Jaeho Kim[1] / **Sunhyung Lee**[1]

An efficient sequential learning algorithm in regime-switching environments

[1] University of Oklahoma, Department of Economics, 308 Cate Center Drive, Room 158, CCD1, Norman, United States of America, E-mail: jaeho@ou.edu

Abstract:
We provide a novel approach of estimating a regime-switching nonlinear and non-Gaussian state-space model based on a particle learning scheme. In particular, we extend the particle learning method in Liu, J., and M. West. 2001. "Combined Parameter and State Estimation in Simulation-Based Filtering." In *Sequential Monte Carlo Methods in Practice*, 197–223. Springer. by constructing a new proposal distribution for the latent regime index variable that incorporates all available information contained in the current and past observations. The Monte Carlo simulation result implies that our approach categorically outperforms a popular existing algorithm. For empirical illustration, the proposed algorithm is used to analyze the underlying dynamics of US excess stock return.

Keywords: regime switching models, sequential Monte Carlo estimation, particle filters, parameter learning, volatility models

DOI: 10.1515/snde-2018-0016

1 Introduction

A linear state-space model with Markov switching is widely used in the application where dramatic changes in model parameters are prevalent. Kim (1994), for instance, estimates the model with a Kalman filter-based technique. The traditional linear model is no longer adequate, however, because of the presence of non-linearity in many modern day applications. Another feature of current time-series observations is a surge in real-time data, compounding at a rapid rate. A more refined model, consequently, requires two areas of improvement: first, consideration of non-linearity and non-Gaussian shocks in the state-space model under a regime-switching environment; and second, the inclusion of sequential parameter learning and state filtering methodology to accommodate the high rate of real-time data update. We propose a sophisticated sequential parameter learning algorithm that can be used in a generalized regime-switching environment. Then we test the estimation accuracy of our algorithm against a popular alternative method.

The foundations of our proposed approach are novel works of Carvalho and Lopes (2007) and Liu and West (2001). In their seminal paper, Liu and West (2001) introduce a sequential parameter learning method by combining the auxiliary particle filter (APF) of Pitt and Shephard (1999)[1] with a kernel smoothing approach that approximates the posterior distribution of model parameters. Extending the method of Liu and West (2001), Carvalho and Lopes (2007) develop a widely applicable and easily implementable particle learning algorithm to estimate a regime-switching state-space model. Our main contribution is improving the estimation performance of the algorithm in Carvalho and Lopes (2007), which is a culmination of several seminal works in the Sequential Monte Carlo (SMC) literature.[2]

Under the framework of Carvalho and Lopes (2007), we improve the estimation accuracy and the computational efficiency by carefully designing a particle re-sampling procedure and a candidate generating distribution for a regime index variable (i.e. s_t). In their algorithm, the particle re-sampling process and the candidate generating distribution are mainly determined by the regime transition probability. Specifically, a particular regime state that has the highest regime transition probability is selected at time t *given* the regime state at time $t-1$. Based on the chosen regime state at time t, the predictive density of the current data (i.e. y_t) is calculated. The predictive density, which is largely determined by the transition probability, is the main factor in the particle re-sampling step. Unfortunately, the existing approach cannot efficiently identify regime-switching because any regime transition probability higher than 0.5, for instance, would indicate that there is no change in regimes between the two time periods in the re-sampling step.[3] Moreover, the existing approach can be inefficient since the regime transition probability is the only factor that determines particles of the regime index variable at time t.

Jaeho Kim is the corresponding author.

To mitigate the strong dependence on the regime transition probability, we combine the re-sampling step and the particle drawing step of the regime index variable, while utilizing the information set available up to the current period. By using both the regime transition probability and the current data in the combined step, the estimation performance is no longer sensitive to the regime transition probability. Given a reasonable number of particles, the SMC simulation results indicate that the estimation accuracy in Carvalho and Lopes (2007) is greatly compromised when regimes are frequently changing, whereas our estimation strategy performs well regardless of the regime persistence.

For an empirical illustration, we apply the proposed algorithm to investigate the dynamics of US excess stock market returns. We extend the model of Brandt and Kang (2004) to investigate whether dramatic regime changes exist in the conditional variance and mean process. Numerous studies document regime changes in the conditional variance process of the S&P 500 data (e.g. Beltratti and Morana, 2006; So, Lam and Li, 1998) and European stock markets (e.g. Morana and Beltratti, 2002). On the other hand, only a few studies allow for the regime-dependent conditional mean for the US stock returns (e.g. Marcucci 2005) and foreign exchange rates (e.g. Bollen, Gray and Whaley, 2000).

Based on marginal likelihood values, we find that models with regime changes in volatility can best explain the underlying dynamics of the process. We do not, however, find any evidence that regime changes exist in the conditional mean. Moreover, we highlight that the leverage effect itself is an integral element of the stock return analysis even though the model with regime changes in both volatility and the leverage effect is not selected as the best model.

2 Sequential estimation of markov switching state-space models

A nonlinear and non-Gaussian state-space model with regime-switching for a N-dimensional time series, y_t, and state vector, x_t, generally adopts the following specification:

$$y_t = h(x_t, s_t, \epsilon_t) \tag{1}$$

$$x_t = g(x_{t-1}, s_t, \eta_t), \tag{2}$$

where the error terms ϵ_t and η_t are i.i.d. random variables and their means are assumed to be zeros. The measurement equation, $h(\cdot)$, relates the state vector x_t to the observed data, while the transition equation, $g(\cdot)$, shows the dynamics of x_t. Both the measurement and transition equations are determined by the parameter set, β_{s_t}, whose values depend on the regime state, $s_t \in \{0, 1, \dots, K-1\}$. The dynamic system changes between K regimes over time, while the latent variable that determines the current regime, s_t, follows a first-order Markovian process as given below:

$$\pi_{k,j} = p(s_t = j | s_{t-1} = k), \tag{3}$$

where $\prod_{j=0}^{K-1} \pi_{k,j} = 1$. We can also allow for structural breaks in the nonlinear state-space model by imposing restrictions on the transition probabilities as in Chib (1998). Let π be the set of transition probabilities. The main goal of our proposed estimation strategy is to sequentially estimate the unknown model parameters, $\theta = [\beta_0', \beta_1', \dots, \beta_{K-1}', \pi]'$, and the latent states, $[s_t, x_t]$, by incorporating new observations at each time period.[4]

Liu and West (2001) combine the auxiliary particle filter with kernel smoothing methods, which is the foundation of our sequential parameter learning approach. We summarize the Liu-West (LW) filter in Algorithm 1 with the inclusion of the distribution for s_t.[5]

Algorithm 1 Liu-West (LW) Filter for Markov Switching state-space Models

i. Generate $\{s_0^{(i)}, x_0^{(i)}, \theta_0^{(i)}\}$ with the importance weight, $\widehat{\omega}_0^{(i)} = \frac{1}{N}$ for $i = 1, 2, \dots, N$.

ii. Compute the re-sampling weight $\widehat{\omega}_{t-1|t}^{(i)}$ for $i = 1, 2, \dots, N$.

iii. Re-sample N particles $\{\hat{s}_{t-1}^{(i)}, \hat{x}_{t-1}^{(i)}, \hat{\theta}_{t-1}^{(i)}\}_{i=1}^N$ from $\{s_{t-1}^{(i)}, x_{t-1}^{(i)}, \theta_{t-1}^{(i)}\}_{i=1}^N$ and draw $\{s_t^{(i)}, x_t^{(i)}, \theta_t^{(i)}\}$ conditional on $\{\hat{s}_{t-1}^{(i)}, \hat{x}_{t-1}^{(i)}, \hat{\theta}_{t-1}^{(i)}\}$ for $i = 1, 2, \dots, N$.

iv. Compute the normalized importance weight, $\hat{\omega}_t^{(i)}$, for the particle set $\{s_t^{(i)}, x_t^{(i)}, \theta_t^{(i)}\}$ for $i = 1, 2, \dots, N$. Iterate steps (ii), (iii), and (iv) at $t = 1, 2, \dots, T$.

