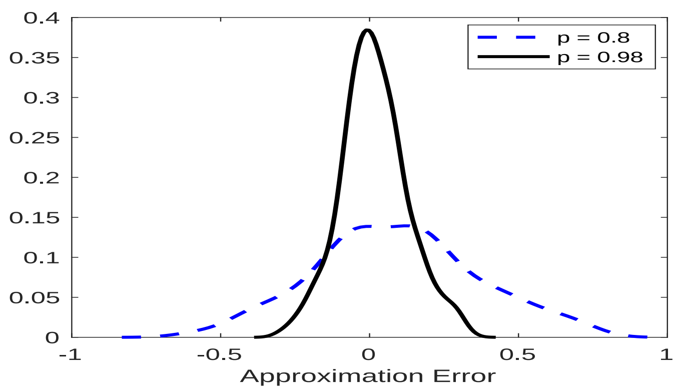
5.1 Friedman's Plucking Model

Milton Friedman's plucking model is one of the popular business cycle models to explain the nature of recessions. The most distinguishing feature of this model is that business cycles are extremely asymmetric. The time-path of the output can only move along a potential output process. However, the output is plucked downwards at irregular intervals. The plucking shocks are temporary demand shocks, whereas the fluctuations of the potential output are driven by supply shocks. From this model's point of view, recessions are caused by uncommon negative and temporary shocks such as financial crises

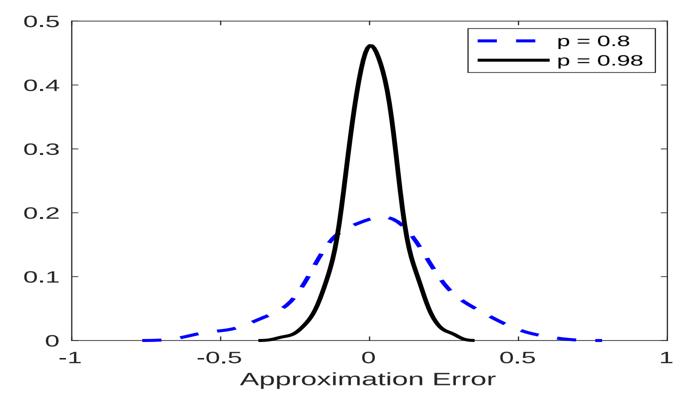
Figure 4: **Approximation Error Kernel Density** This figure plots the kernel density estimates of likelihood discrepancy based on the KF and APF. The approximation error densities corresponding to high and low regime persistence are the solid and dashed lines, respectively.



(a) DCF



(b) TVP



(c) UC

and oil shocks. Kim and Nelson (1999a) propose an econometric model incorporating the asymmetric behavior of business cycles. Following Kim and Nelson (1999a), we decompose the log of real GDP y_t into a transitory component, x_t, and trend component, z_t:

$$y_t = x_t + z_t \tag{39}$$

where

$$x_t = \delta_{s_t} + \phi_1 x_{t-1} + \phi_2 x_{t-2} + u_t, \quad u_t \sim \text{i.i.d.}\mathcal{N}(0, \sigma^2_{u,s_t}), \tag{40}$$

$$z_t = g + z_{t-1} + v_t, \quad v_t \sim \text{i.i.d.}\mathcal{N}(0, \sigma^2_v). \tag{41}$$

To identify asymmetric business cycle dynamics, we impose restrictions as follows:

$$\delta_{s_t=1} < 0, \;\; \delta_{s_t=2} = 0, \;\; \sigma^2_{u,s_t=2} = 0.$$

The business cycle corresponds to transitory deviations in the output process away from its permanent component. Given that the economy stays in regime 1, the transitory component has a negative unconditional mean because $\delta_{s_t=1}$ is constrained to be negative. The period $s_t = 1$ can be interpreted as the state of the plucked-down economy. Meanwhile, in regime 2, the output fluctuates according to the stochastic trend only, not the business cycle component, because the transitory shock variance, $\sigma^2_{u,s_t=2}$, is equal to zero. ϕ_1 and ϕ_2 determine the persistence of the recession component x_t. g is a deterministic drift in trend.

