Feedback in Regime Formation

Naoya Nagasaka*

October 31, 2025
Click here for the latest version.

Abstract

This paper proposes regime-switching state space models with feedback from lagged continuous state variables to regime formation. Regime transition probabilities implied from such a regime rule can be incorporated into the Kalman filter with regime-switching coefficients. It is shown that the truncation step introduced in the filter to circumvent the path dependence problem has an asymptotically negligible impact on the resulting log likelihood. Consistency of the maximized likelihood estimator can be established as well. Two simulation exercises confirm the finite sample performance of the filter. I then study the monetary-fiscal policy mix using the regime-switching DSGE model with the proposed regime determination rule to achieve a better forecasting performance, especially around the time when a regime change is likely.

JEL Classification Code: C32, C34, E63
Keywords: Regime-switching models, State space models, DSGE models, Time-variant transition probabilities, Monetary/fiscal policy mix

---

*Department of Economics, Indiana University (Email: naonagas@iu.edu). I am grateful to Yoosoon Chang, Christian Matthes, and Joon Park for their continued support. I would also like to thank Chaojun Li, Shi Qiu, Felipe Schwartzman, and the participants of Indiana Macro Brown Bag, the Hoosier Economic Conference, the Midwest Econometrics Group Meeting, SETA, and RISE Workshop for their insightful comments. This paper won the SETA Best Young Scholar Award. I thank Ken Weakley for the contribution to the Award.

Surging inflation and rising demands on governments have brought a pivotal moment for economic policy. ("Regime change; The world economy", The Economist, October 8 2022)

1. Introduction

Since the seminal work by Hamilton (1989), regime-switching models are widely used in macroeconomic and financial applications as a convenient way to capture discrete changes in the economic environment. In particular, the regime-switching structure is combined with state space models to study unobserved factors and regimes jointly. This framework enables us to estimate the regime-switching dynamic stochastic general equilibrium (DSGE) models, which are used to investigate the monetary/fiscal policy mix, the financial friction, and the economic uncertainty to name a few.

One of the important drawbacks of the traditional regime-switching model is in time-invariant regime transition probabilities. In other words, probabilities of moving from one regime to another are assumed to be fixed constants. However, a regime shift itself might be an endogenous event influenced by other economic indicators. For example, central banks employ a lot of economists to monitor economic conditions and policymakers decide their policy stance based on reports by those economists. If this is the case, ignoring such a source of information by assuming constant transition probabilities would lead econometricians to misspecify the model.

The primary objective of this paper is to propose a way to model regime-switching allowing for feedback from economic conditions to regime determination. I present two asymptotic properties of this class of models, preciseness of approximated likelihood and consistency of the maximum likelihood estimator. The simulation exercises verify that the proposed model works well in finite samples. Finally, I apply this model to the DSGE model with the monetary/fiscal policy mix.

My model builds on the usual linear-Gaussian state space model with regime-switching coefficients. Instead of the typical time-invariant Markovian assumption on the regime transition, I specify the threshold-type regime rule which produces time-varying transition probabilities. More specifically, the regime rule consists of the constant term governing the tendency of being in one regime compared to another at the steady state, the linear combination of the lagged continuous state variables, and the random variable independent of fundamentals. The second component represents the feedback channel from economic conditions to regime formation, while I interpret the third component as the forces driving regime shifts for reasons which are not modeled explicitly, such as the political environment and "sentiments".

The likelihood of these regime-switching state space models can be computed by the extended Kalman filter. The algorithm is based on the usual Kalman filter with regime-switching coefficients except for one additional step: computing the transition probabilities for each period. I show that the transition probabilities conditional on the history of observations are functions of updated mean and variance of the continuous state variables. Under the assumption of the Gaussian distributed error in the regime determination rule, the transition probabilities can be written using the bivariate normal cumulative distribution function. Due to this property, we can evaluate the likelihood fast enough to estimate the model in a reasonable amount of time.

It is well known that we need to introduce the approximation into the regime-switching Kalman filter. This is because the exact likelihood evaluation is subject to the path-dependence problem: We need to keep track of the entire history of regimes which grows exponentially with the sample length. The common trick employed in the literature, which is adopted in my filter as well, is to truncate the history of the regimes to follow. We keep track of the most recent r periods instead of the entire history and introduce the collapsing step to integrate out the regime $r+1$ period ago. The first econometric result confirms that this approximation works well asymptotically. More precisely, I show that the absolute difference between the approximated and exact log-likelihood converges in probability to zero if the truncation order r grows with the sample length. The speed of convergence is exponential with r. Intuitively, since the contribution of more distant past information becomes smaller, we may safely ignore the regime realization older than the threshold r. We will see that the order of convergence depends on the VAR(1) coefficient in the transition equation in the state space model. As the system becomes more persistent, we have to keep track of the longer history of regimes to ensure convergence.

The second set of results discusses the consistency of the exact maximum likelihood estimator. I establish this claim for regime transition probabilities more general than the ones proposed in this paper. As long as transition probabilities have the property similar to the irreducibility in the usual Markov process context in addition to being continuous with respect to the parameters, the maximum likelihood estimator is shown to be consistent. These two requirements are satisfied by not only the transition probabilities implied by my regime rule but also those introduced by other preceding works.

To evaluate the finite sample performance, two Monte Carlo simulation exercises are conducted. The first simulation design is the model with a scalar unobserved state variable and a scalar observation. Since the exact likelihood cannot be calculated computationally, I use the relatively large truncation order, $r = 10$, as the baseline. Comparing the log-likelihood from this baseline and from smaller r, we see the difference between the two is reasonably small even with $r = 1$ and shrinks as we increase r. Consistent with the theoretical investigation, the less persistent system delivers a smaller difference. The computational burden grows exponentially as r gets larger, and the filter with $r = 10$ needs a long enough time to prohibit us from using this case in the estimation. The second laboratory is the state space representation derived from the regime-switching small-scale DSGE model. The purpose of this exercise is to investigate the role of truncation in a more realistic model. I confirm that the likelihood with a small r gives us a good approximation in this case as well.

As an empirical illustration, the proposed regime determination rule is applied to the DSGE model with the monetary/fiscal policy mix by Bianchi and Ilut (2017). There are two possible policy regimes: the active monetary/ passive fiscal (AM/PF) policies and the passive monetary/ active fiscal (PM/AF) policies. They differ in the mechanism of how inflation is determined and how the transversality condition is satisfied. Several key variables related to the monetary and fiscal policies—output, potential output, inflation, nominal interest rate, and debt-to-output ratio—are allowed to affect the policy regime determination. The model is solved allowing the agents to take into account the possibility of regime shifts, and the resulting state space representation is fed into the regime-switching Kalman filter developed above.

Estimating the parameters using the sequential Monte Carlo method, I find the regime probabilities and impulse response functions similar to Bianchi and Ilut (2017). I find feedback from some macroeconomc variables to regime determination. For instance, a higher debt-to-GDP ratio and higher inflation make the AM/PF regime more likely to happen. This observation is interpretable from the optimal policy perspective: An increase in the debt level under the PM/AF regime raises the inflation rate at impact. Since aggregate welfare is a decreasing function of the deviation of the inflation rate from its target, the (consolidated) government has the incentive to take the AM/PF regime in order to stabilize inflation. Also, tax rate and interest rate are associated with the AM/PF regime, which is also consistent with the observation that the AM/PF regime is associated with contractionary monetary and fiscal policies.

Unlike the traditional exogenous regime switching model where regime transition probabilities are kept constant over time, this framework features endogenous and time-varying transition probabilities. This feature is useful especially in the context of forecasting. Using the same DSGE model, I show that we can predict regime changes by keeping track of transition probability. Also, by taking time-varying probabilities of regime changes into account, the baseline model offers better forecasts of GDP growth and inflation than the exogenous switching model, at the time when a regime change is likely to happen.

This paper contributes to two strands of the literature: the asymptotic properties of regime-switching models in general and the structural macroeconomic models incorporating endogenous regime-switching framework. Douc et al. (2004) provide the asymptotic properties for hidden Markov models with discrete state variables, whose results are extended by Kasahara and Shimotsu (2019). The generalization of their claims to the case with time-varying transition probabilities is done by Li and Liu (2023) and Pouzo et al. (2022). Since all of these papers do not consider continuous state variables, their framework cannot be applied to state space models. As an extension to the models with continuous state variables, Douc and Moulines (2012) show the consistency of the maximum likelihood estimator with the time-invariant transition kernel. Using the framework by Douc and Moulines (2012), Li (2023) examines the linear-Gaussian state space model with time-invariant regime transition probabilities. However, none of these works consider the models with non-discrete state variables whose transition kernel is time-varying.

There are a couple of works investigating regime-switching state space models in particular, especially focusing on the collapsing step in the Kalman filter. The regime truncation is introduced by Kim (1994) to avoid the computational issue related to path dependence. Kim and Kang (2019) examine the preciseness of the Kim filter by conducting simulation exercises. They compare the likelihood computed from the Kim filter with the one from the particle filter, the latter of which is expected to be closer to the exact likelihood, and numerically show that those two produce similar outcomes. Theoretically, Li (2023) proves the asymptotic negligibility of the difference between the exact and truncated log-likelihood. This paper can be regarded as a generalization of those two papers to the models with endogenous regime-switching.

There is an accumulation of literature employing regime-switching DSGE models to study discrete changes in the economic environment, including Liu et al. (2011), Bianchi (2012), Nimark (2014), Bianchi and Melosi (2017), Bianchi et al. (2018), and Aruoba et al. (2018)[1]. Recognizing the limitation of the regime-switching framework with time-invariant transition probabilities, several works deviate from this assumption

[1]See Hamilton (2016) for a survey of the literature.

and formulate endogenous regime shifts. Ascari et al. (2022) study the state space model in which the equation representing the long-run Phillips curve has a kink at the threshold of the trend inflation rate. Chang et al. (2021) incorporate the threshold-type regime-switching model proposed by Chang et al. (2017) into the small-scale New Keynesian model to examine the monetary policy stance in the postwar U.S. Section 2 provides a further discussion on the relationship between these two papers and mine. Another related work is Benigno et al. (2020) who assume that the transition probabilities are logistic functions of the endogenous variables in the model and use this model to analyze the Mexican business cycle and financial market.

This paper is organized as follows. In Section 2, I introduce the econometric framework which incorporates the feedback from lagged continuous state variables to regime formation. Sections 3 and 4 examine the asymptotic properties of the proposed model: The former establishes the asymptotic equivalence between exact and truncated likelihood and the latter shows the consistency of the exact maximum likelihood estimator. Two simulation exercises are conducted in Section 5. Section 6 discusses the empirical application using the DSGE model focusing on the monetary/fiscal policy mix. Section 7 concludes.

Notation. We denote the history of a variable z_t by $z_{t_1}^{t_2} = (z_{t_1}, z_{t_1+1}, \cdots, z_{t_2})$ where $t_1 \leq t_2$. For any matrix A, let $\|A\|$ denote the operator norm of A. For any symmetric matrices A and B with the same size, we write $A \geq B$ ($A > B$) if $A - B$ is positive semidefinite (definite). The function $\phi(x; \mu, \Sigma)$ is the normal probability density function (PDF) with mean μ and variance Σ evaluated at x, and $\Phi(x; \mu, \Sigma)$ is the corresponding cumulative distribution function (CDF). We simply write $\phi(x)$ and $\Phi(x)$ to denote the PDF and CDF of the standard normal. When we want to be explicit about the size of $x \in \mathbb{R}^k$, we write $\phi_k(x)$ and $\Phi_k(x)$. For any real-valued function f, we denote $f^+(x) = \max\{f(x), 0\}$ and $f^-(x) = \max\{-f(x), 0\}$. For a transition kernel L on a measurable space $(\mathbb{X}, \mathcal{X})$, we denote

$$Lf(x) = \delta_x Lf = \int L(x, dx')f(x'), \quad \mu L(A) = \mu L\mathbf{1}_A = \int \mu(dx)L(x, A)$$

where $f : \mathbb{X} \to \mathbb{R}$ is bounded and μ is a measure on $(\mathbb{X}, \mathcal{X})$. We can combine two transition kernels L_1 and L_2 on $(\mathbb{X}, \mathcal{X})$ to construct a new transition kernel L_1L_2 such that

$$L_1L_2(x, A) = \int L_1(x, dx')L_2(x', A)$$

2. Baseline Specification

This section elaborates on the model specification considered in this paper. I will also introduce the filtering algorithm to derive the likelihood.

2.1. Regime-Switching State Space Model

Consider the linear-Gaussian state space model with regime-switching parameters consisting of the observed variable $y_t \in \mathbb{R}^{d_y}$ and the unobserved state variable $x_t \in \mathbb{R}^{d_x}$.

$$
\begin{aligned}
x_t &= A_{s_t} x_{t-1} + Q_{s_t} \varepsilon_t \\
y_t &= B_{s_t} x_t + R_{s_t} u_t
\end{aligned}
\tag{1}
$$

where $u_t \in \mathbb{R}^{d_u}$ and $\varepsilon_t \in \mathbb{R}^{d_\varepsilon}$ are independent and follow the standard Gaussian. The first equation is the transition equation specifying the dynamics of the latent x_t using the VAR(1) model. The second equation is the observation equation which relates the observed y_t with the unobserved x_t. The coefficients $(A_{s_t}, B_{s_t}, Q_{s_t}, R_{s_t})$ depend on the discrete latent regime $s_t \in \mathcal{S} \equiv \{0, \cdots, S-1\}$ whose process will be described shortly. Although the model (1) does not include constant terms, it is fairly straightforward to incorporate them. As is well known, the linear-Gaussian state space model is the most commonly used empirical framework to estimate dynamic stochastic general equilibrium (DSGE) models. The regime-switching structure allows us to model changes in the coefficients across time, such as shifts of monetary policy stance (Dovish vs Hawkish) and economic volatilities.

In the traditional regime-switching model such as Hamilton (1989), the regime (s_t) is assumed to follow a Markov process with time-invariant transition probabilities. This specification might be restrictive because the regime shifts might occur in response to changes in economic circumstances. For example, in the context of monetary policy, it is more natural to believe that the central bank settles on the monetary policy stance based on various economic conditions such as the inflation rate, unemployment rate, and the stability of the financial market.