In Algorithm 1, N represents the number particles of the latent variables. At step (i) of Algorithm 1, the particle set $\{s_0^{(i)}, x_0^{(i)}, \theta_0^{(i)}\}$ is orderly drawn from $p(\theta_0^{(i)})$, $p(s_0^{(i)} \mid \theta_0^{(i)})$, and $p(x_0^{(i)} \mid s_0^{(i)}, \theta_0^{(i)})$. Note that $p(s_0^{(i)} \mid \theta_0^{(i)})$ and $p(x_0^{(i)} \mid s_0^{(i)}, \theta_0^{(i)})$ are the unconditional distribution of $s_0^{(i)}$ and $x_0^{(i)}$. To estimate the unknown parameter set $\theta = [\beta_0', \beta_1', \dots, \beta_{K-1}', \pi]'$ at time $t-1$, the LW filter incorporates the following mixture of multivariate normal distributions:

$$p(\theta|y_{1:t-1}) \approx \sum_{i=1}^{N} \hat{\omega}_{t-1}^{(i)} f(\theta; m_{t-1}^{(i)}, h^2 V_{t-1}), \tag{4}$$

where $f(\theta; m_{t-1}^{(i)}, h^2 V_{t-1})$ is a normal distribution with the mean, $m_{t-1}^{(i)}$, and the variance, $h^2 V_{t-1}$. Note that $m_{t-1}^{(i)} = \alpha\theta_{t-1}^{(i)} + (1-\alpha)\bar{\theta}_{t-1}$, where $\bar{\theta}_{t-1} = \sum_{i=1}^{N} \hat{\omega}_{t-1}^{(i)} \theta_{t-1}^{(i)}$. Also $h^2 V_{t-1} = h^2 \sum_{i=1}^{N} \hat{\omega}_{t-1}^{(i)} (\theta_{t-1}^{(i)} - \bar{\theta}_{t-1})(\theta_{t-1}^{(i)} - \bar{\theta}_{t-1})'$, where $h^2 = (1 - \alpha^2)$. The tuning parameter, α, appears in the mean through $m_{t-1}^{(i)}$ (i.e. a shrinkage factor) and in the variance through h^2 (i.e. a smoothing factor).

Carvalho and Lopes (2007) extend the LW filter by including the regime-state, s_t. An important feature of their algorithm is that the estimation of s_t is heavily dependent on the previous state through the prior transition density. We explain Carvalho and Lopes (2007)'s approach in Algorithm 2. Notice that the derivation in Algorithm 2 starts from step (ii.1) because the first step is the same as step (i) of the LW filter derivation in Algorithm 1.

Algorithm 2 Carvalho and Lopes (2007)

ii.1 Compute posterior mean and variance statistics, $m_{t-1}^{(i)}$ and $h^2 V_{t-1}$.

ii.2 Generate $\tilde{s}_t^{(i)} = argmax_{k \in \{0,\dots,K-1\}} p(s_t = k \mid s_{t-1}^{(i)}; m_{t-1}^{(i)})$ for $i = 1, 2, \dots, N$.

ii.3 Compute $\tilde{x}_t^{(i)} = g(x_{t-1}^{(i)}, \tilde{s}_t^{(i)}, \eta_t = 0; m_{t-1}^{(i)})$ for $i = 1, 2, \dots, N$.

ii.4 Compute the re-sampling importance weight, $\hat{\omega}_{t-1|t}^{(i)} = \frac{\omega_{t-1|t}^{(i)}}{\sum_{j=1}^{N} \omega_{t-1|t}^{(j)}}$ for $i = 1, 2, \dots, N$, where ${\omega_{t-1|t}}^{(i)} \propto p(y_t|\tilde{x}_t^{(i)}, \tilde{s}_t^{(i)}; m_{t-1}^{(i)}) \hat{\omega}_{t-1}^{(i)}$, and $p(y_t|\tilde{x}_t^{(i)}, \tilde{s}_t^{(i)}; m_{t-1}^{(i)})$ is the conditional density of y_t given $\tilde{x}_t^{(i)}$, $\tilde{s}_t^{(i)}$, and $m_{t-1}^{(i)}$.

iii.1 Re-sample the particle set, $\{\tilde{s}_t^{(i)}, \hat{x}_t^{(i)}, \hat{s}_{t-1}^{(i)}, \hat{x}_{t-1}^{(i)}, \hat{\theta}_{t-1}^{(i)}, \hat{m}_{t-1}^{(i)}\}_{i=1}^{N}$ from $\{\tilde{s}_t^{(i)}, \tilde{x}_t^{(i)}, s_{t-1}^{(i)}, x_{t-1}^{(i)}, \theta_{t-1}^{(i)}, m_{t-1}^{(i)}\}_{i=1}^{N}$ using $\{\hat{\omega}_{t-1|t}^{(i)}\}_{i=1}^{N}$. Define $\tilde{s}_t^{(i)} = \hat{s}_t^{(i)}$; $\tilde{x}_t^{(i)} = \hat{x}_t^{(i)}$; $s_{t-1}^{(i)} = \hat{s}_{t-1}^{(i)}$; $x_{t-1}^{(i)} = \hat{x}_{t-1}^{(i)}$; $\theta_{t-1}^{(i)} = \hat{\theta}_{t-1}^{(i)}$; and $m_{t-1}^{(i)} = \hat{m}_{t-1}^{(i)}$ for $i = 1, 2, \dots, N$.

iii.2 Generate $\theta_t^{(i)}$ from $N(m_{t-1}^{(i)}, h^2 V_{t-1})$ for $i = 1, 2, \dots, N$.

iii.3 Generate $s_t^{(i)}$ from $p(s_t \mid s_{t-1}^{(i)}; \theta_t^{(i)})$ for $i = 1, 2, \dots, N$.

iii.4 Generate $x_t^{(i)}$ from $p(x_t \mid s_{t-1}^{(i)}, s_t^{(i)}, x_{t-1}^{(i)}; \theta_t^{(i)})$ for $i = 1, 2, \dots, N$.

iv Compute the importance weight $\hat{\omega}_t^{(i)} = \frac{\omega_t^{(i)}}{\sum_{j=1}^{N} \omega_t^{(j)}}$, where $\omega_t^{(i)} \propto \frac{p(y_t|x_t^{(i)}, s_t^{(i)}; \theta_t^{(i)})}{p(y_t|\tilde{x}_t^{(i)}, \tilde{s}_t^{(i)}; m_{t-1}^{(i)})}$.

It is important to note that $\tilde{s}_t^{(i)}$, which is solely determined by the regime transition probability, $p(s_t \mid s_{t-1}^{(i)})$, is used to generate $\tilde{x}_t^{(i)}$ and to compute the re-sampling weight $\hat{\omega}_{t-1|t}^{(i)}$. Moreover, the new particle, $s_t^{(i)}$, is generated from the same transition probability. Because of the heavy dependency of Algorithm 2 on the transition probability, we conjecture that its performance is substantially influenced by the degree of persistence in each regime. A small number of particles, in particular, would exacerbate the problem.

To mitigate the problem, we relax the restriction of relying on the regime transition probability as a main factor that determines the re-sampling weights and s_t in the proposal distribution. Instead, we utilize all of the available information set, including the current observation y_t in the merged process that combines the step for re-sampling particles at time $t-1$ with the step of generating new s_t particles. Algorithm 3 summarizes of our approach.[6] Notice that the derivation in Algorithm 3 starts from step (ii.1) because the first step is the same as step (i) of the LW filter derivation in Algorithm 1.

Algorithm 3 Proposed Algorithm

ii.1 Compute posterior mean and variance statistics, $m_{t-1}^{(i)}$ and $h^2 V_{t-1}$.

ii.2 Compute $\tilde{x}_{t|k}^{(i)} = g(x_{t-1}^{(i)}, s_t = k, \eta_t = 0; m_{t-1}^{(i)})$ for $k = 0, 1, \dots, K-1$ and $i = 1, 2, \dots, N$.

ii.3 Compute the importance weight $\hat{\omega}_{t-1|t,k}^{(i)} = \frac{\omega_{t-1|t,k}^{(i)}}{\sum_{k=0}^{K-1}\sum_{i=1}^{N}\omega_{t-1|t,k}^{(i)}}$ for $k = 0, 1, \dots, K-1$ and $i = 1, 2, \dots, N$, where $\omega_{t-1|t,k}^{(i)} \propto p(y_t \mid \tilde{x}_{t|k}^{(i)}, s_t = k; m_{t-1}^{(i)}) p(s_t = k | s_{t-1}^{(i)}; m_{t-1}^{(i)}) \hat{\omega}_{t-1}^{(i)}$.