The plucking model does not contain measurement error, but the measurement equation should include measurement error to apply the APF. To resolve this problem, we should reexpress the model as another state-space form with measurement error. To do this, we eliminate the trend component by multiplying both sides of equation (39) by $(1 - L)$,

$$y_t = g + x_t - x_{t-1} + y_{t-1} + v_t, \quad u_t \sim \text{i.i.d.}\mathcal{N}(0, \sigma^2_v) \tag{42}$$

where L is a lag operator and

$$x_t = \delta_{s_t} + \phi_1 x_{t-1} + \phi_2 x_{t-2} + u_t, \quad u_t \sim \text{i.i.d.}\mathcal{N}(0, \sigma^2_{u,s_t}). \tag{43}$$

For the likelihood inference, the resulting state-space form given the regime is expressed by

$$\beta_t|\beta_{t-1}, \Theta, s_t \sim \mathcal{N}(\mu_{s_t} + G_{s_t}\beta_{t-1}, Q_{s_t}),$$
$$y_t|\mathcal{F}_{t-1}, \Theta, \beta_t, s_t \sim \mathcal{N}(H\beta_t + Fx_t, R_{s_t})$$

where

$$\beta_t = [\ x_t \quad x_{t-1}\ ]',\ \mu_{s_t} = [\ \delta_{s_t} \quad 0\ ]',\ G_{s_t} = \begin{bmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{bmatrix},\ Q_{s_t} = diag(\ \sigma^2_{u,s_t} \quad 0\ ),$$

and

$$H = [\ 1 \quad -1\ ],\ F = [\ g \quad 1\ ],\ x_t = [\ 1 \quad y_{t-1}\ ]',\ R_{s_t} = \sigma^2_v.$$

We now estimate the model using the log of quarterly real GDP data for the United States ranging from 1951:Q1 to 2016:Q1.[4] This model is estimated by the MLE where the likelihoods are computed by the KF method and the log likelihood functions are numerically maximized. To reduce the local-maxima problem, we follow the approach of Chib and Ramamurthy (2010) and use a simulated annealing algorithm combined with a Newton-Raphson method. This numerical optimizer is especially useful in cases where the likelihood is high-dimensional and irregular. Once the likelihood is maximized, the likelihood is computed at the ML estimates using the KF and APF methods. MLE results for the model parameters and likelihood values at the estimates are reported in Table 3.

The log likelihood from the KF method is -567.57 and that from the APF method is -567.29. The approximation error is measured at 0.28, so the likelihood values calculated by the KF and APF methods are almost equal. Moreover, the estimation results are convincing. Figures 5 and 6 plot the estimated stochastic trend and asymmetric business cycles of the log real GDP. These estimates are the filtered values of unobserved components (i.e., $x_{t|t}$ and $z_{t|t}$). During the normal period, the economy is subject mostly to permanent shocks, whereas the transitory component plays an important role in the output fluctuations during recession. On the other hand, Figure 7 presents the filtered

[4]The data come from the FRED database at the Federal Reserve Bank of St. Louis. They are in billions of chained 2009 USD with a seasonally adjusted annual rate.

Table 3: **Parameter Estimates: Friedman's Plucking Model** This table presents the parameter estimates, the log likelihoods (ln L), and the computing time (Time) in seconds.

Parameter		Estimates	t value
g		3.01	2.89
δ_1		-5.60	-1.18
ϕ_1		1.16	0.43
ϕ_2		-0.32	-0.49
σ_v^2		7.19	2.94
$\sigma_{u,1}^2$		0.01	0.06
p_{11}		0.61	19.99
p_{22}		0.96	15.83
$\ln L$	APF		-567.57
	KF		-567.29
Time	APF		386.79
	KF		0.06

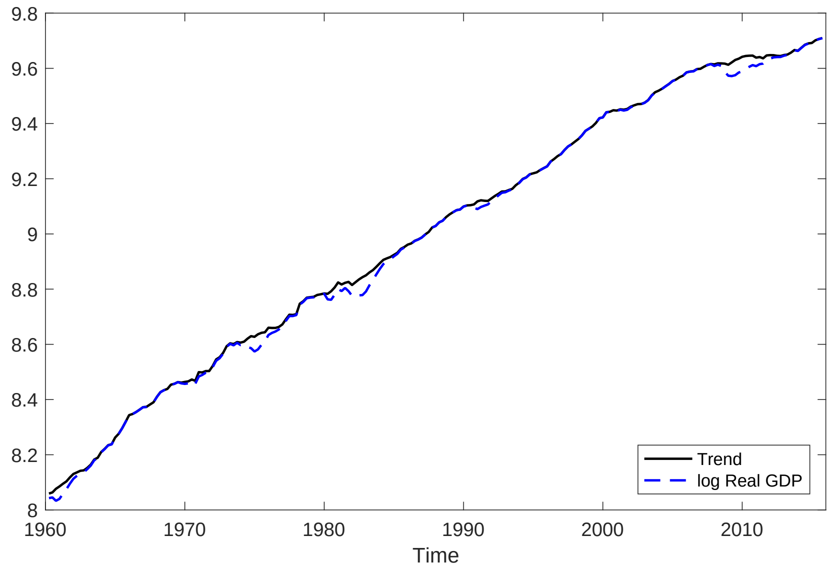
probabilities of recession, $\Pr[s_t = 1|\mathcal{F}_t]$, along with the NBER indicator. The filtered probabilities are strongly correlated with the NBER business cycle dates, and historical recessions seem to be precisely detected by this plucking model.