Given this consideration, I specify the regime (s_t) as a function of the continuous state variable (x_t). More specifically, $s_t \in \{0, 1\}$ is determined by the following threshold type rule[2].

$$s_t = \mathbf{1}\left\{\tau + \lambda' x_{t-1} + \eta_t \geq 0\right\} \tag{2}$$

where $\tau \in \mathbb{R}$, $\lambda \in \mathbb{R}^{d_x}$, and (η_t) is i.i.d. and follows a distribution with a distribution function F. In the baseline specification, I will assume F to be the standard normal to get the simple expression for the regime transition probabilities. Meanwhile, other choices for F, such as the logistic function, are also possible. Also, η_t can follow the AR(1) process so that we allow the persistency of regimes irrelevant to the economic model.

The regime specification by equation (2) allows a linear combination of lagged continuous state variables $\lambda' x_{t-1}$ to influence the regime shifts. This part captures the feedback from economic conditions to the regime. Going back to the monetary policy example, the policy stance is influenced by information about various economic indicators. The constant term τ governs the inclination for being in regime 1. The specification (2) also includes the random component η_t independent of lags and leads of (x_t, y_t). One way to interpret this term is the regime shifter unmodeled in the state space model, such as political forces. Alternatively, this term might capture sentiments, which do not directly drive the business cycle but have an implication for the economic outcome by causing regime shifts[3].

We may interpret this regime rule from the perspective of discrete choice models popular in microeconomic applications. Suppose there is a decision maker who determines which regimes to take. In the context of macroeconomic policy analysis, he/she can be a policy authority. In other applications absent of such an actual decision maker (e.g., time-varying volatility), we may regard him/her as "nature". This decision maker chooses a regime giving a higher payoff. In this context, the left hand side of the inequality in (2) can be interpreted as a net utility from choosing regime 1 over regime 0. Especially, this framework fits with optimal policy analyses in which a policy authority determines the policy variables based on observed state variables[4].

[2] Regime is assumed to be binary throughout this paper, i.e., $S = 2$, while one can extend the framework to allow more than two regimes.

[3] In the literature, sentiment is characterized as a shock or variable which affects agents' decisions through their information sets without affecting fundamentals. Among various ways to include sentiments in structural models, some works in the literature exploit sentiments as equilibrium selection devices. Conceptually, this idea is close to the regime-switching state space models.

[4] There is a difference in the setting from the Ramsey style optimal policy analysis. The most prototypical optimal policy problem, such as Galí (2015), assumes that contemporaneous structural shocks are observed. To the contrary, the ε_t is not observed for the decisionmaker in this context.

2.2. Filtering Algorithm

In order to estimate the parameters, we need to derive the likelihood implied by equations (1). I introduce the modified Kalman filter allowing for the regime rule given by equation (2). Here, I describe only the key steps in the algorithm. A detailed description is given in Appendix A.

The filter iterates the forecasting and updating steps for each $t = 1, \cdots, T$. Each step is similar to the Kalman filter with regime-switching coefficients proposed by Kim (1994). The algorithm by Kim assumes that the regime transition probabilities are time-invariant and are given as parameters. As our transition probabilities depend on the lagged state variables x_{t-1}, our model needs an extra step to calculate the filtered transition probabilities for each period.

As in the Kim filter, we need to introduce the collapsing step at the end of each iteration. Otherwise, the likelihood evaluation is subject to path dependence: the likelihood of y_t depends on the entire history of the regime up to period t, which grows with t exponentially. In the collapsing step, we truncate the number of periods whose regime we keep track of. More specifically, I assume that the algorithm tracks only the most recent $r(T)$ periods. We compress the $S^{r(T)+1}$ moments into $S^{r(T)}$ expressions to avoid path dependence. Section 3 proves that this truncation works well: the likelihood derived from the truncated algorithm is asymptotically equivalent to the exactly evaluated likelihood as we increase $r(T)$ with T. For brevity, I simply write r to express $r(T)$ unless I need to stress the dependence of the truncation number on the sample size.

Appendix A derives the transition probability from s_{t-r+1}^{t-1} to s_t given $\mathcal{F}_{t-1} = \sigma(y_1^{t-1})$, the σ-field generated by the history of observations. It is shown that we have to consider the lag-augmented state space model:

$$\begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix} = \begin{bmatrix} A_{s_t} & O \\ I & O \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \end{bmatrix} + \begin{bmatrix} Q_{s_t} \\ O \end{bmatrix} \varepsilon_t$$

$$y_t = \begin{bmatrix} B_{s_t} & O \end{bmatrix} \begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix} + R_{s_t} u_t$$

We re-define $x_t = [x_t', x_{t-1}']'$ and so forth. Assuming $\eta_t \sim N(0,1)$, the transition probability from $s_{t-1} = 0$ to s_t is approximated as

$$
\begin{aligned}
&p_r(s_t = 0 | s_{t-1} = 0, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}) \\
\approx &\frac{\Phi\left(-\tau\iota_2; \Lambda\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2}), I + \Lambda\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\Lambda'\right)}{\Phi\left(-\tau; \lambda'\left(\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x)}, 1 + \lambda'\left(\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x, d_x+1:2d_x)}\lambda\right)}
\end{aligned}
$$

where ι_n ($n \in \mathbb{N}$) is a $n \times 1$ vector whose elements are all unity, $\Lambda = \begin{bmatrix} \lambda' & 0 \\ 0 & \lambda' \end{bmatrix}$ is a $2 \times 2d_x$ matrix, $\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2})$ and $\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})$ are conditional mean and variance of the state vector in the lag-augmented state space model given s_{t-r+1}^{t-2} and $\mathcal{F}_{t-1}$. In the denominator, we extract $(d_x + 1)$-th to $2d_x$-th elements from $\bar{x}_{t-1|t-1}(\cdot)$ and the corresponding rows and columns from $\bar{\Omega}_{t-1|t-1}(\cdot)$. Appendix A shows the transition probabilities for other combinations of (s_{t-1}, s_t) as well. Due to the trancation explained above, those two expressions approximate $x_{t-1|t-1}(s_1^{t-2})$ and $\Omega_{t-1|t-1}(s_1^{t-2})$, the objects from the exact filter. The bars in the conditional mean and variance are placed to be explicit about the approximation.

Augmenting these approximated transition probabilities in the forecasting step enables us to derive the likelihood of the model, which is the ingredient of maximum likelihood estimation as well as Bayesian inference.

2.3. Comparison with Related Regime-Switching Rules

Before concluding this section, it might be instructive to compare the proposed method with the other endogenous regime-switching frameworks.

To examine the long-run Phillips curve, the relationship between trend inflation and trend output, Ascari et al. (2022) incorporate a particular type of endogenous regime-switching. Their long-run Phillips curve has a kink depending on trend inflation: The slope of the long-run Phillips curve changes when the trend inflation crosses a certain threshold. More specifically, their regime indicator can be expressed as $s_t = \mathbf{1}\{\bar{\pi}_t \geq \tau\}$. This specification has similarities with equation (2) except that they are incorporating the feedback from the current state variable to the regime and they shut down the stochastic component η_t. Since it is straightforward to incorporate the current x_t into (2), the empirical model by Ascari et al. (2022) can be regarded as a special case of my framework. In terms of the estimation strategy, while their filter is based on particle filtering, this paper develops the algorithm based on the Kalman filter[5]. Although I do not compare these two, the latter is more straightforward and expected to be faster.

Extending Chang et al. (2017) (CCP henceforth) approach of modeling endogenous regime-switching, Chang et al. (2021) consider the regime specification in which the unobserved regime $s_t \in \{0, 1\}$ is determined by the latent regime factor w_t.

$$\begin{aligned} s_t &= \mathbf{1}\{w_t \geq -\tau\} \\ w_t &= \alpha w_{t-1} + \rho' \varepsilon_{t-1} + \nu_t \sqrt{1 - \rho' \rho}, \quad \nu_t \sim N(0, 1) \end{aligned} \tag{3}$$

where $\rho \in \mathbb{R}^{d_\varepsilon}$ satisfies $\rho' \rho < 1$ and ν_t is independent of (u_t) and (ε_t). The regime factor is dependent on the past structural shock ε_{t-1}, which specifies the feedback from economic conditions to regime determinations. By sequential iteration, the CCP regime factor w_t can be written as

$$w_t = \alpha^t w_0 + \sum_{j=1}^{t} \alpha^{j-1} \rho' \varepsilon_{t-j} + \sqrt{1 - \rho' \rho} \times \sum_{j=1}^{t} \alpha^{j-1} \nu_{t+1-j}$$

On the other hand, we can express the terms inside the indicator function in equation (2) as

$$\lambda' x_{t-1} + \eta_t = \lambda' \left(\prod_{j=1}^{t-1} A_{s_{t-j}} \right) x_0 + \lambda' \sum_{j=1}^{t-1} \left(\prod_{k=1}^{j} A_{s_{t-j}} \right) Q_{s_{t-j}} \varepsilon_{t-j} + \eta_t$$

Both of these two equations involve the summation of initial conditions, the path of structural shocks (ε_t), and stochastic terms. Although the initial structural shock ε_0 matters for the former but does not for the latter, its contribution is expected to be negligible under $|\alpha| < 1$. Hence, my framework can be regarded as an unweighted version of CCP-type regime-switching.

It is straightforward to see that equation (2) is identical to equation (3) when $\lambda = \rho = 0$ and η_t follows an AR(1) process[6]. As shown by CCP, this case reduces to the traditional Hamilton (1989) filter. In other words, when we shut down the endogenous feedback in the regime determination, my and CCP's frameworks are nothing but regime-switching models with time-invariant transition probabilities.

Although which of the regime specifications is suitable would be an application-dependent question, an advantage of my approach is the robustness to non-invertibility.

[5]Since Ascari et al. (2022) deal with nonstationary variables, the asymptotic properties discussed later do not hold for their model.

[6]It is straightforward to allow η_t to follow an AR(1) process, which is discussed in Appendix A.

When the model is non-invertible, agents cannot recover current structural shocks from current and past observations. Under such a situation, it might be more natural to think that regime shifts happen based on lagged observation rather than structural shocks.

3. Asymptotic Equivalence between Exact and Truncated Likelihood Functions

This section establishes the asymptotic equivalence between the likelihood inferred from the Kalman filter presented in the previous section and the exact likelihood. We fix the collection of parameters θ, which lies in a compact parameter space Θ. Let

$$\psi(s_{t_1}, \cdots, s_{t_2}) = \begin{cases} A_{s_{t_1}} \cdots A_{s_{t_2}} & \text{if } t_1 > t_2 \\ I & \text{if } t_1 = t_2 \\ A_{s_{t_1}}^{-1} \cdots A_{s_{t_2}}^{-1} & \text{if } t_1 < t_2 \end{cases}$$

The following two assumptions are standard in the control theory and required to establish the stability of the filter.

ASSUMPTION 1 (Uniform Complete Observability). *There exists $N \in \mathbb{N}$ and $\beta_{UCO} \geq \alpha_{UCO} > 0$ such that for any $s_0^N \in \{0,1\}^{N+1}$,*

$$0 < \alpha_{UCO} I \leq \sum_{t=0}^{N} \psi(s_{t+1}, \cdots, s_N)' B_{s_t}' (R_{s_t} R_{s_t}')^{-1} B_{s_t} \psi(s_{t+1}, \cdots, s_N) \leq \beta_{UCO} I$$

ASSUMPTION 2 (Uniform Complete Controllability). *There exists $N \in \mathbb{N}$ and $\beta_{UCC} \geq \alpha_{UCC} > 0$ such that for any $s_1^N \in \{0,1\}^N$,*

$$0 < \alpha_{UCC} I \leq \sum_{t=0}^{N-1} \psi(s_N, \cdots, s_{t+2}) Q_{s_{t+1}} Q_{s_{t+1}}' \psi(s_N, \cdots, s_{t+2})' \leq \beta_{UCC} I$$

Intuitively, the observability implies that we can recover the initial continuous state variable given the observed variables (y_t). The controllability guarantee that any desirable states is attained by manipulating error terms (ε_t).

To see how these two assumptions work in our framework, we consider the Kalman filter with the different initial variance-covariance matrices Ω^1 and Ω^2 but with the same path of regime realizations $(s_1, \cdots, s_t)$. Denote $\Omega^1_{t|t}(s_1^t)$ and $\Omega^2_{t|t}(s_1^t)$ to be the conditional variance associated with Ω^1 and Ω^2 respectively. We let $\Psi(s_1^t) = (I - K(s_1^t)B_{s_t})A_{s_t}$ where $K(s_1^t) = \Omega_{t-1|t-1}(s_1^{t-1})B'_{s_t}(B_{s_t}\Omega_{t-1|t-1}(s_1^{t-1})B'_{s_t} + R_{s_t}R'_{s_t})$ is the Kalman gain given s_1^t. For $k \in \mathbb{N}$, we let $\Psi^k(s_1^t) = \Psi(s_1^t)\Psi(s_1^{t-1})\cdots\Psi(s_1^{t-k+1})$ and define $\Psi^0(s_1^t) = I$. Then, we can write

$$\begin{aligned}
\Omega^1_{t|t}(s_1^t) - \Omega_{t|t}(s_1^t) &= \Psi(s_1^t)\left(\Omega^1_{t-1|t-1}(s_1^{t-1}) - \Omega^2_{t-1|t-1}(s_1^{t-1})\right)\Psi(s_1^{t-1})' \\
&= \cdots \\
&= \Psi^{t-1}(s_1^t)\left(\Omega^1 - \Omega^2\right)\Psi^{t-1}(s_1^{t-1})'
\end{aligned}$$

Therefore, we should examine the property of $\Psi^k(\cdot)$ to see whether the filter is convergent. Jazwinski (1970) shows the exponential convergence of $\Psi^k(\cdot)$ with respect to k under the uniformly complete observability and controllability.

LEMMA 1 (Theorem 7.4 in Jazwinski (1970)). *Assume Assumptions 1 and 2. Then, the filter is uniformly asymptotically stable. In other words, there exist positive constants c_1, c_2 such that*

$$\max_{s_1^t}\left\|\Psi^k(s_1^t)\right\| \le c_1 \exp(-c_2 k)$$

The stability of the Kalman filter established in this lemma plays a crucial role in our proof.

Set the initial state variables arbitrarily: $x_1 = \tilde{x}$ and $s_1 = \tilde{s}$. Although these two initial variables are fixed at this moment, we can allow x_1 and s_1 to follow known distributions with a minor modification of the proof. Let $p_{r(T),\theta}(y_1^T|x_1 = \tilde{x}, s_1 = \tilde{s})$ be the likelihood with the truncated regime with parameters θ and $p_\theta(y_1^T|x_1 = \tilde{x}, s_1 = \tilde{s})$ be the exact likelihood. The following proposition shows that the difference between the two is asymptotically negligible.