iii.1 Draw $\{s_t^{(i)}\}_{i=1}^N$ and re-sample the particle set, $\{\hat{x}_{t|k}^{(i)}, \hat{s}_{t-1}^{(i)}, \hat{x}_{t-1}^{(i)}, \hat{\theta}_{t-1}^{(i)}, \hat{m}_{t-1}^{(i)}\}_{i=1}^N$ simultaneously using $\hat{\omega}_{t-1|t,k}^{(i)}$. The re-sampling is preformed on $\{\tilde{x}_{t|k}^{(i)}, s_{t-1}^{(i)}, x_{t-1}^{(i)}, \theta_{t-1}^{(i)}, m_{t-1}^{(i)}\}_{i=1}^N$. Define $\tilde{x}_{t|k}^{(i)} = \hat{x}_{t|k}^{(i)}$; $s_{t-1}^{(i)} = \hat{s}_{t-1}^{(i)}$; $x_{t-1}^{(i)} = \hat{x}_{t-1}^{(i)}$; and $\theta_{t-1}^{(i)} = \hat{\theta}_{t-1}^{(i)}$.

iii.2 Generate $\theta_t^{(i)}$ from $N(m_{t-1}^{(i)}, h^2 V_{t-1})$ for $i = 1, 2, \dots, N$.

iii.3 Generate $x_t^{(i)}$ from $p(x_t \mid s_t^{(i)}, x_{t-1}^{(i)}; \theta_t^{(i)})$ for $i = 1, 2, \dots, N$.

iv Compute the importance weight $\hat{\omega}_t^{(i)} = \frac{\omega_t^{(i)}}{\sum_{j=1}^N w_t^{(j)}}$, where $\omega_t^{(i)} \propto \frac{p(y_t | x_t^{(i)}, s_t^{(i)}; \theta_t^{(i)}) p(s_t^{(i)} | s_{t-1}^{(i)}; \theta_t^{(i)})}{p(y_t | \tilde{x}_{t|s_t^{(i)}}^{(i)}, s_t^{(i)}; m_{t-1}^{(i)}) p(s_t^{(i)} | s_{t-1}^{(i)}; m_{t-1}^{(i)})}$.

The main difference between the existing particle learning algorithm in Carvalho and Lopes (2007) (i.e. Algorithm 2) and our proposed approach in Algorithm 3 lies on steps (ii), (iii), and (iv). Specifically, at step (iii.1) of Algorithm 3, we combine the re-sampling step for the existing particles at time $t-1$ and the sampling step for s_t. This step is the key to understanding how the proposed algorithm can overcome the aforementioned problem of the existing algorithm. First, our approach does not require the generation of deterministic $\tilde{s}_t^{(i)}$, which critically depends on the transition probability. Second, our approach employs more information contained in the current observation y_t in generating s_t compared to Algorithm 2.

3 Simulation study

We consider a two-state (i.e. $K = 2$) Markov Switching Stochastic Volatility (MSSV) model proposed in So, Lam, and Li (1998) to evaluate the performance of the two described online estimation algorithms:

$$y_t = \beta + \exp(\frac{x_t}{2})\epsilon_t, \ \epsilon_t \sim N(0,1) \tag{5}$$

$$x_t = \alpha_{s_t} + \phi(x_{t-1} - \alpha_{s_{t-1}}) + \eta_t, \ \eta_t \sim N(0, \sigma^2), \tag{6}$$

where $\alpha_{s_t} = \alpha_0 + \alpha_d s_t$ and $s_t \in \{0, 1\}$ for $t = \{1, 2, \dots, T\}$. Without loss of generality, regime 0 or $s_t = 0$ refers to a low-volatility state, whereas regime 1 or $s_t = 1$ refers to a high-volatility regime. Following the data generating process in equations (5) and (6), we use the parameter values of $\{\alpha_0 = 1, \alpha_d = 3, \phi = 0.5, \sigma^2 = 0.5\}$ with $T = 1000$ over a set of 100 simulation studies. We explore two different sets of transition probabilities. Case 1 considers a volatility process that experiences a relatively frequent regime switch. The corresponding transition probability values are $\pi_{00} = 0.9$ and $\pi_{11} = 0.85$ with the expected regime duration of 10 and 7 for regimes 0 and 1, respectively.[7] On the other hand, Case 2 considers a volatility process that has a more persistent state and uses the probability values of $\pi_{00} = 0.99$ and $\pi_{11} = 0.9$. The corresponding expected regime duration is 100 for regime 0 and 10 for regime 1. For both cases, we generate the entire sequence of true values of s_t based on the computed expected regime duration.

To compare the estimation accuracy of the simulation results, we define the Mean Squared Error (MSE) for volatility as $MSE_V^{(j)} = \frac{1}{T}\sum_{t=1}^{T}(V_t - \hat{V}_t^{(j)})^2$, which represents the difference between the real volatility process (i.e. $V_t = exp(x_t)$) and the filtered process (i.e. $\hat{V}_t = exp(\hat{x}_t)$) in the j-th simulation.[8] We compute the average volatility MSE (i.e. $\bar{MSE}_V = \frac{1}{100}\sum_{j=1}^{100} MSE_V^{(j)}$) to summarize all simulation results.

Moreover, to compare how well each algorithm is able to correctly capture changes in regimes and successfully recognize the current state, we define the Quadratic Probability Score (QPS) for the j-th simulation as $QPS^{(j)} = \frac{1}{T}\sum_{t=1}^{T}(s_t - \hat{Pr}(s_t = 1)^{(j)})^2 \times 100$, where $\hat{Pr}(s_t = 1)$ is the filtered probability of the high volatility regime (i.e. estimated state) and s_t is the true value. We can summarize the QPS by averaging over the repeated simulation number of 100 (i.e. $\bar{QPS} = \frac{1}{100}\sum_{j=1}^{100} QPS^{(j)}$). The score ranges between 0 and 100. QPS is equal to 0

if the online estimation algorithm is able to perfectly capture and recognize the regime changes, while 1 being the other side of the extreme.

To compare the performance of the parameter estimation, we define the MSE for model parameters[9] as $MSE_P^{(j)} = \frac{1}{T}\sum_{t=1}^{T}(p-\hat{p}_t^{(j)})^2$ where p is the true parameter value and $\hat{p}_t$ is the estimated parameter path. Similar to the volatility MSE and QPS, the average MSE for each parameter can be written as $\bar{MSE}_P = \frac{1}{100}\sum_{i=j}^{100} MSE_P^{(j)}$.

We compare the performance of each algorithm under different regime frameworks based on the average values of MSE_V, QPS, and MSE_P in Table 1. For a direct comparison, each filter utilizes the same number of particles ($N = 5000$). We use non-informative priors for the model parameters in the simulation.[10] Columns (1) and (4) are from the algorithm of Carvalho and Lopes (2007) and columns (2) and (5) are from our proposed algorithm. Column (3) and (6) represent the percent change from the algorithm used in Carvalho and Lopes (2007) to our algorithm for Case 1 and Case 2, respectively.

Table 1: Averages of MSE_V, QPS, and MSE_P for the 2-state MSSV Model.

	Case 1: $\pi_{00} = 0.9$, $\pi_{11} = 0.85$			Case 2: $\pi_{00} = 0.99$, $\pi_{11} = 0.9$		
	(1) Carvalho and Lopes (2007)	(2) Proposed	(3) %Δ	(4) Carvalho and Lopes (2007)	(5) Proposed	(6) %Δ
$\bar{MSE}_V$	3811	3787	−0.63%	640.67	644.72	0.63%
$\bar{QPS}$	34.773	16.081	−53.75%	6.254	6.075	−2.86%
$\bar{MSE}_P$: α_0	1.535	0.974	−36.55%	0.0641	0.0636	−0.78%
$\bar{MSE}_P$: α_d	0.540	0.421	−22.04%	0.489	0.440	−10.02%
$\bar{MSE}_P$: ϕ_0	0.0231	0.0161	−30.30%	0.0238	0.0242	1.68%
$\bar{MSE}_P$: σ^2	0.990	0.205	−79.29%	0.0369	0.0206	−44.17%
$\bar{MSE}_P$: π_{00}	0.00203	0.00108	−46.80%	0.00062	0.00059	−4.84%
$\bar{MSE}_P$: π_{11}	0.0124	0.0026	−79.03%	0.00110	0.00109	−0.91%

$\bar{MSE}_V$ refers to the average mean squared error for volatility. $\bar{QPS}$ refers to the average quadratic probability score. The QPS index ranges between 0 and 100, with 0 being the case of correct assignment of the state variable for all time periods and 100 being the opposite case. $\bar{MSE}_P$ refers to the MSE of each parameter. The results are based on the 2-state MSSV model with $\alpha_0 = 1$, $\alpha_d = 3$, $\phi = 0.5$, $\sigma^2 = 0.5$ using 5000 particles, averaged over 100 simulations. %Δ refers to the percent changes from column (1) to column (2) and from column (4) to column (5), where the negative values indicate a percent decrease, whereas positive values indicate a percent increase.