5.2 Business Cycle Turning Points Model

The business cycle turning points model used in the work of Kim and Nelson (1998) encompasses the two key features of business cycles: one common factor among macroeconomic observations containing information on the state of the economy and the regime-switching dynamics of the common factor. To deal with these characteristics, the model considers a dynamic common factor of economic variables and turning points of the business cycle. Assume that y_{it} represents the first difference of the natural log of the i-th observed coincident indicators at time t. The observations are assumed to be generated by one common factor, denoted by c_t, and their idiosyncratic components are as follows:

$$y_{it} = \gamma_i c_t + e_{it}, \quad i = 1, 2, 3, 4. \tag{44}$$

Figure 5: **The log Real GDP and Estimates of the Permanent Component**
This figure plots the log U.S. real GDP (dotted, blue) along with its stochastic trend (dashed, black).



The common factor c_t and error term e_{it} are assumed to follow an AR(1) process.

$$c_t = \delta_{s_t} + \phi c_{t-1} + v_t, \quad v_t \sim \text{i.i.d.}\mathcal{N}(0, \sigma_v^2), \tag{45}$$

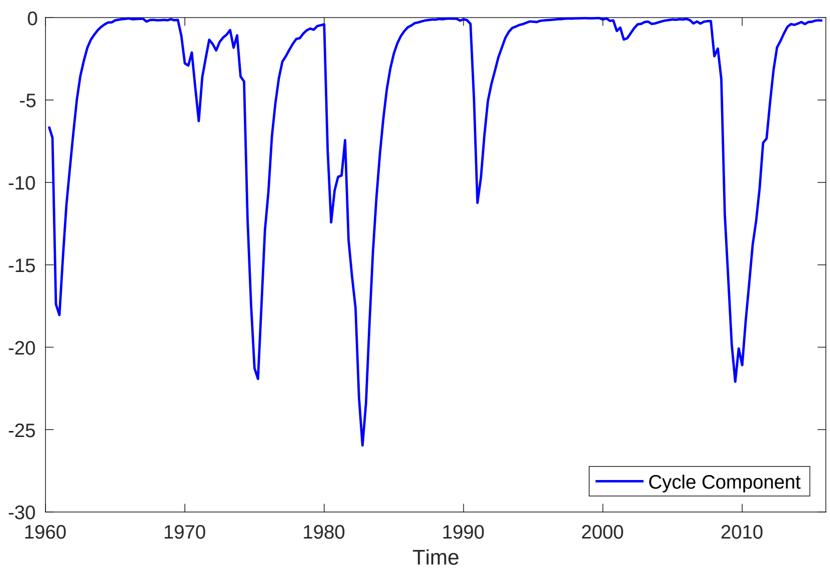
$$e_{it} = \psi_i e_{i,t-1} + \epsilon_{it}, \quad \epsilon_{it} \sim \text{i.i.d.}\mathcal{N}(0, \sigma_i^2), \quad i = 1, 2, 3, 4 \tag{46}$$

where the intercept term in the factor process, δ_{s_t}, is subject to regime shifts, ϕ determines the persistence of the common factor, σ_v^2 is the variance of the factor shock, ψ_i captures the persistence of each shock, and σ_i^2 is the idiosyncratic shock variance. We also assume that v_t and ϵ_{it} are mutually independent. For the factor identification, the first factor loading of the common factor is taken to be unity(i.e., $\gamma_1 = 1$).

The dynamic common factor is extracted by the only source of comovements among the observed macroeconomic variables, so the factor can be interpreted as the business cycle. By allowing the intercept term δ_{s_t} to switch between the two regimes, we can detect the turning points of the business cycles.