PROPOSITION 1. *Assume Assumptions 1 and 2 and let F be a standard normal distribution. We have*

$$\left|\log p_{r(T),\theta}(y_1^T|x_1 = \tilde{x}, s_1 = \tilde{s}) - \log p_\theta(y_1^T|x_1 = \tilde{x}, s_1 = \tilde{s})\right| = O_p(\exp(c_2 r))$$

for any $\theta \in \Theta$, where c_2 is given in Lemma 1.

Intuitively, the contribution of past information becomes smaller as time goes by because of the convergent property of the Kalman filter. Due to Lemma 1, we may safely ignore the regime realizations in the distant past. This convergence is attained with exponential speed.

Here I provide the sketch of the proof. The details can be found in Appendix B. We simply denote $r = r(T)$ henceforth. We can write the truncated likelihood function as

$$p_{r,\theta}(y_1^T|x_1 = \tilde{x}, s_1 = \tilde{s}) = p(y_1|\tilde{x}, \tilde{s})$$
$$\times \sum_{s_2^T} \left\{ \prod_{t=r+1}^{T} \left[p_r(y_t|s_{t-r+1}^t, \mathcal{F}_{t-1}) p_r(s_t|s_{t-r+1}^{t-1}, \mathcal{F}_{t-1}) \right] \prod_{t=2}^{r} \left[p(y_t|s_1^t, \mathcal{F}_{t-1}) p(s_t|s_1^{t-1}, \mathcal{F}_{t-1}) \right] \right\}$$

On the other hand, the exact likelihood function is given by

$$p_\theta(y_1^T|x_1 = \tilde{x}, s_1 = \tilde{s}) = p(y_1|\tilde{x}, \tilde{s}) \times \sum_{s_2^T} \prod_{t=2}^{T} \left[p(y_t|s_1^t, \mathcal{F}_{t-1}) p(s_t|s_1^{t-1}, \mathcal{F}_{t-1}) \right]$$

Then we can write

$$\begin{aligned} p_{r,\theta}(y_1^T|x_1 = \tilde{x}, s_1 = \tilde{s}) \leq & p_\theta(y_1^T|x_1 = \tilde{x}, s_1 = \tilde{s}) \\ & \times \prod_{t=r+1}^{T} \exp\left(\max_{s_2^t} \left| \log p_r(y_t|s_{t-r+1}^t, \mathcal{F}_{t-1}) - \log p(y_t|s_1^t, \mathcal{F}_{t-1}) \right| \right) \\ & \times \prod_{t=r+1}^{T} \exp\left(\max_{s_2^t} \left| \log p_r(s_t|s_{t-r+1}^{t-1}, \mathcal{F}_{t-1}) - \log p(s_t|s_1^{t-1}, \mathcal{F}_{t-1}) \right| \right) \end{aligned} \tag{4}$$

After taking the log to both hand sides, the difference between the log two likelihood functions $\log p_{r,\theta}(y_1^T|x_1 = \tilde{x}, s_1 = \tilde{s})$ and $\log p_\theta(y_1^T|x_1 = \tilde{x}, s_1 = \tilde{s})$ can be attributed to the difference in the log likelihood of y_t given the regime realization (the second line) and the difference of the regime transition probabilities (the third line). Since both of them are functions of the conditional mean and variance implied from the filter, we may compare the differences of the mean and variance from the updated and exact filters.

Let $\bar{x}_{t|t}(s_{t-r+2}^t)$ and $x_{t|t}(s_1^t)$ be the updated mean of x_t from the truncated and exact filters, respectively. Likewise, let $\bar{\Omega}_{t|t}(s_{t-r+2}^t)$ and $\Omega_{t|t}(s_1^t)$ be the updated variance of x_t. Define $\Delta_{r,t}^{\Omega} = \bar{\Omega}_{t|t}(s_{t-r+2}^t) - \Omega_{t|t}(s_1^t)$ and $\Delta_{r,t}^{x} = \bar{x}_{t|t}(s_{t-r+2}^t) - x_{t|t}(s_1^t)$. Utilizing Lemma 1, Proposition A1 in the appendix shows that there exist positive constants $c_{\Omega,\Delta}$ and c_2 (identical to c_2 in Lemma 1), as well as a positive and stochastically bounded random variable $M_{x,t}$ such that

$$\max_{s_1^t} \|\Delta_{r,t}^{\Omega}\| \leq c_{\Omega,\Delta} \exp(-2c_2(r-1))$$
$$\max_{s_1^t} \|\Delta_{r,t}^{x}\| \leq M_{x,t} \exp(-c_2(r-1))$$

Utilizing these inequalities, it can be shown the boundedness of the second and third lines of (4), establishing our claim.

REMARK 1. *Li (2023) shows the identical result when (i) $d_x = d_y = d_\varepsilon = d_u = 1$ and (ii) (s_t) follows a usual Markov chain with constant transition probabilities. Proposition 1 generalizes her proposition in these two respects.*

REMARK 2. *Suppose again $d_x = d_y = d_\varepsilon = d_u = 1$. One can show that the convergence rate of the difference between two likelihood functions is $O_p\left(A_{max}^{r(T)-1}\right)$ where $A_{max} = \max_s |A_s|$. Thus, the speed of convergence depends on the maximum persistence of the continuous state variable x_t across regimes. This observation is intuitive because the past information would be preserved more as we increase the persistence. Loss of information caused by the regime truncation is expected to be large when the system is more persistent. In the general vector case, the convergence rate depends on the constant c_2 which is hard to describe analytically.*

Note that this proof can be generalized to other forms of regime transition probabilities. As explained in Appendix B, to show the asymptotic negligibility of exact and approximated transition probabilities, we interpret those probabilities as functions of conditional mean and variance of the state vector x_t. Under differentiability with respect to these mean and variance, we apply the mean value theorem and show the stochastic boundedness of the gradient. Even if we formulate the transition probabilities differently, we can show the asymptotic negligibility as long as we can establish that the gradient is stochastically bounded.

4. Consistency of Maximum Likelihood Estimator

This section shows the consistency of the exact maximum likelihood estimator. I prove the consistency for a class of regime rules more general than (2). Before discussing the main proposition of this section, I introduce the framework with examples.

4.1. Specification

Consider state space models with regime-switching coefficients introduced by (1). The transition probability of regimes is specified as $q^\theta(s_{t+1}|s_t, x_t, y_t)$ whose time-varying property is expressed as the dependence on the current observation y_t and the continuous state variable x_t. This specification includes the regime rule (2). I provide two other examples whose regime transition probabilities are written in this form.

EXAMPLE 1. *Chang et al. (2021) incorporate the endogenous regime-switching structure à la CCP, equations (3), into state-space models. Consider the augmented state space system so that we can characterize the transition probability consistently with the notation introduced above.*

$$\begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix} = \begin{bmatrix} \underbrace{A_{s_t}}_{d_x \times d_x} & \underbrace{O}_{d_x \times d_x} \\ \underbrace{I}_{d_x \times d_x} & \underbrace{O}_{d_x \times d_x} \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \end{bmatrix} + \begin{bmatrix} \underbrace{Q_{s_t}}_{d_x \times d_\varepsilon} \\ \underbrace{O}_{d_x \times d_\varepsilon} \end{bmatrix} \varepsilon_t$$

$$y_t = \begin{bmatrix} \underbrace{B_{s_t}}_{d_y \times d_x} & \underbrace{O}_{d_y \times d_x} \end{bmatrix} \begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix} + R_{s_t} u_t$$

The transition probability can be written as

$$q^\theta(s_{t+1}|s_t, [x_t', x_{t-1}']', y_t) = (1 - s_{t+1})\omega_\rho + s_{t+1}(1 - \omega_\rho)$$

where

$$\omega_\rho = \frac{\left[(1 - s_t) \int_{-\infty}^{-\tau\sqrt{1-\alpha^2}} + s_t \int_{-\tau\sqrt{1-\alpha^2}}^{\infty}\right] \Phi\left(\frac{-\tau - \rho' Q_{s_t}^{-1}(x_t - A_{s_t} x_{t-1})}{\sqrt{1-\rho'\rho}} - \frac{\alpha z}{\sqrt{1-\alpha^2}\sqrt{1-\rho'\rho}}\right) \phi(z) dz}{(1 - s_t)\Phi\left(-\tau\sqrt{1-\alpha^2}\right) + s_t\left(1 - \Phi\left(-\tau\sqrt{1-\alpha^2}\right)\right)}$$

EXAMPLE 2. *In the context of regime-switching models without continuous state variables, Diebold et al. (1994) specify the transition matrix of $s_t \in \{0, 1\}$ as*

$$\begin{bmatrix} p_{00}(x_t) & p_{01}(x_t) \\ p_{10}(x_t) & p_{11}(x_t) \end{bmatrix}, \quad p_{ij}(x_t) = 1 - p_{ii}(x_t), \quad for \quad i \neq j$$

where $x_t \in \mathbb{R}^k$ is observed and $p_{00}, p_{11} : \mathbb{R}^k \to [0, 1]$ are the weighting functions, typically modeled as the logistic function.

If we instead consider state-space models, we may model the transition probability as

$q^\theta(s_{t+1} = j|s_t = i, x_t, y_t) = p_{ij}(x_t)$. The difference here is that the transition probability is depending on the continuous state variables x_t. The application to the monetary policy analysis can be found in Davig and Leeper (2006) where the coefficient in the Taylor rule changes discretely based on whether the inflation rate is below or above a certain threshold.

The same discussion holds when the transition depends on the observed variable instead of the latent state variable: we may simply replace x_t with y_t. For example, Auerbach and Gorodnichenko (2012) investigate state-contingent effects of fiscal stimulus using the structural autoregression, a special case of state-space models. To specify whether the economy is in a boom or recession, they use a logistic function that translates the moving average of the output growth rate into a weight between zero and one.

Let $\xi_t = (x_t', s_t)'$ characterize the unobserved state variables whose support is given by $\mathbf{X} = \mathbb{R}^{d_x} \times \mathcal{S}$. We define

$$\tilde{q}^\theta(\xi_{t+1}|\xi_t, y_t) = \phi\left(x_{t+1}; A_{s_{t+1}}x_t, Q_{s_{t+1}}Q'_{s_{t+1}}\right) q^\theta(s_{t+1}|s_t, x_t, y_t)$$

and $Q^\theta(\xi_{t+1}|\xi_t, y_t)$ to be the probability measure associated with $\tilde{q}^\theta(\xi_{t+1}|\xi_t, y_t)$, which is the transition kernel of ξ_t given y_t. Let $\mathcal{X} \equiv \mathcal{B}(\mathbb{R}^{d_x}) \times \sigma(\mathcal{S})$ where $\mathcal{B}(\mathbb{R}^{d_x})$ is the Borel σ-field generated by $\mathbb{R}^{d_x}$ and $\sigma(X)$ is a σ-field generated by a set X. Given a probability measure χ on $(\mathbf{X}, \mathcal{X})$ which can be interpreted as the initial distribution of state variables ξ_1, we define the likelihood function of $y_m^n = (y_m', \cdots, y_n')'$ as

$$\begin{aligned} p_\chi^\theta(y_m^n) = \int \cdots \int \chi(d\xi_m)\phi\left(y_m; B_{s_m}x_m, R_{s_m}R'_{s_m}\right) \\ \times \prod_{p=m+1}^{n} Q^\theta(d\xi_p|\xi_{p-1}, y_{p-1})\phi\left(y_p; B_{s_p}x_p, R_{s_p}R'_{s_p}\right) \end{aligned} \tag{5}$$

Then, the maximum likelihood estimator of θ is defined as[7]

$$\hat{\theta}_{\chi,T} \equiv \arg\max_{\theta\in\Theta} p_\chi^\theta(Y_1^T)$$

4.2. Consistency

We need the stationarity of (y_t) and (x_t).

ASSUMPTION 3. *(y_t) and (x_t) are stationary processes.*

[7] In this section, the variables denoted by uppercase letters are random variables, and the corresponding lowercase letters are realizations.

Here I list additional assumptions to establish consistency.

ASSUMPTION 4. *There exists a positive integer k such that for any $y_1^k \in (\mathbb{R}^{d_y})^k$ and $x_1^k \in (\mathbb{R}^{d_x})^k$, we have*

$$\inf_{\theta \in \Theta} \inf_{s_1, s_{k+1}} q^\theta(s_{k+1}|s_1, x_1^k, y_1^k) > 0$$

where $q^\theta(s_{k+1}|s_1, x_1^k, y_1^k) = \sum_{s_1^k} \prod_{i=0}^{k-1} q^\theta(s_{i+1}|s_i, x_i, y_i)$.

This assumption implies that any regime can be reached with a positive probability after k periods for any $\theta \in \Theta$, in parallel with the irreducibility in usual Markov processes. This is essential for the asymptotic analysis because, otherwise, there might be a regime that cannot be reached forever. In fact, this assumption holds not only for my specification (2) but also for the examples given in this section with $k = 1$.

We also require continuity of the transition probability q^θ with respect to θ.

ASSUMPTION 5. *The function $\theta \mapsto q^\theta(s_{t+1}|s_t, x_t, y_t)$ is continuous on Θ for any (s_{t+1}, s_t, x_t, y_t).*

We define the unnormalized transition kernel of (ξ_t) given the observations y_1^t.

$$\mathbf{L}^\theta \langle y_1^t \rangle (\xi_1, A) = \int \cdots \int \left[\prod_{i=1}^{t} \phi\left(y_i; A_{s_i} x_i, R_{s_i} R_{s_i}'\right) Q^\theta(d\xi_{i+1}|\xi_i, y_i) \right] \mathbf{1}_A(\xi_{t+1})$$

Note that $\mathbf{L}^\theta \langle y_1^t \rangle (\xi_1, A) = p^\theta_{\delta_{\xi_1}}(y_1^t)$. Under some assumptions on the initial distribution, Proposition 2 shows the forgetting of the initial distribution χ in (i), the convergence of likelihood function in (ii), and finally the consistency in (iii).