The overall result indicates that when there is a high turnover rate in regime-switching (i.e. Case 1), the proposed algorithm clearly outperforms the existing algorithm in estimation accuracy. Particularly, the average MSE_V value decreases for Case 1. This implies that the proposed approach can better estimate the volatility process given the same number of particles. Furthermore, the average QPS value of the proposed filter is nearly twice as small as that of the existing algorithm (i.e. a decrease of 53.75%). This reduction is significantly greater than the reduction in Case 2 (i.e. −2.86%). Because a lower value of the average QPS equates to less erroneous state estimations, the proposed algorithm can better determine the position of the current state with the reasonable number of particles. When regime-switching is stagnant (i.e. Case 2), the average MSE_V and the average QPS values of the proposed approach are not very different from those of the existing approach.[11]

The parameter estimation accuracy of the proposed approach is also significantly better than the existing algorithm given $N = 5000$ in Case 1. For instance, regime-dependent volatility means (i.e. α_0 and α_d) are better estimated with the proposed approach with the average MSE value of 0.974 and 0.421 compared with values of 1.535 and 0.540 of the existing filter, respectively. Moreover, the existing filter produces a higher average MSE for σ^2. This result can be explained by the fact that Algorithm 2 often spuriously attributes variations in volatility to the shock, η_t, because it fails to capture regime changes in the mean of the volatility process. On the other hand, the values of average MSE for the model parameters under persistent regimes (i.e. Case 2) are similar between the two approaches.[12] The comparison of the percent changes between Case 1 and Case 2 clearly indicates that the estimation accuracy is far better improved in Case 1 than in Case 2 when we use our proposed algorithm.

One important implication from the findings is that the existing algorithm performs very differently depending on time series characteristics of regimes. Notably the new algorithm constantly performs well. In many different empirical analyses, we may observe various transition probabilities for regimes. The proposed algorithm will be useful in practice regardless of underlying persistence of regimes.

4 Application

4.1 Empirical specification

For an empirical illustration, we apply the proposed algorithm to investigate whether the dynamics of US stock market excess returns can be well characterized by a model with abrupt regime changes.[13] For our analysis, we employ the value-weighted portfolio returns of NYSE, AMEX, and NASDAQ firms minus the 1-month Treasury bill rate.[14] The data are observed at the weekly level from the week of January 4, 1980 to the week of May 26, 2017, a total of 1952 observations.

We consider the following model of excess stock returns:

$$r_t^e = \mu_t^r + \sigma_t \epsilon_t,\ \epsilon_t \sim NID(0,1), \tag{7}$$

where the excess stock returns, r_t^e, are made up of the expected value of excess returns, μ_t^r, and an unexpected random shock, $\sigma_t \epsilon_t$. An economic agent forms her expectations of r_t^e at time $t-1$ using the information available up to time $t-1$. However, μ_t^r is assumed to be a latent variable and is estimated within the model because the information set available to a researcher is substantially smaller than the amount of information available to the economic agent. The unobserved (log) mean, m_t, and the (log) volatility, x_t, of excess returns are determined by the following latent vector autoregressive (VAR) process:

$$m_t = \mu^m_{s_t^m} + \phi_{11}(m_{t-1} - \mu^m_{s_{t-1}^m}) + \phi_{12}(x_{t-1} - \mu^x_{s_{t-1}^x}) + e_t^m, \tag{8}$$

$$x_t = \mu^x_{s_t^x} + \phi_{21}(m_{t-1} - \mu^m_{s_{t-1}^m}) + \phi_{22}(x_{t-1} - \mu^x_{s_{t-1}^x}) + \phi_{23,s_t^x}\epsilon_{t-1} + e_t^x, \tag{9}$$

$$\begin{bmatrix} e_t^m \\ e_t^x \end{bmatrix} \sim NID\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_m^2 & \rho\sigma_m\sigma_x \\ \rho\sigma_m\sigma_x & \sigma_x^2 \end{bmatrix} \right), \tag{10}$$

where $\mu^m_{s_t^m} = \mu_0^m + \mu_d^m\, s_t^m$; $\mu^x_{s_t^x} = \mu_0^x + \mu_d^x\, s_t^x$; $s_t^m \in \{0,1\}$; and $s_t^x \in \{0,1\}$.[15] Under the VAR representation, the conditional mean and volatility are specified as $\mu_t^r = \exp(m_t)$ and $\sigma_t^2 = \exp(x_t)$. We are extending the latent VAR model in Brandt and Kang (2004) by incorporating abrupt but recurring regime changes in μ^m and μ^x with the regime indicator variables s_t^m and s_t^x.[16] The regime-dependent parameter, ϕ_{23,s_t^x}, plays an important role in capturing the leverage effect. For example, the financial leverage of a firm increases when negative news lowers the market value of the firm. An increase in the finantial leverage leads to a higher risk of holding the equity claims, which results in a rise in the volatility.[17] By allowing for the regime switching in ϕ_{23,s_t^x}, we can test for whether the magnitude of the leverage effect depends on the high and low volatility regimes.

The discrete state variables follow first-order Markovian processes of

$$\pi_{00}^m = p[s_t^m = 0 \mid s_{t-1}^m = 0], \quad \pi_{11}^m = p[s_t^m = 1 \mid s_{t-1}^m = 1], \tag{11}$$

$$\pi_{00}^x = p[s_t^x = 0 \mid s_{t-1}^x = 0], \quad \pi_{11}^x = p[s_t^x = 1 \mid s_{t-1}^x = 1] \tag{12}$$

For the purpose of identifying regimes, we assume that $\mu_d^m > 0$ and $\mu_d^x > 0$. Accordingly, $s_t^m = 0$ ($s_t^x = 0$) represents a state of low conditional mean (low volatility), whereas $s_t^m = 1$ ($s_t^x = 1$) represents a state of high conditional mean (high volatility). Algorithm 1 and Algorithm 3 allow us to estimate the above nonlinear state-space model with regime-switching. The priors that we employed are reported in Table 2.[18] In estimation, we use 100,000 particles to ensure that our method fully explores the spaces of parameters and the latent states.

Table 2: Parameter comparisons by model specification.

	Model A		**Model B**		**Model C**		**Model D**	
Regime Δ in Vol.	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Regime Δ in Mean	No	No	No	No	No	No	Yes	Yes

Regime Δ in Leverage		No	No	No	No	Yes	Yes	No	No
Prior		**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
μ_0^m	$\mathcal{N}(ln\hat{M}-1,2^2)$	−2.314	0.376	−2.040	0.306	−2.022	0.325	−4.036	0.385
μ_d^m	$\mathcal{TN}(2,2^2)$	–	–	–	–	–	–	2.235	0.298
μ_0^x	$\mathcal{N}(ln\hat{V}-1,2^2)$	1.269	0.082	0.852	0.069	0.804	0.076	0.776	0.078
μ_d^x	$\mathcal{TN}(2,2^2)$	–	–	0.995	0.121	1.114	0.114	1.191	0.101
ϕ_{11}^m	$\mathcal{N}(0.9,0.5^2)$	0.528	0.131	0.502	0.113	0.827	0.042	0.859	0.034
ϕ_{12}^m	$\mathcal{N}(0,1^2)$	−0.218	0.090	−0.064	0.120	0.035	0.090	0.067	0.120
ϕ_{21}^x	$\mathcal{N}(0,1^2)$	−0.205	0.073	0.003	0.100	−0.063	0.086	−0.023	0.089
ϕ_{22}^x	$\mathcal{N}(0.9,0.5^2)$	0.830	0.052	0.838	0.046	0.876	0.026	0.814	0.044
$\phi_{23,0}^x$	$\mathcal{N}(0,0.1^2)$	−0.198	0.020	−0.245	0.024	−0.187	0.031	−0.255	0.024
$\phi_{23,1}^x$	$\mathcal{N}(0,0.1^2)$	–	–	–	–	−0.223	0.031	–	–
ρ	$\mathcal{TN}(0,1^2)$	−0.026	0.217	−0.195	0.258	−0.026	0.255	−0.085	0.298
σ_m^2	$\mathcal{IG}(5,0.05)$	0.042	0.010	0.043	0.011	0.015	0.003	0.016	0.004
σ_x^2	$\mathcal{IG}(5,0.05)$	0.068	0.013	0.054	0.013	0.015	0.005	0.013	0.004
π_{00}^m	$\mathcal{B}e(98,2)$	–	–	–	–	–	–	0.979	0.006
π_{11}^m	$\mathcal{B}e(98,2)$	–	–	–	–	–	–	0.980	0.006
π_{00}^x	$\mathcal{B}e(98,2)$	–	–	0.986	0.004	0.987	0.005	0.988	0.004
π_{11}^x	$\mathcal{B}e(98,2)$	–	–	0.982	0.005	0.983	0.005	0.978	0.006
log(ML)		−4105.9		−4101.3		−4105.5		−4107.2	

Refer to Eqs. (8), (9), (10), (11), and (12). Model A assumes no regime change. Model B assumes a regime change in μ^s only. Model C assumes regime changes in μ^s and ϕ_3^x. Model D assumes regime changes in μ^s and μ^m. The value of log(ML) refers to the log of marginal likelihood. The sample period ranges between the week of January 4, 1980 and the week of May 26, 2017. A total of 100,000 particles are used to obtain the estimates. The values of $ln\hat{M}$ and $ln\hat{V}$ represent the logs of the sample mean and variance of y_t. The truncated normal distributions for μ_d^m and μ_d^x are defined between 0 and infinity. The truncated normal distribution ρ is defined between −1 and 1.