As in the plucking model, we multiply both sides of equation (44) by $(1 - \psi_i L)$, and we can eliminate the serial correlation of the idiosyncratic component for consideration

Figure 6: **Transitory Component of the log Real GDP** This figure plots the cyclical component of the log U.S. real GDP.



of the measurement error in the measurement equation,

$$y_{it} = \psi_i y_{i,t-1} + \gamma_i c_t - \psi_i c_{t-1} + \epsilon_{it}, \quad i = 1, 2, 3, 4. \tag{47}$$

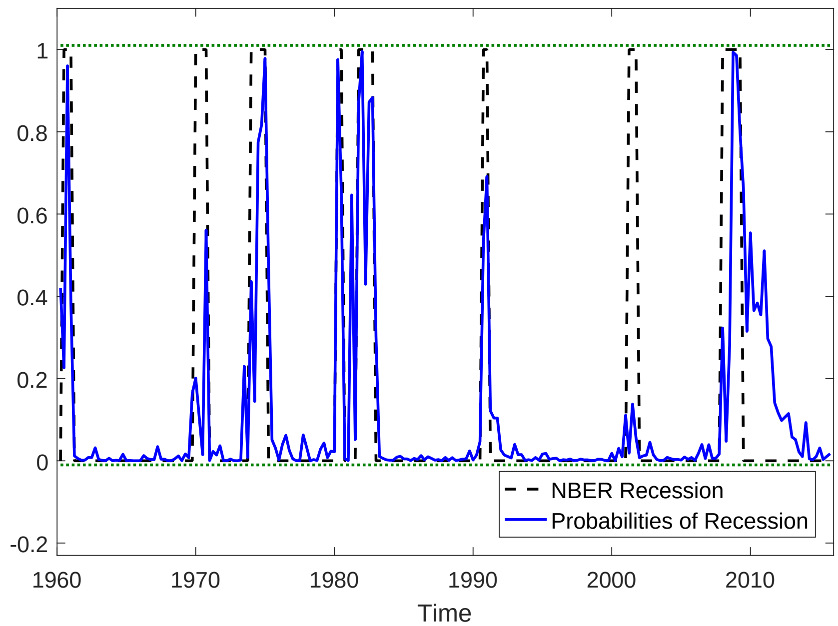
For the likelihood inference, the resulting state-space representation is obtained as

$$\begin{aligned} \beta_t|\beta_{t-1}, \Theta, s_t &\sim \mathcal{N}(\mu_{s_t} + G_{s_t}\beta_{t-1}, Q_{s_t}), \\ y_t|\mathcal{F}_{t-1}, \Theta, \beta_t, s_t &\sim \mathcal{N}(H\beta_t + Fx_t, R_{s_t}) \end{aligned} \tag{48}$$

where

$$\beta_t = [\ c_t \quad c_{t-1}\ ]',\ \mu_{s_t} = [\ \delta_{s_t} \quad 0\ ]',\ G_{s_t} = \begin{bmatrix} \phi & 0 \\ 1 & 0 \end{bmatrix},\ Q_{s_t} = diag[\ \sigma_v^2 \quad 0\ ]',$$

Figure 7: **Probabilities of Recessions: Friedman's Plucking Model** This figure plots the filtered probabilities of recession (dashed, blue) along with the NBER recession indicator (dotted, black) over time



and

$$
\begin{aligned}
H &= \begin{bmatrix} 1 & \gamma_2 & \gamma_3 & \gamma_4 \\ -\psi_1 & -\psi_2\gamma_2 & -\psi_3\gamma_3 & -\psi_4\gamma_4 \end{bmatrix}', \\
F &= diag[\ \psi_1 \quad \psi_2 \quad \psi_3 \quad \psi_4\ ]', \\
x_t &= [\ y_{1,t-1} \quad y_{2,t-1} \quad y_{3,t-1} \quad y_{4,t-1}\ ]', \\
R_{s_t} &= diag[\ \sigma_1^2 \quad \sigma_2^2 \quad \sigma_3^2 \quad \sigma_4^2\ ]'.
\end{aligned}
$$

The data used for estimation pertain to U.S. monthly industrial production, personal income less transfer, manufacturing and trade sales, and non-farm employee data. The time span is January 1967 to August 2014. Table 4 summarizes the estimates of the model parameters and likelihood values from the KF and APF methods. Figures 8 and 9 plot the filtered common factor process and the filtered probability of recession over time, respectively. These recession probabilities are in close agreement with the NBER recession indicator. In addition, the asymmetric nature of business cycle phases is captured because expansions display a longer duration on average than recessions.