PROPOSITION 2. *Assume Assumptions 3–5. For a compact set $D \in \mathcal{X}$ and $k \in \mathbb{Z}$, take $\mathcal{M}(D, k)$ the family of probability measures in $(X, \mathcal{X})$ such that*

$$\mathcal{M}(D,k) = \left\{ \chi : \mathbb{E}\left[\log^- \inf_{\theta \in \Theta} \chi \mathbf{L}^\theta \langle Y_1^u \rangle \mathbf{1}_D \right] < \infty, \text{ for any } u = 1, \cdots, k \right\}$$

(i) For any $\theta \in \Theta$, there exists a measurable function $\pi_Y^\theta : (\mathbb{R}^{d_y})^{\mathbb{Z}^-} \to \mathbb{R}$, which does not depend on χ, such that for any $\chi \in \mathcal{M}(D, k)$,

$$\mathbb{P}\left[\lim_{m \to \infty} \frac{p_\chi^\theta(Y_{-m}^0)}{p_\chi^\theta(Y_{-m}^{-1})} = \pi_Y^\theta\left(Y_{-\infty}^0\right) \right] = 1$$

and moreover,

$$\mathbb{E}\left[\left| \log \pi_Y^\theta(Y_{-\infty}^0) \right| \right] < \infty$$

(ii) *For any $\theta \in \Theta$ and $\chi \in \mathcal{M}(D,k)$,*

$$\lim_{T\to\infty} T^{-1} \log p_{\chi}^{\theta}(Y_1^T) = \ell(\theta)$$

almost surely where $\ell(\theta) \equiv \mathbb{E}\left[\log \pi_Y^{\theta}\left(Y_{-\infty}^0\right)\right]$.

(iii) *Define $\Theta^* \equiv \arg\max_{\theta\in\Theta} \log p_{\chi}^{\theta}\left(Y_1^T\right) \subset \Theta$. For any $\chi \in \mathcal{M}(D,k)$, we have*

$$\lim_{T\to\infty} d\left(\hat{\theta}_{\chi,T}, \Theta^*\right) = 0$$

almost surely.

The proof is provided in Appendix B. Since the framework by Douc and Moulines (2012) can be applied to our model, the Appendix is verifying the assumptions for their Theorem 2.

To get an idea about why the maximum likelihood estimator is consistent, it is instructive to overview the proof by Douc and Moulines (2012), which is composed of two steps. In the first step, they establish that the limit of the likelihood function exists and it does not depend on the initial distribution χ.

$$\lim_{T\to\infty} T^{-1} \log p_{\chi,T}^{\theta}(Y_1^T) = \lim_{T\to\infty} T^{-1}\mathbb{E}\left[\log p_{\chi}^{\theta}(Y_1^T)\right] = \ell(\theta)$$

Note that the log likelihood can be written as $\log p_{\chi}^{\theta}(Y_0^{n-1}) = \sum_{k=0}^{n-1} \log p_{\chi}^{\theta}(Y_k|Y_0^{k-1})$. One can show the existence of the limit of $\log p_{\chi}^{\theta}(Y_0|Y_{-m}^{-1})$, $\mathbb{P} - a.s.$, which is denoted by $\pi_Y^{\theta}(Y_{-\infty}^0)$. Since $\{Y_k\}_{k\in\mathbb{Z}}$ is assumed to be stationary ergodic, the log-likelihood converges to $\ell(\theta)$ by the Birkhoff ergodic theorem.

The proof is done by showing the second step: the sequence of the maximizers of log-likelihood $\{\arg\max_{\theta\in\Theta} T^{-1} \log p_{\chi}^{\theta}(Y_1^T)\}_{T\geq 1}$ converges to the maximizer of $\ell(\theta)$ almost surely.

REMARK 3. *Proposition 2 does not establish the identification of the estimator. It is known that even linear-Gaussian state space models without regime-switching coefficients cannot be globally identified[8]. To the best of my knowledge, the literature has not provided the identification of regime-switching state space models even with time-invariant transition probabilities.*

[8]For example, Hamilton (1994) discusses two simple models that are observationally equivalent.

TABLE 1. Simulation Design

	$s_t = 0$	$s_t = 1$
A_{s_t}	Varies	Varies
B_{s_t}	1	1
Q_{s_t}	1.0	2.0
R_{s_t}	1.0	1.5
λ	0.8	
τ	0.2	

TABLE 2. Log Likelihood from Univariate Model

	(1)	(2)	(3)	(4)
(A_0, A_1)	(0.7,0.95)	(0.5,0.95)	(0.5,0.7)	Time (ss)
$r = 1$	0.0063	0.0056	0.0049	0.070
$r = 2$	4.95e-04	2.67e-04	9.61e-05	0.130
$r = 3$	4.47e-05	1.96e-05	4.91e-06	0.248
$r = 4$	3.98e-06	1.44e-06	2.45e-07	0.481
$r = 10$	—	—	—	29.306

Note: The columns (1)-(3) report the median of the absolute difference between log-likelihood with $r = 10$ and the one with the corresponding r. The last column displays the median implementation time for $(A_0, A_1) = (0.7, 0.95)$.

5. Simulations

This section conducts two simulations to evaluate the finite-sample performance of the filter whose asymptotic properties are investigated in the previous sections. Especially, we are interested in how different assumptions on the truncation number r affect the log-likelihood as well as the parameter inference. The first exercise examines maximum likelihood estimates of parameters of a simple univariate model. The second simulation considers a larger model mimicking a small-scale New-Keynesian model with a regime-switching monetary policy rule to see the sensitivity of the log-likelihood to r.

5.1. Simulation 1: Univariate Model

The first simulation design considers a simple state space model with $d_x = d_y = d_u = d_\varepsilon = 1$. We generate the random sample $\{y_t\}_{t=1}^T$ based on the parameters given by Table 1. Persistency of x_t captured by A_{s_t} will be varied across simulation designs while keeping $A_0 < A_1$. The standard deviations Q_{s_t} and R_{s_t} are also larger in regime 1.

Hence, regime 1 can be interpreted as a more persistent and volatile one. We allow the feedback from x_{t-1} to the regime determination by setting nonzero λ. Regime 1 is more likely to happen since τ is positive. Time length is set as $T = 400$.

To begin with, we evaluate the difference in log-likelihood derived from small r plausible for empirical applications and the one from relatively large truncation order $r = 10$ over 500 simulations. Both of them are evaluated at the true parameters. Given that it is almost infeasible to calculate the exact log-likelihood (i.e., $r = 400$), our benchmark is set to be $r = 10$.[9]

Table 2 reports the median of absolute differences between two log-likelihood values across different (A_0, A_1). In all cases, the difference shrinks as we enlarge r and the rate of shrinkage seems to be exponentially fast. These are as expected given that the order of convergence was A_{max}^{r-1} as we saw in Section 3. Another implication from Table 2 is that the model exhibiting less persistency gives a smaller difference. Although the asymptotic convergence depends sorely on more persistent regime A_{max} theoretically as pointed out in Section 3, the less persistent regime also matters in the simulation as is clear by comparing columns (1) and (2).

To investigate the computational burden, the last column of Table 2 displays the median run time to calculate the likelihood. As we have to keep track of the 2^r history of regimes, the run time is almost doubled as we increase r by one. When $r = 10$, the filter spends almost half minute to derive the likelihood. It is unrealistic to estimate the model in such a setting since the estimation algorithm generally requires us to evaluate the likelihood at least thousands of times.

Next, we estimate the parameters $(A_0, A_1, Q_1, R_1, \tau, \lambda)$ with the maximum likelihood estimation for each combination of (A_0, A_1), whose results are displayed in Table 3.[10] Generally speaking, the parameters are estimated close to the truth and the intervals of the estimates are reasonable. We do not see substantial differences in parameter estimates across different r. Although we might expect that an increase in r leads to a more precise estimation as the likelihood function becomes more accurate, we do not see an improvement in the estimates for larger r.

TABLE 3. Maximum Likelihood Estimates

r	1	2	3	4	Truth
Panel A. $(A_0, A_1) = (0.7, 0.95)$					
A_0	0.6794 (0.50,0.78)	0.6794 (0.50,0.78)	0.6794 (0.50,0.78)	0.6794 (0.50,0.78)	0.7
A_1	0.9458 (0.88,0.97)	0.9458 (0.88,0.97)	0.9458 (0.88,0.97)	0.9458 (0.88,0.97)	0.95
Q_1	2.0384 (1.75,2.40)	2.0379 (1.75,2.40)	2.0379 (1.75,2.40)	2.0379 (1.75,2.40)	2.0
R_1	1.4737 (1.12,1.77)	1.4735 (1.12,1.77)	1.4735 (1.12,1.77)	1.4735 (1.12,1.77)	1.5
Panel B. $(A_0, A_1) = (0.5, 0.95)$					
A_0	0.4723 (0.25,0.62)	0.4727 (0.25,0.62)	0.4727 (0.25,0.62)	0.4727 (0.25,0.62)	0.5
A_1	0.9470 (0.89,0.97)	0.9470 (0.89,0.97)	0.9470 (0.89,0.97)	0.9470 (0.89,0.97)	0.95
Q_1	2.0349 (1.74,2.40)	2.0345 (1.74,2.40)	2.0345 (1.74,2.40)	2.0345 (1.74,2.40)	2.0
R_1	1.4791 (1.13,1.77)	1.4787 (1.13,1.77)	1.4788 (1.13,1.77)	1.4788 (1.13,1.77)	1.5
Panel C. $(A_0, A_1) = (0.5, 0.7)$					
A_0	0.4809 (0.29,0.61)	0.4809 (0.29,0.61)	0.4809 (0.29,0.61)	0.4809 (0.29,0.61)	0.5
A_1	0.6914 (0.52,0.82)	0.6917 (0.52,0.82)	0.6917 (0.52,0.82)	0.6917 (0.52,0.82)	0.7
Q_1	2.0468 (1.54,2.57)	2.0456 (1.54,2.57)	2.0455 (1.54,2.57)	2.0455 (1.54,2.57)	2.0
R_1	1.4539 (0.29,1.89)	1.4519 (0.31,1.89)	1.4519 (0.31,1.89)	1.4519 (0.31,1.89)	1.5

Note: This table reports the median of maximized likelihood estimates over 500 simulations along with 5 and 95% percentiles in parentheses.

TABLE 4. Log-Likelihood from Chang et al. (2021) Model

	Abs Diff of Loglik	Time (ss)
$r = 1$	0.1014	0.1248
$r = 2$	0.0111	0.2366
$r = 3$	0.0026	0.4615
$r = 4$	7.55e-04	0.9134
$r = 10$	—	59.1491

Note: The left column shows the median of the absolute difference in log-likelihood between $r = 1, 2, 3, 4$ and $r = 10$ over 500 replications. The right column shows the median implementation time.

5.2. Simulation 2: Log-Likelihood from DSGE

The purpose of the second simulation exercise is to examine whether the regime truncation matters in the more realistic model. Given that DSGE models are rarely estimated with the maximum likelihood method, we only provide the log-likelihood with different r. It is instructive to check whether the log-likelihood is precisely calculated with relatively small r as it is the most important ingredient of Bayesian inference.

I use the model by Chang et al. (2021) as a laboratory, which is a small-scale New-Keynesian model resembling An and Schorfheide (2007) with the regime-switching Taylor rule. The monetary policy regime is determined by the CCP-type threshold rule. Given that we use the proposed filter in the empirical application in the next section, it is instructive to see whether the filter works well in such a prototypical model.

I solve the Chang et al. (2021) model at the posterior mean to deduce the state space representation[11]. For each element in x_t, I draw λ randomly from the uniform distribution from -0.5 to 0.5 and fix these values across simulations. The constant term τ is assumed to be -0.75. The time length is $T = 200$, which is close to the sample length of most DSGE applications using the post-WWII quarterly data.

Table 4 lists the median of the absolute value of the difference between log-likelihood with $r = 1, 2, 3, 4$ and the one with $r = 10$ over 1,000 replications along with the median implementation time to run the filter. The absolute differences are enlarged

[9] Another possible strategy to approximate the exact log-likelihood is to use the filtering algorithm allowing for nonlinearity. Kim and Kang (2019) use the particle filter to examine the accuracy of Kim (1994) filter.

[10] The estimates for τ and λ are omitted since they are unstably estimated possibly due to issues in the numerical optimizer.

[11] This step is done with the RISE Toolbox. I describe this toolbox in the next section.

compared to the simple univariate application (Table 2), but they are still close to zero. We also see the shrinkage of the difference for larger r, consistent with both theory and the previous simulation result. In terms of the run time displayed in the right column, we can calculate the likelihood with a reasonable amount of time even for the relatively large state-space model studied here. However, the computational burden grows exponentially as r gets larger and the case with $r = 10$ ends up spending one minute to calculate the likelihood just once.

6. Empirical Application

This section provides the empirical application. I incorporate the regime rule (2) into the New-Keynesian model with the monetary-fiscal policy mix by Bianchi and Ilut (2017). After describing the overview of the model, I will discuss the prior distributions of the parameters as well as the estimation strategy. Then I provide the posterior distributions and show that the proposed model is useful in the forecasting context. I assume $r = 1$ throughout this section[12]. Additional results, including impulse response functions, can be found in Appendix C.

6.1. Model

We begin by introducing the structure of the model. Then we discuss the specification of regime switching.

6.1.1. Overview

This subsection discusses the basic framework of the structural model. The full description is given in Appendix C.

The model is based on Bianchi and Ilut (2017) which studies the monetary/fiscal policy mix in the postwar U.S. with a particular emphasis on the Great Inflation from the late 1970s and the early 1980s. The representative household consumes the final good and supplies labor for each period. In addition to the one-period government bond, he/she is allowed to hold the long-term securities provided by the government. There is a continuum of firms producing differentiated goods, which are combined to

[12] Most of existing papers on regime-switching DSGE models use the filter with $r = 1$ with the exception of Nimark (2014) who considers $r = 4$.

form the final consumption good. These firms are subject to monopolistic competition as well as quadratic price adjustment costs.

The government operates fiscal and monetary policies. The one-period government bond is assumed to have zero net supply. Total government expenditure is a combination of direct lump-sum transfer to the household and government spending. The total government expenditure and interest payment of the government bond will be financed by the lump-sum tax which follows the prespecified tax rule. The nominal interest rate is determined according to the Taylor rule depending on the current inflation rate and the output gap. The tax rule and Taylor rule are subject to regime-switching, which will be discussed later.

6.1.2. Regime-Switching

We introduce two discrete variables governing regimes: $s_t^{vol} \in \{0, 1\}$ related to volatility and $s_t^{pol} \in \{AM/PF, PF/AM\}$ related to policy rules. The volatility of the structural shocks changes over time depending on s_t^{vol}. Regime $s_t^{vol} = 1$ can be regarded as volatile times. Regime shifts for volatility occur following a time-invariant transition probability matrix P^{vol}.

$$P^{vol} = \begin{bmatrix} 1 - p_{1,2}^{vol} & p_{1,2}^{vol} \\ p_{2,1}^{vol} & 1 - p_{2,1}^{vol} \end{bmatrix}$$

The exogenous switching is assumed for s_t^{vol} to reduce the computational cost. In the meantime, this assumption might be reasonable if we regard the changes in economic volatility as events caused chiefly by reasons outside the model. For example, the estimation result by Bianchi and Ilut (2017) suggests that the oil crisis around the mid-1970s is classified as a high volatility regime and this episode is mainly driven by geopolitical forces. Also, large volatility around the Great Recession is a result of the instability in the financial market, while we do not model the financial friction explicitly.