4.2 Empirical results

Our empirical analysis accounts for various possibilities of innovations under which the conditional mean and volatility may perform. In particular, the generalized environment can be categorized into three broad cases: a regime change in volatility, a regime change in the conditional mean, and a regime change in the leverage effect. Model A assumes no regime change at all. All other cases assume regime-switching either in the conditional mean and volatility of the market excess return. For instance, Model B assumes a regime change in μ^s only; Model D assumes regime changes in both μ^s and μ^m; and Model C assumes regime changes in μ^s and ϕ_{23} to incorporate a possibility of a regime change in the leverage effect. To determine which model best fits with the data in a parsimonious model setup, we compute the log marginal likelihoods (log(ML)) at the terminal time period of each case.[19]

Comparing the log(ML) values of the four models determines if regime switching occurs and where it occurs in the excess return process. Table 2 presents the log(ML) values. The log(ML) value of −4105.9 in Model A (i.e. no regime-switching) is smaller than Model B (i.e. regime-switching in volatility), which is evidence that the regime change exists in the conditional volatility process. Comparing Model B (i.e. a change in volatility only) and Model D (i.e. changes in the conditional mean and the volatility) determines whether a regime change exists in the conditional mean process. The log(ML) value of −4101.3 in Model B is higher than the log(ML) value of −4107.2 in Model D. Again, this finding suggests that the regime change occurs only in the volatility process. Given the evidence in favor of regime-switching in volatility, we can examine whether a regime-switching occurs in both the volatility process and the leverage effect. Comparing Model B (i.e. the leverage effect is constant) versus Model C (i.e. the leverage effect changes according to the volatility regimes) allows us to determine whether the leverage effect changes over time. The log(ML) of −4101.3 in Model B is greater than the log (ML) of −4105.5 in Model C, which implies that the regime switching in the leverage effect is not a main feature of the excess market return.[20]

This evidence does not discount the importance of the leverage effect in the model. Rather, the leverage effect itself is an integral element of the stock return analysis, as $\phi_{23,0}^x$ is significant across all model specifications. Note

that according to the model comparison criterion suggested by Kass and Raftery (1995) and Raftery (1995), Model B is strongly or very strongly preferred to other models.[21]

We depict the movements of the weekly excess return, the filtered stochastic volatility, $E[\sigma_t^2 \mid y_{1:t}]$, and the filtered probability of the high-volatility regime, $\Pr(s_t = 1 \mid y_{1:t})$, from the week of January 4, 1980 to the week of May 26, 2017 in Figure 1 using the specification from Model B, the best model. The weekly return (i.e. top graph) oscillates quite violently throughout the sample period. The degree of fluctuation is particularly large for periods between the years of '87-89', '99-01', and '09-11' (i.e. periods of the second oil shock, the dot-com boom, and the global financial crisis, respectively). When matching the weekly market return with the volatility graph (i.e. the mid graph), it is evident that the cycles with high fluctuations are closely associated with the periods that exhibit high volatility. The volatility is at its peak during the recent financial crisis. When we compare the volatility graph with the filtered regime state graph (i.e. the bottom graph), it is evident that the level of volatility and the probability of the high-volatility regime exhibit strong co-movements. Note that the volatility regime is not well identified in the early sample periods due to the insufficient number of observations. However, since the mid-1990s, when observations are accumulated, the volatility regime is more accurately estimated by incorporating more observations that contain regime switching signals. Our estimated regimes fairly match up to those of Ang and Timmermann (2012) over the same sample periods.

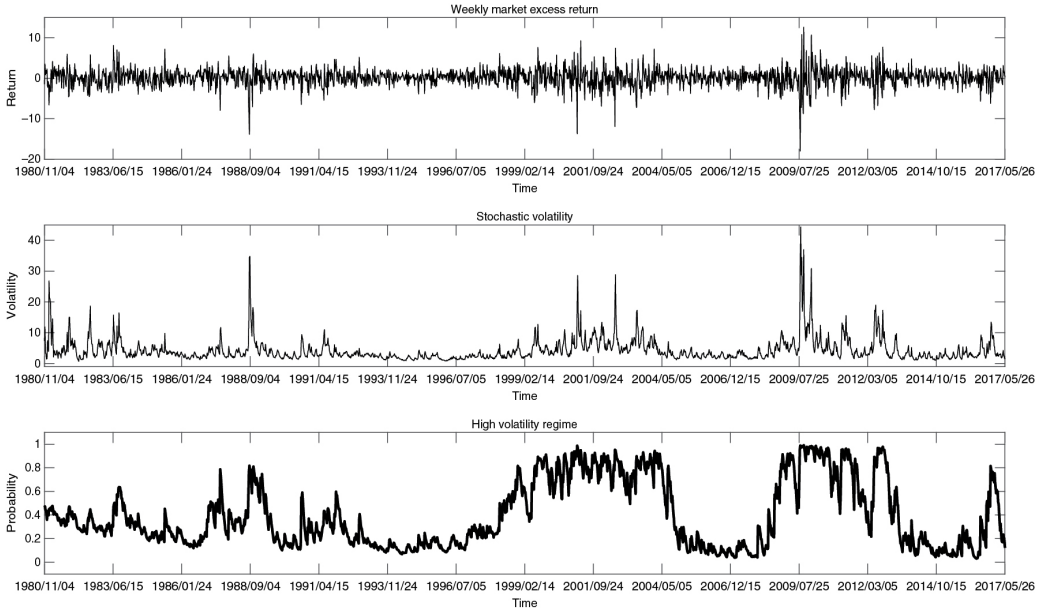

Figure 1: Time series of stock returns, volatility, and regime probability.
Notes: The sample period ranges from the week of 1980/01/04 and 2017/05/26. Stochastic volatility refers to the extracted $exp(x_t)$. High volatility regime depicts the probability of the estimated switches in regimes.

Continuing from the working example of Model B, we compare the 1 year average excess return and the estimated μ_t^r in Figure 2. The red dotted line is the average excess return, whereas the blue solid line is the extracted μ_t^r from the data. For a better comparison of the movements in the average excess return and the estimated μ_t^r, we report the result for the latter sample period from the week of August 12, 2011 to the week of May 26, 2017. Given that the sample period starts from 1980, we conjecture that there is sufficient information to accurately estimate μ_t^r by August of 2011. Using a slightly earlier or slightly later date than August 2011 does not substantially change our finding. Figure 2 shows that there is a similarity in the long-term movements between the two series. Even though the magnitude of fluctuation may be different, the overall vertical movements of these two series co-align with each other.[22] The overall comparison result indicates that the estimated μ_t^r can accurately capture the behaviors of the average excess return.

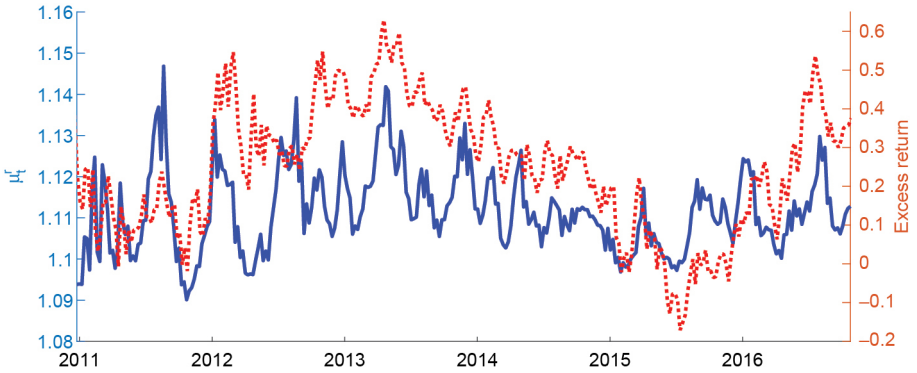

Figure 2: Comparison of yearly average excess return and filtered μ_t^r.
Notes: The solid blue line is for the estimated μ_t^r, while the red dashed line is the for the yearly average excess return. The graph depicts the periods between 2011/08/12 and 2017/05/26.