Table 4: **Parameter Estimates: Business Cycle Turning Points Model** This table presents the parameter estimates, the log likelihoods (ln L), and the computing time (Time) in seconds.

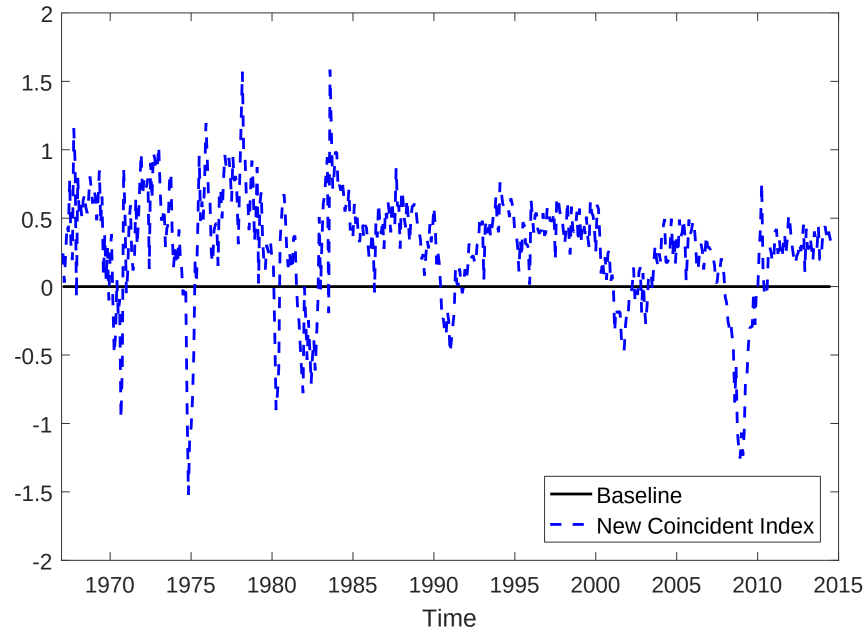
Parameter		Estimates	t value
δ_1		-0.17	-2.76
δ_2		0.16	4.18
ϕ		0.64	9.53
σ_v^2		0.06	5.07
ψ_1		0.11	2.30
ψ_2		-0.16	-3.81
ψ_3		-0.24	-5.79
ψ_4		-0.42	-4.73
γ_2		0.60	11.45
γ_3		0.89	11.90
γ_4		0.46	18.37
σ_1^2		0.32	15.86
σ_2^2		0.33	16.56
σ_3^2		0.75	16.58
σ_4^2		0.01	4.05
p_{11}		0.91	18.59
p_{22}		0.98	136.06
$\ln L$	APF		-1476.43
	KF		-1476.07
Time	APF		4994.36
	KF		0.37

According to Table 4, the KF approximation error in this application is around 0.37, which is very small, similar to the experiment with the plucking model. This finding also strongly supports the reliability of the KF approximation approach. We can conclude that, although the KF approximation does not integrate the entire history of the regimes, this method produces an accurate likelihood value that is close enough to the exact likelihood value.

6 Conclusion

In this paper, we evaluate the KF approximation method for the likelihood inference of dynamic linear models with Markov switching parameters. Its approximation error is measured by the difference between the likelihood values obtained by the KF and APF

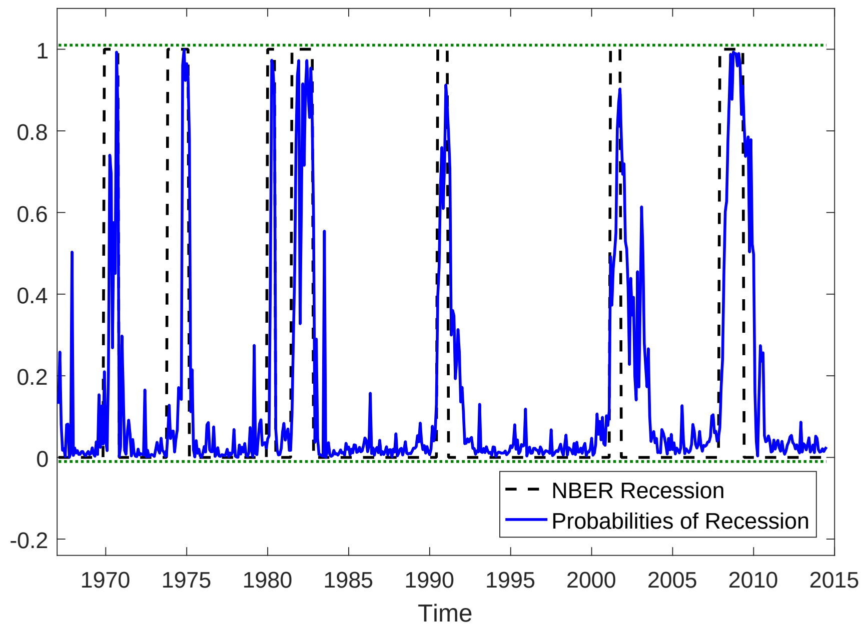
Figure 8: **New Coincident Index** This figure plots the new coincident index extracted from the estimation of the dynamic common factor model with Markov switching parameters.