Another regime $s_t^{pol} \in \{AM/PF, PM/AF\}$ is related to the monetary/fiscal policy mix pioneered by Leeper (1991)[13]. AM/PF stands for the combination of active monetary and passive fiscal policies, and PM/AF represents passive monetary and active fiscal policies. To illustrate the role of the policy regime, we look at the tax rule and the Taylor rule after linearization.

$$\tilde{\tau}_t = \rho_\tau(s_t^{pol})\tilde{\tau}_{t-1} + \left(1 - \rho_\tau(s_t^{pol})\right) \left[\delta_b(s_t^{pol})\tilde{b}_{t-1}^m + \delta_e \tilde{e}_t + \delta_y \left(\hat{y}_t - \hat{y}_t^*\right)\right] + \sigma_\tau(s_t^{vol})\varepsilon_t^\tau$$

[13]See Cochrane (2023) for a textbook treatment.

$$\tilde{R}_t = \rho_R(s_t^{pol})\tilde{R}_{t-1} + \left(1-\rho_R(s_t^{pol})\right)\left[\psi_\pi(s_t^{pol})\tilde{\pi}_t + \psi_y(s_t^{pol})\left(\hat{y}_t - \hat{y}_t^*\right)\right] + \sigma_R(s_t^{vol})\varepsilon_t^R$$

where we assume $\delta_b(AF) < \delta_b(PF)$ and $\psi_\pi(PM) < \psi_\pi(AM)$. The tax rate $\tilde{\tau}_t$ responds to the lagged debt-to-output ratio $\tilde{b}_{t-1}^m$, government expenditure $\tilde{e}_t$, and output gap $\hat{y}_t - \hat{y}_t^*$. The policy interest rate $\tilde{R}_t$ is influenced by inflation rate $\tilde{\pi}_t$ as well as output gap. On the one hand, the first policy regime, AM/PF, is the usual assumption in the New-Keynesian framework where fiscal policy is responsible for stabilizing the debt-to-output ratio $\tilde{b}_{t-1}^m$ by increasing the tax rate $\tilde{\tau}_t$ while monetary policy controls the nominal interest rate $\tilde{R}_t$ to make the inflation rate close to its target level. On the other hand, fiscal policy plays a role to determine the inflation rate in the PM/AF regime. A fiscal adjustment in response to an increase in the debt-to-output ratio is insufficient under this regime because the coefficient on the debt level $\delta_b(AF)$ is small. In order to satisfy the transversality condition, there must be inflation enough to inflate away the public debt. The central bank gives up stabilizing the price level and responds to changes in the inflation rate weakly. Note that our framework does not include the AM/AF regime considered in the original Bianchi and Ilut (2017) model. The periods classified as AM/AF by Bianchi and Ilut (2017) are very short compared to the other two regimes and our regime determination rule does not allow to have more than two regimes whose order cannot be defined[14].

The regime indicator s_t^{pol} is determined according to equation (2). We need to restrict some elements of λ to zero, however. This is because we have a few dozen endogenous variables x_t in the solved DSGE system, and allowing all elements of λ to be unconstrained would add plenty of parameters to be estimated. Although there is no criterion on how to select variables mattering for regime shifts, we use the following regime determination rule.

$$s_t^{pol} = \begin{cases} AM/PF & \text{if } \tau^{pol} + \lambda_y(\hat{y}_{t-1} - \hat{y}_{t-1}^*) + \lambda_\pi\tilde{\pi}_{t-1} + \lambda_R\tilde{R}_{t-1} + \lambda_b\tilde{b}_{t-1} + \lambda_\tau\tilde{\tau}_{t-1} + \eta_t \geq 0 \\ PM/AF & \text{otherwise} \end{cases} \tag{6}$$

where $\eta_t = \rho_\eta\eta_{t-1} + \varepsilon_{\eta,t}$, $\varepsilon_{\eta,t} \sim N(0,1)$. Output gap $\hat{y} - \hat{y}^*$ is included in order to capture the business cycle fluctuation. The rest of the variables $\tilde{\pi}$, $\tilde{R}$, $\tilde{b}$, and $\tilde{\tau}$ play

[14] One may think about introducing two regime factors representing the monetary and fiscal policy stance respectively. It enables us to investigate two other policy combinations not considered here: AM/AF and PM/PF. Instead, I assume that the one regime indicator determines the monetary/fiscal policy mix jointly to reduce the computational burden. The consensus in the literature is that the periods classified as AM/AF and PM/PF are short compared to the others, possibly because they do not have unique rational expectations equilibrium (AM/AF gives an explosive equilibrium, and PM/PF is subject to the indeterminacy of equilibria).

important roles in the monetary/fiscal policy mix[15]. Allowing for serial correlation in the error term η_t crucially matters for the estimation results because, if not, the persistence of s_t^{pol} depends sorely on that of economic variables.

6.2. Data, Solution, and Estimation

We construct the dataset based on Bianchi and Ilut (2017) which consists of quarterly observations of real output growth, inflation rate, debt-to-GDP ratio, federal tax revenues to GDP ratio, federal expenditure to GDP ratio, and government purchases to GDP ratio[16]. The sample covers 1954Q4-2009Q3.

The model needs to be solved for each draw of the parameters to get state space representations. I employ the perturbation method proposed by Maih and Waggoner (2018) which accommodates regime-switching DSGE with time-varying transition probabilities as in our application. Maih and Waggoner's perturbation method is available in the Rationality In Switching Environments (RISE) Toolbox developed by Junior Maih[17]. I utilize the built-in functions in RISE to solve the model and then plug in the resulting state space representation into my filtering algorithm to obtain the likelihood. The Maih and Waggoner (2018) perturbation method is discussed in more detail in Appendix C.

Due to the nonlinearity of our model, the posterior distribution might be irregular or multimodal. Given this consideration, the model is estimated via Sequential Monte Carlo (SMC) which is a variant of the Markov Chain Monte Carlo methods and is known to be robust to such a nonstandard posterior distribution. Appendix C provides the details. Another attractive feature of SMC is that we can parallelize the likelihood evaluation of particles, which is the most computationally demanding part. This algorithm is executed by the Big Red 200, the supercomputer possessed by Indiana University. With 24 cores, it takes about 26 hours to complete the whole run[18].

[15]We may employ machine learning techniques to allow every element of λ to be unrestricted. Although this extension is beyond the scope of this paper, it might be interesting to investigate how we can improve our estimation strategy by incorporating machine learning methods.

[16]More rigorously, the dataset is constructed following the description by Bianchi and Melosi (2022). This is because Bianchi and Ilut (2017) relies on the old definition in NIPA, which does not perfectly align with the currently available NIPA table.

[17]https://github.com/jmaih/RISE_toolbox

[18]Big Red 200 equips with 2.25 GHz AMD EPYC 7742 processors.

6.3. Priors and Calibration

The right columns of Tables 5 and 6 show the prior distributions. Other than the parameters governing the transition of policy regime, I adopt exactly the same prior as in Bianchi and Ilut (2017). The persistency of long-term expenditure and its volatility are fixed at $\rho_{e^L} = 0.99$ and $\sigma_{e^L} = 0.01$. The discount factor is set to be $\beta = 0.9965$, and ρ is calibrated to match the average debt maturity of 5 years: $\rho = 0.9513$. We assume $\delta_b(AF) = 0.0$. For the parameters related to the regime determination, the prior means of τ^{pol} and ρ_η imply that the probability of staying in the same regime is 0.85 absent of the feedback from the endogenous variables. Each element of λ follows the uniform distribution centered at zero with wide enough ranges so that we do not reflect any information on signs of the feedback a priori.

6.4. Posteriors

The left columns of Tables 5 and 6 display the posterior distribution of each parameter for the baseline model and exogenous switching model where the transition probability of policy regimes is constant. We specify the exogneous switching model by fixing all λ's to be zero. Qualitatively, the posterior means are similar to those reported in Bianchi and Ilut (2017), although there are some quantitative differences between the two. For example, the coefficients of inflation rate in the regime-switching Taylor rule are estimated to be lower: 0.2251 vs 0.5343 in PM and 1.8745 vs 2.6787 in AM. The regime transition probabilities for volatility are close to the estimates in Bianchi and Ilut (2017). The parameter governing the persistency of the policy regime, ρ_η, is estimated to be 0.9516, suggesting that the policy regime is likely to stay the same in the subsequent period in the absence of the feedback.

The baseline model and exogenous switching model give the similar posterior distribution broadly speaking, while we find disagreement in τ^{pol}, the threshold in the policy regime rule. This parameter is estimated to be smaller in the baseline model than the exogenous switching model, which suggests that the policy regime is more likely to be PM/AF in the baseline model. This difference is reflected in the policy regime probabilities inferred from two models, as we see below.

6.4.1. Feedback Coefficients λ

The posterior mean of λ_b is positive and its 90% posterior band does not include zero. This estimate means that the policy regime is more likely to be AM/PF after observing

TABLE 5. Prior and Posterior Distributions: Part1

	Posterior (Baseline)			Posterior (Exog. Switch)			Prior		
	Mean	5%	95%	Mean	5%	95%	Type	Para (1)	Para (2)
δ_y	0.2931	0.2387	0.3502	0.2945	0.2393	0.3509	N	0.2	0.2
δ_e	0.4029	0.2983	0.5060	0.2877	0.1788	0.3911	N	0.5	0.25
ι_y	0.0909	-0.1123	0.2965	0.0950	-0.1156	0.3089	N	0.1	0.2
ϕ_y	-0.6181	-0.6800	-0.5571	-0.6052	-0.6714	-0.5397	N	0.1	0.2
ς	0.5242	0.4690	0.5789	0.5044	0.4487	0.5577	B	0.5	0.25
Φ	0.4068	0.3648	0.4486	0.3836	0.3390	0.4269	B	0.5	0.25
κ	0.0016	0.0013	0.0018	0.0015	0.0012	0.0017	G	0.3	0.15
ρ_χ	0.9955	0.9931	0.9978	0.9954	0.9928	0.9978	B	0.5	0.2
ρ_a	0.6566	0.5939	0.7189	0.6982	0.6361	0.7597	B	0.5	0.2
ρ_d	0.9694	0.9651	0.9737	0.9719	0.9675	0.9759	B	0.5	0.2
ρ_{eS}	0.1700	0.1271	0.2158	0.1816	0.1353	0.2292	B	0.2	0.05
ρ_μ	0.0454	0.0189	0.0750	0.0467	0.0191	0.0765	B	0.5	0.2
ρ_{tp}	0.2165	0.1497	0.2830	0.2166	0.1445	0.2874	B	0.5	0.2
100π	0.5599	0.5243	0.5943	0.5861	0.5519	0.6194	N	0.5	0.05
100γ	0.4995	0.4595	0.5400	0.5008	0.4599	0.5401	N	0.42	0.05
b^m	0.6977	0.6485	0.7453	0.7344	0.6748	0.7938	N	1.0	0.1
g	1.0805	1.0764	1.0845	1.0792	1.0746	1.0838	N	1.08	0.04
τ	0.1716	0.1699	0.1733	0.1709	0.1693	0.1724	N	0.18	0.005

Note: This table describes the posterior mean as well as the posterior 5 and 95 percentiles for each parameter along with the prior distributions. The characters 'N', 'B', 'G', 'IG', and 'U' in 'Type' refer to normal, beta, gamma, inverse gamma, and uniform distributions respectively. Except for U, Para (1) and (2) are the prior mean and standard deviation respectively. For U, those two parameters are the lower and upper bound.

an increase in the debt-to-output ratio in the previous quarter. This feedback channel can be justified from the optimal policy perspective. Suppose that the policy authority observes a high debt-to-output ratio when they are taking the PM/AF policy. Under the PM/AF regime, public debt will be inflated away to satisfy the transversality condition. High inflation leads to large price dispersion, and it causes welfare loss in the New Keynesian context. To stabilize the inflation rate, the policy authority has the incentive to switch to the AM/PF regime. The positively estimated λ_π can also be interpreted from the view that high inflation leads to the combination of hawkish monetary policy and dovish fiscal policy.