In addition, we report the posterior means and standard deviations of model parameters at the terminal time period in Table 2. While the parameters that represent persistence, ϕ_{11} and ϕ_{22}, are significant, other parameters that govern inter-temporal relationships between the conditional mean and volatility, ϕ_{12} and ϕ_{21}, are not significant. Even if the trade-off between risk and return is not our direct concern, the empirical result may be the consequence of the intrinsically inconclusive evidence on the risk-return relation in the return data, which has been actively discussed in the previous literature. For instance, Ghysels, Santa-Clara, and Valkanov (2005), Lundblad (2007), and Ludvigson and Ng (2007) show a positive risk-return trade-off, whereas Nelson (1991), Glosten, Jagannathan, and Runkle (1993), and Brandt and Kang (2004) present a negative risk-return trade-off.

5 Conclusion

In this paper, we propose a novel online estimation method for regime-switching state-space models. Particularly, we improve the estimation performance of the existing method developed by Carvalho and Lopes (2007) by utilizing all of the available and most recent information when re-sampling existing particles and generating new particles. Empirically, we apply the proposed method to understand the dynamics of the conditional mean and volatility of US excess stock market returns under a regime switching framework. Our empirical study indicates that the model that incorporates regime-switching (particularly in the volatility process) is the most viable candidate for correctly capturing the stock return movement.

Our new approach opens doors for future research. A combination of a full parameter learning approach (e.g. sufficient statistics) in Carvalho et al. (2010) with our proposed approach of handling the regime index variable may offer a more reliable estimation method for various regime switching models. Additionally, it would be interesting to merge more refined SMC smoothing methods of Bernardo et al. (2011), Rios and Lopes (2013), and Yang, Stroud, and Huerta (2017) with our proposed algorithm.

Acknowledgement

We are grateful to two anonymous referees and the editor for helpful suggestions and constructive comments that greatly improved the paper. All errors are our own.

Appendix A

Volatility MSE and QPS comparison

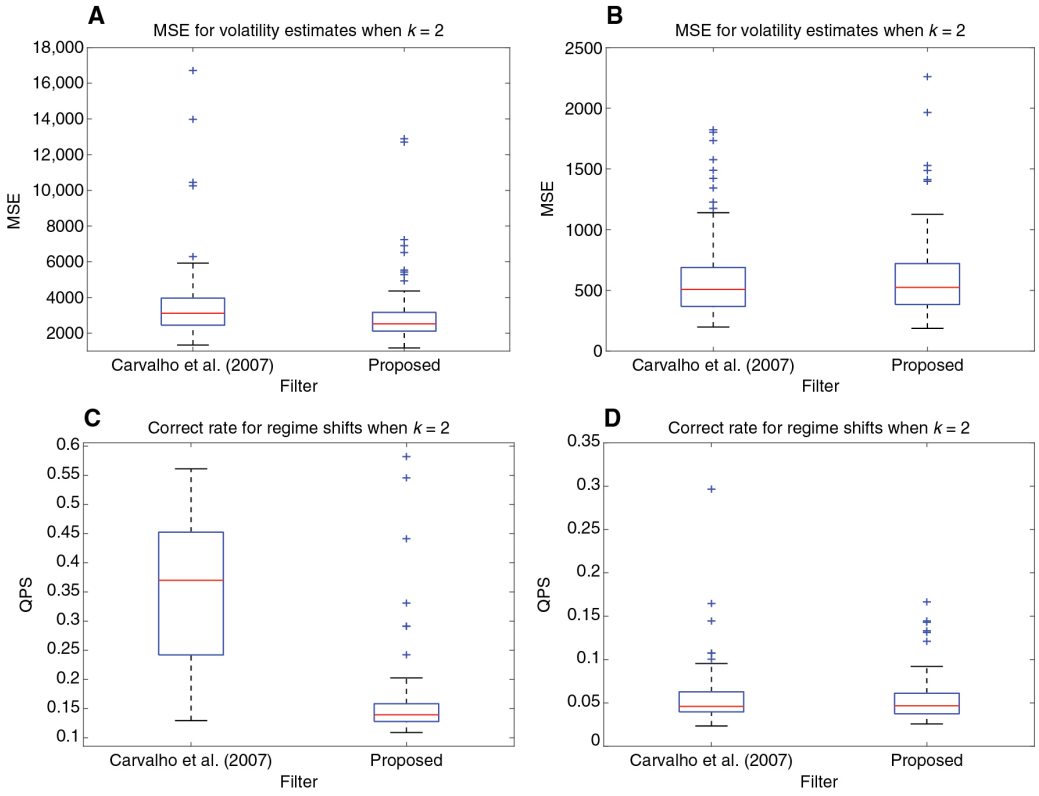

Figure 3: Top: box plots of the MSE of the estimated volatility process compared to the real simulated volatility process for each filter. Bottom: box plots of the QPS for each filter. (A) Case 1: $\pi_{00} = 0.9, \pi_{11} = 0.85$. (B) Case 2: $\pi_{00} = 0.99, \pi_{11} = 0.9$. (C) Case 1: $\pi_{00} = 0.9, \pi_{11} = 0.85$. (D) Case 2: $\pi_{00} = 0.99, \pi_{11} = 0.9$.
Notes: All plots present the results for the MSSV $k = 2$ model with the parameter set of $\{\alpha_0 = 1, \alpha_1 = 3, \phi = 0.5, \sigma^2 = 0.5\}$. Case 1 uses the transition probability of $\{\pi_{00} = 0.9, \pi_{11} = 0.85\}$ and Case 2 uses $\{\pi_{00} = 0.99, \pi_{11} = 0.9\}$. The total simulation runs are 100. Each simulation uses 5000 particles in 1000 time periods.

Appendix B

Comparison of parameter MSE: case 1

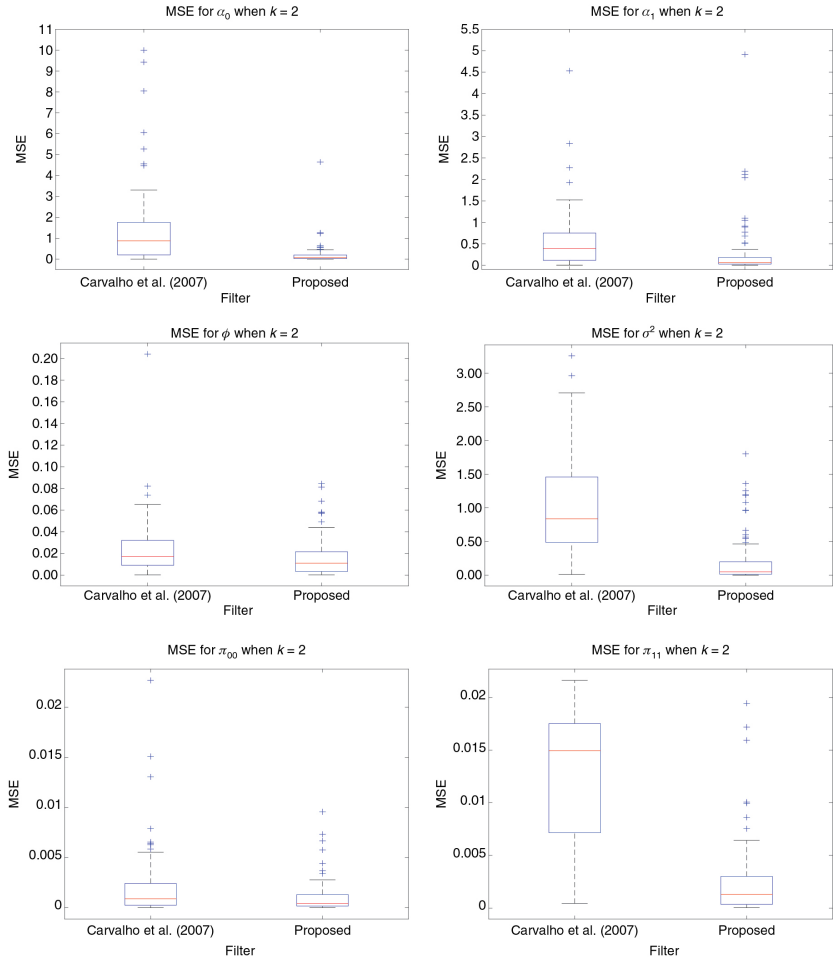

Figure 4: Box plots of the MSE of each parameter for each filter.
Notes: All plots present the results for the MSSV $k = 2$ model with the parameter set of $\{\alpha_0 = 1, \alpha_d = 3, \phi = 0.5, \sigma^2 = 0.5\}$. The total simulation runs are 100. Each simulation uses 5000 particles in 1000 time periods.