methods, given that the likelihood from the APF can be rendered sufficiently close to the exact likelihood value obtained when using a large simulation size. According to the likelihood comparison based on our extensive simulation and empirical studies, the KF approximation error is sufficiently small, particularly when the regimes are persistent, and the performance is robust to the sample size and model specifications. Therefore, the KF method is a fast and efficient likelihood inference approach for dynamic linear models with Markov-regime switching parameters. We believe that our work justifies the use of the KF method in the existing literature on MLE of the models.

The KF approximation can be useful for the Bayesian posterior simulation as well. Typically, model parameters are sampled from their full conditional distributions, given regimes and factors. Because of the dependency on the regimes and factors, the parameter sampling efficiency is not high, especially when the models are highly nonlinear to parameters. By using the KF method, we can optimize the likelihood or posterior density integrating over the latent variables, and obtain a proposal distribution that successfully approximates the posterior distribution. Consequently, we may achieve a

Figure 9: **Probabilities of Recession: Business Cycle Turning Points Model**
This figure plots the filtered probability of recession (dashed, blue) along with the NBER recession indicator (dotted, black) over time.



high efficiency of posterior sampling. We will consider the application of the KF method in the Bayesian posterior sampling in future work.

References

Chauvet, M. (1998), "An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching," *International Economic Review*, 39(4), 969–996.

Chib, S., Nardari, F., and Shephard, N. (2002), "Markov chain Monte Carlo methods for stochastic volatility models," *Journal of Econometrics*, 108, PII S0304–4076(01)00137–3.

Chib, S. and Ramamurthy, S. (2010), "Tailored Randomized-block MCMC Methods with Application to DSGE Models," *Journal of Econometrics*, 155, 19–38.

Hamilton, J. (1989), "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 57, 357–84.

Herbst, E. and Schorfheide, F. (2015), "Bayesian Estimation of DSGE Models," *Princeton University Press*.

Kahn, J. A. and Rich, R. W. (2007), "Tracking the new economy: Using growth theory to detect changes in trend productivity," *Journal of Monetary Economics*, 54, 1670–1701.

Kang, K. H. (2014), "Estimation of state-space models with endogenous Markov regime-switching parameters," *Econometrics Journal*, 17, 56–82.

Kang, K. H., Kim, C. J., and Morley, J. (2009), "Changes in U.S. Inflation Persistence," *Studies in Nonlinear Dynamics and Econometrics*, 13, 1–23.

Kim, C. J. (1994), "Dynamic linear models with Markov-switching," *Journal of Econometrics*, 60, 1–22.

Kim, C. J. and Nelson, C. R. (1998), "Business Cycle Turning Points, A New Coincident Index, and Tests of Duration Dependence Based on a Dynamic Factor Model With Regime Switching," *Review of Economics and Statistics*, 80, 188–201.

— (1999a), "Friedman's Plucking Model of Business Fluctuations: Tests and Fstimates of Permanent and Transitory Components," *Journal of Money, Credit, and Banking*, 31(3), 317–334.

— (1999b), "State-Space Models with Regime- Switching: Classical and Gibbs-Sampling Approaches with Applications," *MIT Press*, Cambridge.

Morley, J. and Piger, J. (2012), "The Asymmetric Business Cycle," *Review of Economics and Statistics*, 94(1), 208–221.

Pitt, M. K. and Shephard, N. (1999), "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, 94(446), 590–599.

Timmermann, A. (2001), "Structural Breaks, Incomplete Information, and Stock Prices," *Journal of Business and Economic Statistics*, 19(3), 299–314.