Another parameter regarding the fiscal policy, λ_τ takes positive posterior mean, although its 90% posterior band includes zero. The positive point estimate is consistent

TABLE 6. Prior and Posterior Distributions: Part2

	Posterior (Baseline)			Posterior (Exog. Switch)			Prior		
	Mean	5%	95%	Mean	5%	95%	Type	Mean	Std
$\psi_\pi(PM)$	0.2251	0.1584	0.2942	0.2325	0.1704	0.2986	G	0.8	0.3
$\psi_\pi(AM)$	1.8745	1.4786	2.2771	1.7605	1.3634	2.1701	N	2.5	0.5
$\psi_y(PM)$	0.2125	0.1884	0.2370	0.1875	0.1671	0.2088	G	0.15	0.1
$\psi_y(AM)$	1.0368	0.8386	1.2490	1.0929	0.8938	1.3004	G	0.4	0.2
$\rho_R(PM)$	0.7395	0.7057	0.7710	0.6755	0.6390	0.7137	B	0.5	0.2
$\rho_R(AM)$	0.8559	0.8319	0.8787	0.8674	0.8449	0.8885	B	0.5	0.2
$\delta_b(PF)$	0.0498	0.0368	0.0634	0.0470	0.0340	0.0607	G	0.07	0.02
$\rho_\tau(AF)$	0.8154	0.7705	0.8609	0.7921	0.7406	0.8417	B	0.5	0.2
$\rho_\tau(PF)$	0.9799	0.9707	0.9885	0.9779	0.9679	0.9870	B	0.5	0.2
$100\sigma_R(1)$	0.0732	0.0673	0.0789	0.0725	0.0662	0.0790	IG	0.5	0.5
$100\sigma_R(2)$	0.3743	0.3364	0.4134	0.3900	0.3463	0.4354	IG	0.5	0.5
$100\sigma_\chi(1)$	1.9554	1.8201	2.0970	1.9874	1.8320	2.1505	IG	1.0	1.0
$100\sigma_\chi(2)$	4.6480	4.1861	5.1357	4.7152	4.2359	5.2266	IG	1.0	1.0
$100\sigma_a(1)$	0.3829	0.3311	0.4355	0.3498	0.2991	0.4013	IG	1.0	1.0
$100\sigma_a(2)$	0.6981	0.5602	0.8410	0.6184	0.4991	0.7438	IG	1.0	1.0
$100\sigma_\tau(1)$	0.2582	0.2400	0.2768	0.2548	0.2359	0.2745	IG	2.0	2.0
$100\sigma_\tau(2)$	0.7366	0.6653	0.8094	0.7310	0.6557	0.8115	IG	2.0	2.0
$100\sigma_d(1)$	6.2614	5.6140	6.9122	6.5964	5.9123	7.3294	IG	10.0	2.0
$100\sigma_d(2)$	10.7919	9.4271	12.2127	10.9300	9.5206	12.4497	IG	10.0	2.0
$100\sigma_{eS}(1)$	0.2256	0.1905	0.2616	0.2257	0.1908	0.2637	IG	2.0	2.0
$100\sigma_{eS}(2)$	0.3872	0.3103	0.4661	0.3877	0.3052	0.4727	IG	2.0	2.0
$100\sigma_{tp}(1)$	2.5930	2.4328	2.7597	2.5849	2.4203	2.7555	IG	1.0	1.0
$100\sigma_{tp}(2)$	3.3134	2.9318	3.7040	3.3227	2.9250	3.7285	IG	1.0	1.0
$100\sigma_\mu(1)$	0.1438	0.1305	0.1567	0.1369	0.1259	0.1484	IG	1.0	1.0
$100\sigma_\mu(2)$	0.2616	0.2303	0.2947	0.2739	0.2435	0.3064	IG	1.0	1.0
$p^{vol}_{1,2}$	0.0911	0.0686	0.1140	0.1013	0.0763	0.1277	B	0.17	0.1
$p^{vol}_{2,1}$	0.2284	0.1670	0.2928	0.2482	0.1857	0.3164	B	0.17	0.1
τ^{pol}	-11.9810	-14.1410	-9.8778	-1.6169	-5.1271	1.6398	U	-50.0	50.0
$0.01\lambda_y$	-0.6144	-1.0558	-0.1975	—	—	—	U	-10.0	10.0
$0.01\lambda_\tau$	0.6545	-0.1323	1.5137	—	—	—	U	-10.0	10.0
$0.01\lambda_\pi$	1.7754	0.2565	3.3060	—	—	—	U	-10.0	10.0
$0.01\lambda_R$	1.2554	0.3728	2.1370	—	—	—	U	-10.0	10.0
$0.01\lambda_b$	0.0539	0.0390	0.0697	—	—	—	U	-10.0	10.0
ρ_η	0.9516	0.9325	0.9702	0.9935	0.9891	0.9973	B	0.9	0.05

Note: See the footnote for Table 5.

with the view that the fiscal authority cares fiscal discipline in the passive fiscal policy regime. The point estimate of λ_R being positive is associated with the fact that aggressive monetary policy usually comes with high interest rate.

The relationship of policy regime and business cycle is captured by λ_y. This parameter is point-estimated to be negative, which implies that expansion, i.e., positive output gap, is tied to the PM/AF regime. Although associating expansion (recession) with accommodate (contractionary) policy regime seems to be counterintuitive, it is not necessarily the case. For example, the hawkish monetary policy under Paul Volcker as a chair of Federal Reserve took place even after entering the recession.

6.5. Regime Probability

Figure 1 reports the updated regime probabilities at the posterior mean for our baseline model (top two panels) and exogenous switching model (bottom two panels). The first and third panels plot the probability of AM/PF regime, while the second and fourth panel plot the probability of high volatility regime. Two models give the similar probabilities of volatility regime. A part of the reasons is related to our specification: We assume exogenous switching for macroeconomic volatility even in the endogenous switching model. We see the rise in uncertainty around the first oil crisis in the early 1970s, the appointment of Volcker as the Fed chairman (around 1980), the collapse of the Dot-Com bubble (around 2000), and the Great Recession (around the end of sample).

Turning to the policy regime, the two figures share some characteristics. We observe the PM/AF policy at 1960s and 70s. In the early 1980s, the macroeconomic policy switches to AM/PF and stays there by around 2000. We finally come back to the AM/PF regime at the end of sample. The notable differences appear in the mid 1950s and mid 1960s when we do not observe the AM/PF regime in the baseline model while we do in the exogenous switching model, and early-to-mid 2000s when the PM/AF policy takes place in the baseline model while we stay in the AM/PF regime in the exogenous switching model.

In the mid 1950s, fiscal policy returned to normalcy after the Korean war by cutting down defense expenditure. On the monetary policy side, "[t]he Fed might well have intended to be vigilant against inflation, but it appears not to have acted to prevent the 1955 inflation", as described by Davig and Leeper (2006). These narratives convince us to categorize this period as the PM/PF regime. The baseline (exogenous switching) model captures the PF (PM) nature in the regime probability. The policy stance in the mid 1960s is also mixed. The increase in the Federal funds rate was related to the concern to

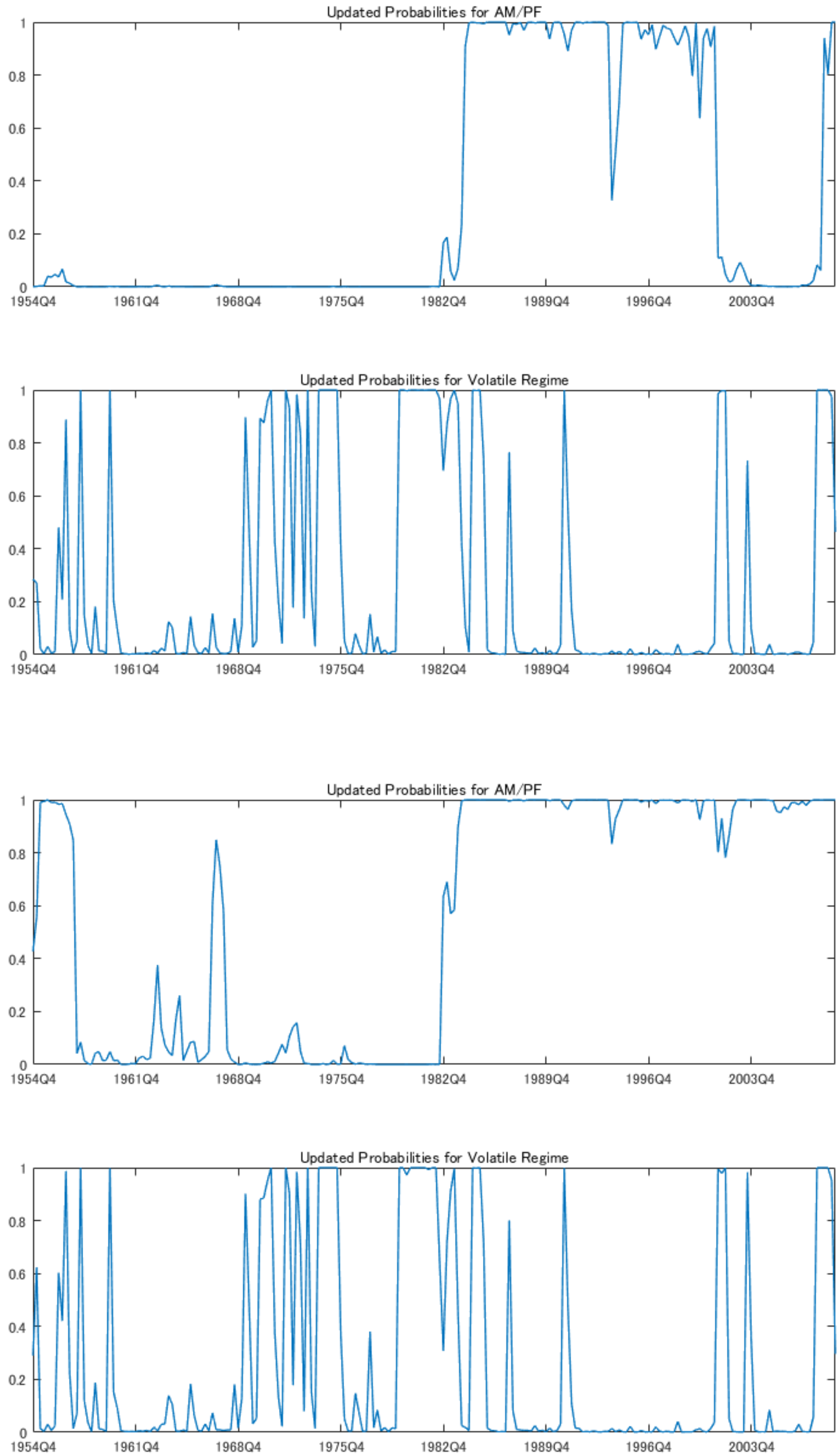

FIGURE 1. Updated Regime Probabilities from Baseline Model (Top Panel) and Exogenous Switching Model (Bottom Panel)

surging inflation due mainly to escalated fiscal spending related to the Vietnam war. There was a conflict between contractionary monetary policy and expansionary fiscal policy (Blinder 2022), which might have caused instability in the regime probability in the exogenous switching model.

The narrative episodes are in favor the PM/AF regime in the pre Great Recession. The policy rate was kept low to help the US economy recover from the Dot-Com bubble collapse in the early 2000s. The expansionary monetary policy is accompanied with George Bush's tax cut in 2002 and 2004. The high probability of the PM/AF regime from the baseline model aligns better with such historical accounts compared with the probability from the exogenous switching model.

6.6. Forecasting

The endogenous regime switching model is useful especially in the context of forecasting. Unlike the exogenous switching model where the regime transition probabilities are time-invariant, the framework presented here is able to predict the regime transition probability, improving the forecasting performance around the time of regime change.

6.6.1. Transition Probability

Figure 2 shows the transition probability $p(s_t|s_{t-1}, \mathcal{F}_{t-1})$ and the updated regime probability of AM/PF regime $p(s_t = AM/PF \mid \mathcal{F}_t)$. The left panel plots the probability of staying in AM/PF regime for the baseline model (solid line labeled "Endo") and the exogenous switching model (dashed line labeled "Exog") and the AM/PF regime probability (dotted line). By construction, the exogenous switching model exhibits the flat transition probability. In contrast, the transition probability under the baseline model decreases at 2001Q3, indicating that the regime likely to switch from the AM/PF to PM/AF. Indeed, this is the quarter when the policy stance changes from AM/PF to PM/AF according to the regime probability. Our baseline model succeeds in predicting the policy regime change at that period.

The right panel plots the probability of staying in PM/AF regime along with the AM/PF regime probability. Although the regime switching occurs at 2008Q4, the drop in transition probability happens 2009Q1. Such a lagged response in the transition probability happens due to the different information treatment in these probabilities. We predict the transition probability based on the information up to period $t-1$, while the regime probability is calculated with the information up to period t. This implies that

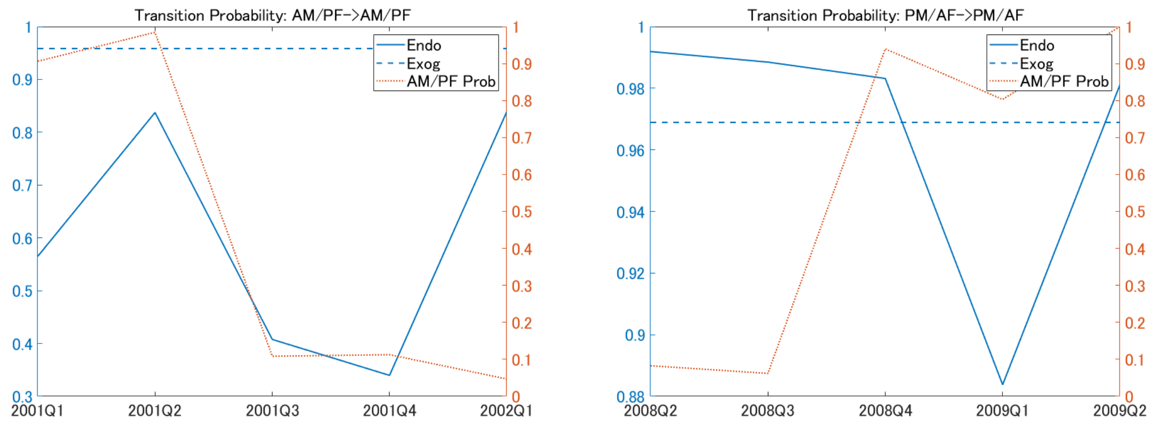

FIGURE 2. Transition Probability

Note: The left panel plots the transition probability from AM/PF at time $t-1$ to AM/PF at time t, for baseline ("Endo") and exogenous switching ("Exog") models along with the updated AM/PF probability in 2001Q1-2002Q1. The right panel plots the transition probability from PM/AF at time $t-1$ to PM/AF at time t along with the updated AM/PF probability in 2008Q2-2009Q2. For both panels, the left axis corresponds to the transition probability, and the right axis corresponds to the regime probability.

shocks happening at 2008Q4 were so extreme to invoke regime switching, which cannot be predicted by the information we have at 2008Q3. After we update the information at 2008Q4, we predict the low probability of staying in PM/AF at 2009Q1.

6.6.2. Real-Time Forecasting of GDP Growth and Inflation

The availability of prediction on regime transition probability improves the forecasting performance of economic variables as well. To see this, we conduct the real time forecasting exercise: Given the information available at those periods, we use the models to forecast real GDP growth and inflation rate. Figure 3 shows the 4-quarter-ahead real time forecast of GDP growth and inflation rate at two periods just before the regime switching: 1983Q4 (upper panels) and 2001Q2 (lower panels). The forecast generated by the baseline model (solid line labeled "Endo") predicts the variables better than those generated by the exogenous switching model (dashed line labeled "Exog"), especially at short horizon. The baseline model is able to capture changes in transition probability, a feature absent in the exogenous switching model. This helps the baseline model to make more accurate prediction.