Appendix C

Comparison of parameter MSE: case 2

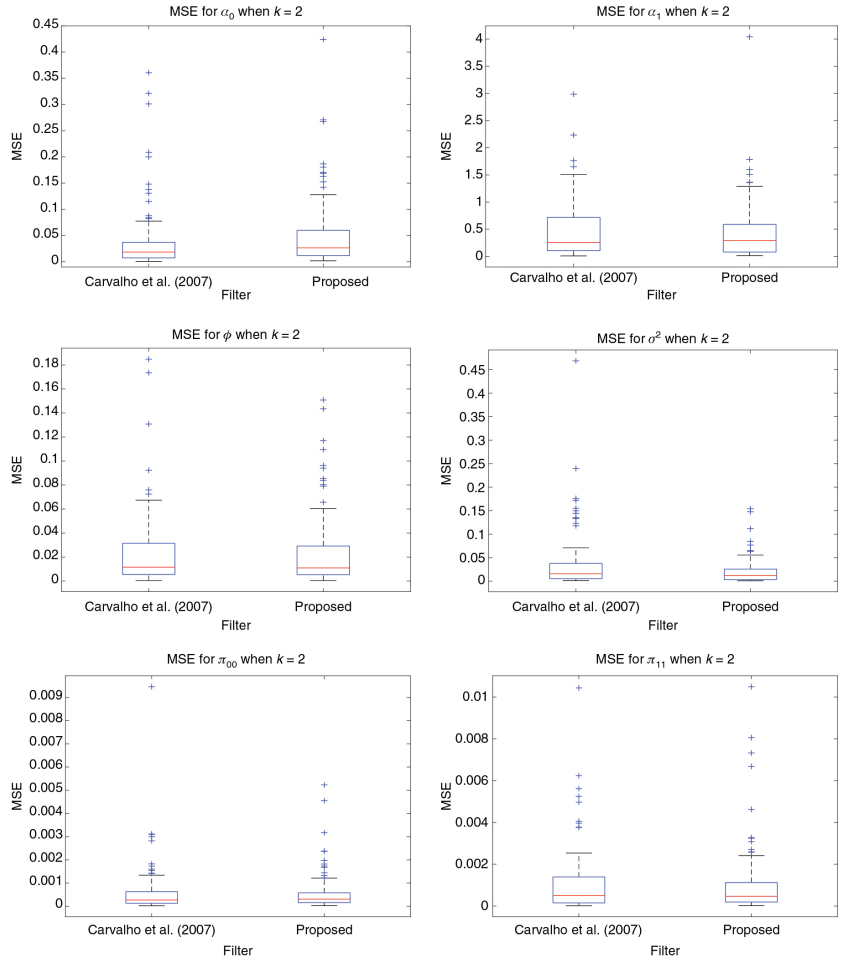

Figure 5: Box plots of the MSE of each parameter for each filter.
Notes: All plots present the results for the MSSV $k = 2$ model with the parameter set of $\{\alpha_0 = 1, \alpha_d = 3, \phi = 0.5, \sigma^2 = 0.5\}$. The total simulation runs are 100. Each simulation uses 5000 particles in 1000 time periods.

Appendix D

Table 2 Robust Check

Table 3: Parameter comparisons by model specification, robust check.

		Model A		**Model B**		**Model C**		**Model D**	
Regime Δ in Vol.		No	No	Yes	Yes	Yes	Yes	Yes	Yes
Regime Δ in Mean		No	No	No	No	No	No	Yes	Yes
Regime Δ in Leverage		No	No	No	No	Yes	Yes	No	No
Prior		**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**	**Mean**	**S.D.**
μ_0^m	$\mathcal{N}(ln\hat{M} - 1, 2^2)$	−2.314	0.376	−1.990	0.282	−2.204	0.372	−3.891	0.357
μ_d^m	$\mathcal{TN}(2, 2^2)$	–	–	–	–	–	–	2.053	0.360
μ_0^x	$\mathcal{N}(ln\hat{V} - 1, 2^2)$	1.269	0.082	0.812	0.068	0.917	0.083	0.828	0.085
μ_d^x	$\mathcal{TN}(2, 2^2)$	–	–	1.026	0.118	1.017	0.124	1.155	0.121

ϕ_{11}^m	$\mathcal{N}(0.9, 0.5^2)$	0.528	0.131	0.324	0.152	0.861	0.032	0.861	0.030
ϕ_{12}^m	$\mathcal{N}(0, 1^2)$	−0.218	0.090	−0.046	0.120	0.002	0.077	0.011	0.104
ϕ_{21}^x	$\mathcal{N}(0, 1^2)$	−0.205	0.073	0.032	0.114	−0.063	0.067	−0.020	0.087
ϕ_{22}^x	$\mathcal{N}(0.9, 0.5^2)$	0.830	0.052	0.843	0.040	0.881	0.024	0.824	0.049
$\phi_{23,0}^x$	$\mathcal{N}(0, 0.1^2)$	−0.198	0.020	−0.231	0.023	−0.182	0.030	−0.262	0.033
$\phi_{23,1}^x$	$\mathcal{N}(0, 0.1^2)$	–	–	–	–	−0.256	0.042	–	–
ρ	$\mathcal{TN}(0, 1^2)$	−0.026	0.217	−0.104	0.236	0.020	0.259	−0.050	0.252
σ_m^2	$\mathcal{IG}(5, 0.05)$	0.042	0.010	0.044	0.012	0.013	0.004	0.016	0.005
σ_x^2	$\mathcal{IG}(5, 0.05)$	0.068	0.013	0.065	0.017	0.013	0.003	0.012	0.003
π_{00}^m	$\mathcal{B}e(99, 1)$	–	–	–	–	–	–	0.975	0.006
π_{11}^m	$\mathcal{B}e(99, 1)$	–	–	–	–	–	–	0.975	0.007
π_{00}^x	$\mathcal{B}e(99, 1)$	–	–	0.992	0.002	0.992	0.002	0.991	0.003
π_{11}^x	$\mathcal{B}e(99, 1)$	–	–	0.989	0.004	0.984	0.005	0.987	0.004
log(ML)		−4105.9		−4099.6		−4105.2		−4105.0	

Refer to footnotes in Table 2.

Notes

1 Gordon, Salmond, and Smith (1993) introduce the bootstrap filter to draw samples from unobserved states based on a sampling importance re-sampling strategy. Building off the bootstrap filter, Pitt and Shephard (1999) construct the APF, which adopts a sequential importance sampling with re-sampling particle filters.

2 Lopes and Tsay (2011) provide an excellent review of the recent development in the SMC literature, including Carvalho and Lopes (2007) who provide an application to the Markov switching stochastic volatility (MSSV) models. Rios and Lopes (2013) revisit and extend the work of Carvalho and Lopes (2007) and implement the filters to estimate MSSV models. See also Carvalho et al. (2010) and Bernardo et al. (2011) for recent developments in particle learning methods.

3 To illustrate our point, consider two cases with two different regimes, regime 1 and regime 2. For Case 1, $\pi_{11} = 0.51$ and $\pi_{12} = 0.49$. For Case 2, $\pi_{11} = 0.99$ and $\pi_{12} = 0.01$. The parameter π_{11} (π_{12}) represents the transition probability of a hidden state moving from regime 1 to regime 1 (from regime 1 to regime 2). While the transition probabilities for π_{11} are both higher than 0.5, there is a substantial difference in the magnitude between π_{11} in Case 1 versus π_{11} in Case 2. The existing algorithm could not distinguish the difference in calibrating the regime index variable, whereas our algorithm considers the different degrees of transition probabilities.

4 Our proposed algorithms in this section can be modified to incorporate time-varying transition probabilities. For instance, consider a two-state environment with transition probabilities $\pi_{11,t}$ and $\pi_{22,t}$ and assume that they follow a functional form of a logit specification. Then $0 < \pi_{11,t} = exp(Z_t\gamma_1)/(1 + exp(Z_t\gamma_1)) < 1$ and $0 < \pi_{22,t} = exp(Z_t\gamma_2)/(1 + exp(Z_t\gamma_2)) < 1$, where the values of the covariates Z_t vary over time.