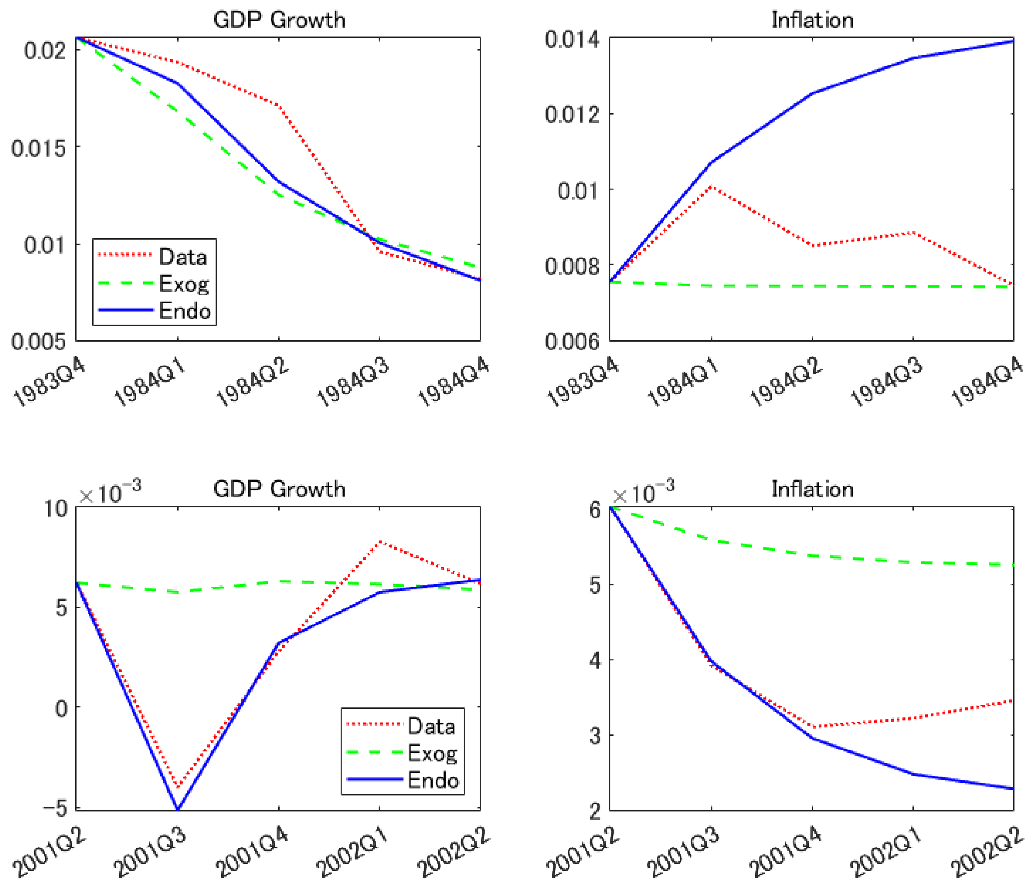

FIGURE 3. Real-Time Forecasting

Note: 4-quarter-ahead forecast of GDP growth and inflation rate given the information at 1983Q4 (upper panels) and 2001Q2 (lower panels).

7. Conclusion

This paper develops state space models with regime-switching coefficients allowing for the feedback from lagged continuous state variables into regime determination. I incorporate such a regime rule into the regime-switching Kalman filter and show that regime transition probabilities are given by functions of updated distributions of the state variables.

To circumvent the path dependence problem, we need to truncate the history of regimes we keep track of. I firstly prove that the likelihood from the filter with such a truncation step is asymptotically equivalent to the one from the exact filter if we increase the number of periods to take into account as the sample length grows. The second set of econometric claims concerns the consistency of the maximum likelihood estimator.

Consistency can be established for regime transition probabilities more general than the proposed ones. These results are confirmed in the simulation exercises using two types of data generating processes, a simple univariate model and a model mimicking the small-scale New Keynesian model.

I finally study the monetary/fiscal policy mix in the post-war U.S. using the regime-switching DSGE model with the proposed regime determination rule. By capturing changes in regime transition probability, we can make a better forecast especially at the times when a regime change is likely to happen.

The framework proposed in this paper can be applied in other models as well, such as the regime-switching DSGE models studying financial friction or macroeconomic volatility. These applications seem to be promising and are left for future work.

References

An, Sungbae and Frank Schorfheide (2007) "Bayesian Analysis of DSGE Models," *Econometric Reviews*, 26 (2-4), 113–172.

Aruoba, Borağan S, Pablo Cuba-Borda, and Frank Schorfheide (2018) "Macroeconomic Dynamics near the ZLB: A Tale of Two Countries," *The Review of Economic Studies*, 85 (1), 87–118.

Ascari, Guido, Paolo Bonomolo, and Qazi Haque (2022) "The Long-Run Phillips Curve Is... a Curve."

Auerbach, Alan J and Yuriy Gorodnichenko (2012) "Measuring The Output Responses to Fiscal Policy," *American Economic Journal: Economic Policy*, 4 (2), 1–27.

Benigno, Gianluca, Andrew Foerster, Christopher Otrok, and Alessandro Rebucci (2020) "Estimating Macroeconomic Models of Financial Crises: An Endogenous Regime-Switching Approach."

Bhatia, Rajendra (1997) *Matrix Analysis*: Springer.

Bianchi, Francesco (2012) "Evolving Monetary/Fiscal Policy Mix in the United States," *American Economic Review*, 102 (3), 167.

Bianchi, Francesco and Cosmin Ilut (2017) "Monetary/Fiscal Policy Mix and Agents' Beliefs," *Review of Economic Dynamics*, 26, 113–139.

Bianchi, Francesco, Cosmin L Ilut, and Martin Schneider (2018) "Uncertainty shocks, Asset Supply and Pricing over the Business Cycle," *Review of Economic Studies*, 85 (2), 810–854.

Bianchi, Francesco and Leonardo Melosi (2017) "Escaping the Great Recession," *American Economic Review*, 107 (4), 1030–58.

——— (2022) "Inflation as a Fiscal Limit," *Jackson Hole Symposium*.

Blinder, Alan S (2022) *A Monetary and Fiscal History of the United States, 1961-2021*: Princeton University Press.

Chang, Yoosoon, Yongok Choi, and Joon Y Park (2017) "A New Approach to Model Regime Switching," *Journal of Econometrics*, 196 (1), 127–143.

Chang, Yoosoon, Junior Maih, and Fei Tan (2021) "Origins of Monetary Policy Shifts: A New

Approach to Regime Switching in DSGE Models," *Journal of Economic Dynamics and Control*, 133, 104235.

Cochrane, John H (2023) *The Fiscal Theory of the Price Level*: Princeton University Press.

Davig, Troy and Eric M Leeper (2006) "Endogenous Monetary Policy Regime Change," in *NBER International Seminar on Macroeconomics*, 2006, 345–391, The University of Chicago Press Chicago, IL.

Diebold, Francis X., Joon-Haeng Lee, and Gretchen C. Weinbach (1994) "Regime Switching with Time-Varying Transition Probabilities," in Hargreaves, Colin P. ed. *Nonstationary Time Series Analysis and Cointegration. (Advanced Texts in Econometrics)*, 283–302, Oxford: Oxford University Press.

Douc, Randal and Eric Moulines (2012) "Asymptotic Properties of the Maximum Likelihood Estimation in Misspecified Hidden Markov Models," *Annals of Statistics*, 40 (5), 2697–2732.

Douc, Randal, Eric Moulines, and Tobias Rydén (2004) "Asymptotic Properties of the Maximum Likelihood Estimator in Autoregressive Models with Markov Regime," *Annals of Statistics*, 32 (5), 2254–2304.

Galí, Jordi (2015) *Monetary Policy, Inflation, and the Business Cycle: An Introduction to the New Keynesian Framework and Its Applications*: Princeton University Press.

Hamilton, James D (1989) "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57 (2), 357–384.

——— (1994) *Time Series Analysis*: Princeton University Press.

——— (2016) "Macroeconomic Regimes and Regime Shifts," *Handbook of Macroeconomics*, 2, 163–201.

Herbst, Edward and Frank Schorfheide (2014) "Sequential Monte Carlo Sampling for DSGE Models," *Journal of Applied Econometrics*, 29 (7), 1073–1098.

——— (2016) *Bayesian Estimation of DSGE Models*: Princeton University Press.

Ipsen, Ilse C. F. and Rizwana Rehman (2008) "Perturbation Bounds for Determinants and Characteristic Polynomials," *SIAM Journal on Matrix Analysis and Applications*, 30 (2), 762–776.

Jazwinski, Andrew H (1970) *Stochastic Processes and Filtering Theory*: Academic Press.

Kasahara, Hiroyuki and Katsumi Shimotsu (2019) "Asymptotic Properties of the Maximum Likelihood Estimator in Regime Switching Econometric Models," *Journal of Econometrics*, 208 (2), 442–467.

Kim, Chang-Jin (1994) "Dynamic Linear Models with Markov-Switching," *Journal of Econometrics*, 60 (1-2), 1–22.

Kim, Young Min and Kyu Ho Kang (2019) "Likelihood Inference for Dynamic Linear Models with Markov Switching Parameters: On the Efficiency of the Kim Filter," *Econometric Reviews*, 38 (10), 1109–1130.

Leeper, Eric M (1991) "Equilibria under 'Active' and 'Passive' Monetary and Fiscal Policies," *Journal of Monetary Economics*, 27 (1), 129–147.

Li, Chaojun (2023) "Asymptotic Properties of Approximated Maximum Likelihood Estimator in Markov-Switching State-Space Models," Unpublished.

Li, Chaojun and Yan Liu (2023) "Asymptotic Properties of the Maximum Likelihood Estimator in Regime-Switching Models with Time-Varying Transition Probabilities," *The Econometrics Journal*, 26 (1), 67–87.

Liu, Zheng, Daniel F Waggoner, and Tao Zha (2011) "Sources of Macroeconomic Fluctuations: A Regime-Switching DSGE Approach," *Quantitative Economics*, 2 (2), 251–301.

Maih, Junior and Daniel Waggoner (2018) "Perturbation Methods for DSGE Models with Time-Varying Coefficients and Transition Matrices."

Nimark, Kristoffer P (2014) "Man-bites-dog Business Cycles," *American Economic Review*, 104 (8), 2320–2367.

Pouzo, Demian, Zacharias Psaradakis, and Martin Sola (2022) "Maximum Likelihood Estimation in Markov Regime-Switching Models With Covariate-Dependent Transition Probabilities," *Econometrica*, 90 (4), 1681–1710.

Appendix A. Filter for Regime-Switching State Space Model

The regime-switching state space model is given by

$$\begin{aligned} x_t &= A_{s_t} x_{t-1} + Q_{s_t} \varepsilon_t \\ y_t &= B_{s_t} x_t + R_{s_t} u_t \end{aligned}$$

where u_t and ε_t are independent and follow the standard Gaussian. We assume that (y_t) is observed while (x_t) is not. The regime s_t is determined by

$$s_t = \mathbf{1}\left\{\tau + \lambda' x_{t-1} + \eta_t \geq 0\right\}$$

To derive the transition probability at time t, we need the conditional mean and variance of x_{t-2} given $\mathcal{F}_{t-1}$. To calculate this, we consider the lag-augmented state space model instead of the original model.

$$\begin{aligned} \begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix} &= \begin{bmatrix} A_{s_t} & O \\ I & O \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \end{bmatrix} + \begin{bmatrix} Q_{s_t} \\ O \end{bmatrix} \varepsilon_t \\ y_t &= \begin{bmatrix} B_{s_t} & O \end{bmatrix} \begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix} + R_{s_t} u_t \end{aligned}$$

We re-define $x_t = [x_t', x_{t-1}']'$, and so forth. Below is the algorithm to compute the approximated likelihood with the regime truncation. As is in the main text, we use the upper bar and subscript r to emphasize that the objects are coming from the truncated filter.

- Initialization: Set $\bar{x}_{1|1}(i)$ and $\bar{\Omega}_{1|1}(i)$ for $i = 0, 1$. We also set the initial regime probability $p_r(s_1 = i)$ for $i = 0, 1$.

- For $t = 2, \cdots, T$,

  (i) Forecasting Step

$$\bar{x}_{t|t-1}(s_{t-r+1}^t) = A_{s_t} \bar{x}_{t-1|t-1}(s_{t-r+1}^{t-1}) \tag{A1}$$

$$\bar{\Omega}_{t|t-1}(s_{t-r+1}^t) = A_{s_t} \bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-1}) A_{s_t}' + Q_{s_t} Q_{s_t}' \tag{A2}$$

$$\bar{y}_{t|t-1}(s_{t-r+1}^t) = B_{s_t} \bar{x}_{t|t-1}(s_{t-r+1}^t) \tag{A3}$$

$$\bar{\Sigma}_{t|t-1}(s_{t-r+1}^{t}) = B_{s_t}\bar{\Omega}_{t|t-1}(s_{t-r+1}^{t})B_{s_t}' + R_{s_t}R_{s_t}' \tag{A4}$$

(ii) Calculating Likelihood

$$p_r(y_t|\mathcal{F}_{t-1}) = \sum_{s_{t-r+1}^{t}} p_r(y_t|s_{t-r+1}^{t}, \mathcal{F}_{t-1})p_r(s_t|s_{t-r+1}^{t-1}, \mathcal{F}_{t-1})p_r(s_{t-r+1}^{t-1}|\mathcal{F}_{t-1}) \tag{A5}$$

The first part is given by $y_t|s_{t-r+1}^{t}, \mathcal{F}_{t-1} \sim N\left(\bar{y}_{t|t-1}(s_{t-r+1}^{t}), \bar{\Sigma}_{t|t-1}(s_{t-r+1}^{t})\right)$. The third part is obtained from equation (A8) in the previous iteration. The second part is the transition probability from s_{t-r+1}^{t-1} to s_t conditional on the information set $\mathcal{F}_{t-1}$, which will be elaborated on later.