5 We have added the regime-switching state-space model to the original work of Liu and West (2001).

6 The incremental target density of the proposed algorithm is given below:

$$p(x_t, s_t, \theta \mid y_{1:t}) \propto p(y_t \mid x_t, s_t; \theta)p(x_t \mid x_{t-1}, s_t; \theta)p(s_t \mid s_{t-1}; \theta)p(\theta|y_{1:t-1}),$$

where $y_{1:t} = [y_1, y_2, \dots, y_t]'$.

7 The expectation of regime duration is computed by $\frac{1}{1-\pi_{kk}}$ for $k = \{0, 1\}$. See Kim and Nelson (1999) for the derivation of regime duration. For the simulation purpose, when the computed regime duration is not an integer, we round it up to the nearest integer.

8 MSE is a widely used and standard measurement for the validity of online filtering methods. Many others, including Carvalho and Lopes (2007), Carvalho et al. (2010), and Rios and Lopes (2013) utilize MSE of volatility and model parameters to compare the validity of different methods.

9 The parameter set is $\theta = \{\alpha_0, \alpha_d, \phi_0, \sigma^2, \pi_{00}, \pi_{11}\}$.

10 The priors for α_0, α_d, and ϕ are normal distributions. The variance parameter takes an inverse gamma distribution as a prior. For the transition probabilities, we assume beta distributions.

11 The average MSE and QPS for each algorithm can also be compared with a box plot representation. Appendix A presents the distribution of observed MSE_V and QPS values obtained from 100 simulations.

12 We show box plots for the distribution of MSE for each parameter in Appendix B.

13 So, Lam, and Li (1998), for instance, highlight the appropriateness of a time-varying stochastic volatility model with regime-switching using the Standard and Poor's 500 weekly return data.

14 The data is retrieved from the CRSP database.

15 We also consider a case in which $\phi_{13}\epsilon_t$ is included in equation (8). However, the posterior distribution of ϕ_{13} is narrowly disperse around zero and the more general model is not preferred by our Bayesian model selection criterion. For the sake of brevity, we do not report the corresponding result.

16 We have considered another comparable model in which all VAR coefficients, $\{\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}\}$ change according to regimes. We do not report the results here because the marginal likelihood of the model is too low compared to others.

17 Christie (1982) provides a comprehensive analysis of the importance of the leverage effect.

18 We impose weakly informative priors for other parameters while using informative priors for regime transition probabilities following the literature.

19 The marginal likelihood estimator of the auxiliary particle filter is defined by:

$$\hat{p}(y_{1:t}) = \hat{p}(y_1) \prod_{t=2}^{T} \hat{p}(y_t \mid y_{1:t-1}),$$

where $\hat{p}(y_t \mid y_{1:t-1}) = (\frac{1}{N} \sum_{i=0}^{N} \omega_t^{(i)})(\sum_{k=0}^{K-1} \sum_{i=1}^{N} \omega_{t-1|t,k}^{(i)})$.
20 Our main finding in Table 2 is based on the prior distribution $\mathcal{B}e(98, 2)$ for the transition probabilities. The expected regime duration within the middle 90% of the assumed prior beta distribution ranges from 21.3 to 277.7 periods. To explore a wider range of the regime duration in the prior distribution, we carry out a robust check that utilizes $\mathcal{B}e(99, 1)$. The middle 90% range expands to 33.3 and 1923.1 periods, which includes cases with much more persistent regime changes. Appendix D presents the results from the sensitivity analysis. All of our findings are robust to the alternative priors.
21 The difference between the log(ML) value for the two compared models is multiplied by two to compute the model comparison criterion. When the value is between 6 and 10, the corresponding model is strongly preferred. When it is larger than 10, the model is very strongly preferred.
22 We also experiment with the monthly excessive return instead of the yearly, and the result does not change.

References

Ang, A., and A. Timmermann. 2012. "Regime Changes and Financial Markets." *Annual Review of Financial Economics* 4: 313–337.

Beltratti, A., and C. Morana. 2006. "Breaks and Persistency: Macroeconomic Causes of Stock Market Volatility." *Journal of Econometrics* 131: 151–177.

Bernardo, J., M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, and M. West. 2011. "Particle Learning for Sequential Bayesian Computation." *Bayesian Statistics 9* 9: 317.

Bollen, N. P., S. F. Gray, and R. E. Whaley. 2000. "Regime Switching in Foreign Exchange Rates:: Evidence from Currency Option Prices." *Journal of Econometrics* 94: 239–276.

Brandt, M. W., and Q. Kang. 2004. "On the Relationship between the Conditional Mean and Volatility of Stock Returns: A Latent var Approach." *Journal of Financial Economics* 72: 217–257.

Carvalho, C. M., and H. F. Lopes. 2007. "Simulation-Based Sequential Analysis of Markov Switching Stochastic Volatility Models." *Computational Statistics & Data Analysis* 51: 4526–4542.

Carvalho, C., M. S. Johannes, H. F. Lopes, and N. Polson. 2010. "Particle Learning and Smoothing." *Statistical Science* 25: 88–106.

Chib, S. 1998. "Estimation and Comparison of Multiple Change-Point Models." *Journal of Econometrics* 86: 221–241.

Christie, A. A. 1982. "The Stochastic Behavior of Common Stock Variances: Value, Leverage and Interest Rate Effects." *Journal of Financial Economics* 10: 407–432.

Ghysels, E., P. Santa-Clara, and R. Valkanov. 2005. "There is a Risk-Return Trade-off after All." *Journal of Financial Economics* 76: 509–548.

Glosten, L. R., R. Jagannathan, and D. E. Runkle. 1993. "On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks." *The Journal of Finance* 48: 1779–1801.

Gordon, N. J., D. J. Salmond, and A. F. Smith. 1993. "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation." in *IEE Proceedings F-Radar and Signal Processing* volume 140, IET, 140: 107–113.

Kass, R. E., and A. E. Raftery. 1995. "Bayes Factors." *Journal of the American Statistical Association* 90: 773–795.

Kim, C.-J. 1994. "Dynamic Linear Models with Markov-Switching." *Journal of Econometrics* 60: 1–22.

Kim, C.-J., and C. R. Nelson. 1999. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*, Edn. 1, Vol. 1. Cambridge, MA: MIT Press Books.

Liu, J., and M. West. 2001. "Combined Parameter and State Estimation in Simulation-Based Filtering." In *Sequential Monte Carlo Methods in Practice*, edited by Arnaud Doucet, Nando de Freitas, and Neil Gordon. 197–223. New York, NY: Springer.

Lopes, H. F., and R. S. Tsay. 2011. "Particle Filters and Bayesian Inference in Financial Econometrics." *Journal of Forecasting* 30: 168–209.

Ludvigson, S. C., and S. Ng. 2007. "The Empirical Risk–Return Relation: A Factor Analysis Approach." *Journal of Financial Economics* 83: 171–222.

Lundblad, C. 2007. "The Risk Return Tradeoff in the Long Run: 1836–2003." *Journal of Financial Economics* 85: 123–150.

Marcucci, J. 2005. "Forecasting Stock Market Volatility with Regime-Switching Garch Models." *Studies in Nonlinear Dynamics & Econometrics* 9: 1–55.

Morana, C., and A. Beltratti. 2002. "The Effects of the Introduction of the Euro on the Volatility of European Stock Markets." *Journal of Banking & Finance* 26: 2047–2064.

Nelson, D. B. 1991. "Conditional Heteroskedasticity in Asset Returns: A New Approach." *Econometrica: Journal of the Econometric Society* 59: 347–370.

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

Raftery, A. E. 1995. "Bayesian Model Selection in Social Research." *Sociological Methodology* 25: 111–163.

Rios, M. P., and H. F. Lopes. 2013. "The Extended Liu and West Filter: Parameter Learning in Markov Switching Stochastic Volatility Models." In *State-Space Models*, 23–61. Springer.

So, M. E. P., K. Lam, and W. K. Li. 1998. "A Stochastic Volatility Model with Markov Switching." *Journal of Business & Economic Statistics* 16: 244–253.

Yang, B., J. R. Stroud, and G. Huerta. 2017. "Sequential Monte Carlo Smoothing with Parameter Estimation." *Bayesian Analysis* 13: 1133–1157.

Supplementary Material: The online version of this article offers supplementary material (DOI: https://doi.org/10.1515/snde-2018-0016).