(iii) Updating step

$$\begin{aligned}&\bar{x}_{t|t}(s_{t-r+1}^{t})\\ =&\bar{x}_{t|t-1}(s_{t-r+1}^{t}) + \bar{\Omega}_{t|t-1}(s_{t-r+1}^{t})B_{s_t}'\left[\bar{\Sigma}_{t|t-1}(s_{t-r+1}^{t})\right]^{-1}\left(y_t - \bar{y}_{t|t-1}(s_{t-r+1}^{t})\right)\end{aligned} \tag{A6}$$

$$\begin{aligned}&\bar{\Omega}_{t|t}(s_{t-r+1}^{t})\\ =&\bar{\Omega}_{t|t-1}(s_{t-r+1}^{t}) - \bar{\Omega}_{t|t-1}(s_{t-r+1}^{t})B_{s_t}'\left[\bar{\Sigma}_{t|t-1}(s_{t-r+1}^{t})\right]^{-1}B_{s_t}\bar{\Omega}_{t|t-1}(s_{t-r+1}^{t})\end{aligned} \tag{A7}$$

and

$$p_r(s_{t-r+2}^{t}|\mathcal{F}_t) = \sum_{s_{t-r+1}} p_r(s_{t-r+1}^{t}|\mathcal{F}_t) \tag{A8}$$

where

$$p_r(s_{t-r+1}^{t}|\mathcal{F}_t) = \frac{p_r(y_t|s_{t-r+1}^{t}, \mathcal{F}_{t-1})p_r(s_{t-r+1}^{t}|\mathcal{F}_{t-1})}{p_r(y_t|\mathcal{F}_{t-1})} \tag{A9}$$

(iv) Truncation

$$\bar{x}_{t|t}(s_{t-r+2}^{t}) = \sum_{s_{t-r+1}} \frac{p_r(s_{t-r+1}^{t}|\mathcal{F}_t)}{p_r(s_{t-r+2}^{t}|\mathcal{F}_t)}\bar{x}_{t|t}(s_{t-r+1}^{t}) \tag{A10}$$

$$\bar{\Omega}_{t|t}(s_{t-r+2}^{t}) = \sum_{s_{t-r+1}} \frac{p_r(s_{t-r+1}^{t}|\mathcal{F}_t)}{p_r(s_{t-r+2}^{t}|\mathcal{F}_t)}\bar{\Omega}_{t|t}(s_{t-r+1}^{t}) \tag{A11}$$

REMARK A1. *When $r = 1$, s_{t-r+2}^{t} is not well defined. In this context, we drop (A8) and write*

(A1), (A2), (A10), and (A11) to be

$$\begin{aligned}
\bar{x}_{t|t-1}(s_t) &= A_{s_t}\bar{x}_{t-1|t-1} \\
\bar{\Omega}_{t|t-1}(s_t) &= A_{s_t}\bar{\Omega}_{t-1|t-1}A'_{s_t} + Q_{s_t}Q'_{s_t} \\
\bar{x}_{t|t} &= \sum_{s_t} p_r(s_t|\mathcal{F}_t)\bar{x}_{t|t}(s_t) \\
\bar{\Omega}_{t|t} &= \sum_{s_t} p_r(s_t|\mathcal{F}_t)\bar{\Omega}_{t|t}(s_t)
\end{aligned}$$

REMARK A2. *Kim (1994) includes the second order adjustment term in (A11). As in Li (2023), we ignore that term because whether this additional term improves the approximation is not theoretically very clear, and the presence of the nonlinear term complicates our asymptotic analysis. As far as the simulation exercise tells, the approximation will be slightly improved by the inclusion of the second order term although the improvement is not drastic.*

A.1. Transition Probability

A.1.1. i.i.d. Gaussian Error

We assume the i.i.d. Gaussian error term: $\eta_t \sim N(0,1)$. We are interested in evaluating the transition probability, the second term in equation (A5).

$$\begin{aligned}
& p_r(s_t = 0|s_{t-1} = 0, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}) \\
=& \frac{p_r\left(s_t = 0, s_{t-1} = 0 \mid s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}\right)}{p_r\left(s_{t-1} = 0 \mid s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}\right)} \\
=& \frac{\int p_r(s_t = 0, s_{t-1} = 0|x_{t-1}, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1})p_r\left(x_{t-1}|s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}\right)dx_{t-1}}{\int p_r(s_{t-1} = 0|x_{t-1}, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1})p_r\left(x_{t-1}|s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}\right)dx_{t-1}} \\
\approx& \frac{\Phi\left(-\tau\iota; \Lambda\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2}), I + \Lambda\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\Lambda'\right)}{\Phi\left(-\tau; \lambda'\left(\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x)}, 1 + \lambda'\left(\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x, d_x+1:2d_x)}\lambda\right)}
\end{aligned}$$

where ι is a vector whose elements are all one, $\Lambda = \begin{bmatrix} \lambda' & 0 \\ 0 & \lambda' \end{bmatrix}$ is a $2 \times 2d_x$ matrix, $\left(\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x)}$ is a sub-vector of $\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2})$ consisting of its $(d+1)$-th to $2d_x$-th elements, and $\left(\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x, d_x+1:2d_x)}$ is a sub-matrix of

$\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})$ consisting of (d_x+1)-th to $2d_x$-th rows and columns. The conditional state mean and variance given s_{t-r+1}^{t-2} are computed as

$$\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2}) = \sum_{s_{t-1}} \frac{p_r\left(s_{t-1}, s_{t-r+1}^{t-2}|\mathcal{F}_{t-1}\right)}{p_r\left(s_{t-r+1}^{t-2}|\mathcal{F}_{t-1}\right)} \bar{x}_{t-1|t-1}(s_{t-1}, s_{t-r+1}^{t-2})$$
$$\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2}) = \sum_{s_{t-1}} \frac{p_r\left(s_{t-1}, s_{t-r+1}^{t-2}|\mathcal{F}_{t-1}\right)}{p_r\left(s_{t-r+1}^{t-2}|\mathcal{F}_{t-1}\right)} \bar{\Omega}_{t-1|t-1}(s_{t-1}, s_{t-r+1}^{t-2})$$

Analogously, the transition probability from $s_{t-1}=1$ to $s_t=1$ is given by

$$p_r(s_t=1|s_{t-1}=1, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1})$$
$$\approx \frac{\Phi\left(\tau\iota; -\Lambda\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2}), I + \Lambda\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\Lambda'\right)}{\Phi\left(\tau; -\lambda'\left(\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x)}, 1 + \lambda'\left(\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x, d_x+1:2d_x)}\lambda\right)}$$

The remaining probabilities are

$$p_r(s_t=1|s_{t-1}=0, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}) = 1 - p_r(s_t=0|s_{t-1}=0, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1})$$
$$p_r(s_t=0|s_{t-1}=1, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}) = 1 - p_r(s_t=1|s_{t-1}=1, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1})$$

A.2. Extension to Serially Correlated Error

Suppose that η_t follows AR(1), i.e., $\eta_t = \rho\eta_{t-1} + e_t$ where $e_t \sim N(0,1)$. The unconditional distribution of $[\eta_t, \eta_{t-1}]'$ is given by

$$\begin{bmatrix} \eta_t \\ \eta_{t-1} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \frac{1}{1-\rho^2} & \frac{\rho}{1-\rho^2} \\ \frac{\rho}{1-\rho^2} & \frac{1}{1-\rho^2} \end{bmatrix}\right)$$

Let Σ_η denote the variance-covariance matrix. Then, the transition probability is

$$p_r(s_t=0|s_{t-1}=0, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1})$$
$$\approx \frac{\Phi\left(-\tau\iota; \lambda'\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2}), \Sigma_\eta + \Lambda\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\Lambda'\right)}{\Phi\left(-\tau; \lambda'\left(\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x)}, \frac{1}{1-\rho^2} + \lambda'\left(\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x, d_x+1:2d_x)}\lambda\right)}$$

When implementing this filter computationally, the RISE requires the transition probability given x_{t-1}. Then, the transition probability of interest is given as

$$
\begin{aligned}
p(s_t = 0|s_{t-1} = 0, x_{t-1}) &= \frac{p(s_t = 0, s_{t-1} = 0|x_{t-1})}{p(s_{t-1} = 0|x_{t-1})} \\
&= \frac{p([\eta_t, \eta_{t-1}]' \leq -\tau\iota - \Lambda x_{t-1}|x_{t-1})}{p(\eta_{t-1} \leq -\tau - \lambda' x_{t-2}|x_{t-1})} \\
&= \frac{\Phi\left(-\tau\iota - \Lambda x_{t-1}; 0, \Sigma_\eta\right)}{\Phi(-\tau - \lambda' x_{t-2}; 0, (1-\rho^2)^{-1})}
\end{aligned}
$$

and

$$
p(s_t = 1|s_{t-1} = 1, x_{t-1}) = \frac{\Phi\left(\tau\iota + \Lambda x_{t-1}; 0, \Sigma_\eta\right)}{\Phi\left(\tau + \lambda' x_{t-2}; 0, (1-\rho^2)^{-1}\right)}
$$

Appendix B. Proofs

B.1. Proofs for Section 3

B.1.1. Auxiliary Lemmas

LEMMA A1. *Assume Assumptions 1 and 2. There exist positive constants c_Ω^+ and c_Ω^- such that $c_\Omega^- \leq \|\Omega_{t|t}(s_1^t)\| \leq c_\Omega^+$ for any $s_1^t \in \{0,1\}^t$.*

PROOF. Let $\alpha = \min\{\alpha_{UCO}, \alpha_{UCC}\}$ and $\beta = \max\{\beta_{UCO}, \beta_{UCC}\}$. Lemmas 7.1 and 7.2 in Jazwinski (1970) show $\frac{\alpha}{1+\alpha\beta} I \leq \Omega_{t|t}(s_1^t) \leq \frac{1+\alpha\beta}{\alpha} I$. Taking the matrix norm for both hand sides establishes our claim. $\square$

LEMMA A2. *Assume Assumptions 1 and 2. There exist a positive constant c_x such that for any $t = 1, 2, \cdots,$*

$$
\mathbb{E}_0 \left\| x_{t|t}(s_1^t) \right\| \leq c_x
$$

for any $s_1^t \in \{0,1\}^t$.

PROOF. The transition of $x_{t|t}$ is characterized as

$$
\begin{aligned}
\|x_{t|t}(s_1^t)\| &= \left\| (I - K(s_1^t) B_{s_t}) A_{s_t} x_{t-1|t-1}(s_1^{t-1}) + K(s_1^t) y_t \right\| \\
&\leq \cdots \\
&\leq \left\| \Psi^t(s_1^t) \right\| \|\tilde{x}\| + \sum_{i=1}^{t} \left\| \Psi^{i-1}(s_1^t) \right\| \cdot \left\| K(s_1^{t-i+1}) \right\| \cdot \|y_{t-i+1}\|
\end{aligned}
$$

Note that for any t, $\|K(s_1^t)\| \leq \|\Omega_{t-1|t-1}(s_1^{t-1})\| \cdot \|B_{s_t}\| \cdot \|(R_{s_t}R_{s_t}')^{-1}\| \leq c_\Omega \cdot \max_{s\in\{0,1\}}\{\|B_s\| \cdot \|(R_sR_s')^{-1}\|\} \equiv c_K$. Taking expectations for both hand sides of the inequality above and applying Lemma 1 yield

$$
\begin{aligned}
\mathbb{E}_0\|x_{t|t}(s_1^t)\| \leq & \left\|\Psi^t(s_1^t)\right\| \|\tilde{x}\| + \sum_{i=1}^{t} \left\|\Psi^{i-1}(s_1^t)\right\| \cdot \left\|K(s_1^{t-i+1})\right\| \cdot \mathbb{E}_0 \left\|y_{t-i+1}\right\| \\
\leq & c_1 \exp(-c_2)\|\tilde{x}\| + c_K\mu_y \frac{\exp(c_2)}{\exp(c_2)-1} \equiv c_x
\end{aligned}
$$

where $\mu_y = \mathbb{E}_0\|y_t\|$ for any t due to the stationarity. □

Note that Lemmas 1, A1 and A2 hold for the objects from the truncated filter as well.

LEMMA A3. *Let Ω_0 and Ω_1 be positive definite matrices. Let $K_i = \Omega_i B' \left(B\Omega_i B' + RR'\right)^{-1}$ for $i = 0, 1$, where B and R are matrices with comfortable sizes. Then,*

$$
\|K_0 - K_1\| \leq \|\Omega_0 - \Omega_1\| \cdot \|B\| \cdot \|(RR')^{-1}\| \left[1 + \|B\| \cdot \|(RR')^{-1}\| \cdot (\|B\| + \|\Omega_0 - \Omega_1\|)\right]
$$

PROOF. We employ $(A + BCB')^{-1} = A^{-1} - A^{-1}B(C^{-1} + B'A^{-1}B)^{-1}B'A^{-1}$. It follows that

$$
\begin{aligned}
K_0 =& \Omega_0 B' \left[\left(B\Omega_1 B' + RR\right) + B(\Omega_0 - \Omega_1)B'\right]^{-1} \\
=& (\Omega_1 + (\Omega_0 - \Omega_1))B' \Big[\left(B\Omega_1 B' + RR\right)^{-1} \\
& - \left(B\Omega_1 B' + RR\right)^{-1} B \left((\Omega_0 - \Omega_1)^{-1} + B'\left(B\Omega_1 B' + RR\right)^{-1} B\right)^{-1} B'\left(B\Omega_1 B' + RR\right)^{-1}\Big] \\
=& \left(\Omega_1 + (\Omega_0 - \Omega_1)\right)\Omega_1^{-1} \left[K_1 - K_1 B \left((\Omega_0 - \Omega_1)^{-1} + B'\left(B\Omega_1 B' + RR\right)^{-1} B\right)^{-1} \Omega_1^{-1} K_1\right] \\
=& K_1 + (\Omega_0 - \Omega_1)\Omega_1^{-1} K_1 \\
& - (I + (\Omega_0 - \Omega_1)\Omega_1^{-1})K_1 B \left((\Omega_0 - \Omega_1)^{-1} + B'\left(B\Omega_1 B' + RR\right)^{-1} B\right)^{-1} \Omega_1^{-1} K_1
\end{aligned}
$$

Note that $\|A^{-1}\| = \lambda_{min}(A)$ for an invertible matrix A where $\lambda_{min}(A)$ is the smallest eigenvalue of A. Note also that for two positive semi-definite $n \times n$ matrices A and B, we have $\lambda_i(A + B) \geq \lambda_i(A)$ for $i = 1, \cdots, n$ where $\lambda_i(A)$ is the i-th largest eigenvalue of A. Then, $\|(A+B)^{-1}\| \leq \|A^{-1}\|$. Therefore, we have

$$
\begin{aligned}
&\|K_0 - K_1\| \\
\leq &\|\Omega_0 - \Omega_1\| \cdot \|\Omega_1^{-1} K_1\| \\
+ &\left\|(I + (\Omega_0 - \Omega_1)\Omega_1^{-1})K_1 B\right\| \cdot \left\|\left((\Omega_0 - \Omega_1)^{-1} + B'\left(B\Omega_1 B' + RR\right)^{-1} B\right)^{-1}\right\| \cdot \|\Omega_1^{-1} K_1\| \\
\leq &\|\Omega_0 - \Omega_1\| \cdot \|B\| \cdot \|(RR')^{-1}\| \\
&+ \left(\|B\|^2 \|(RR')^{-1}\| + \|\Omega_0 - \Omega_1\| \cdot \|B\| \cdot \|(RR')^{-1}\|\right) \cdot \|\Omega_0 - \Omega_1\| \cdot \|B\| \cdot \|(RR')^{-1}\|
\end{aligned}
$$

□

LEMMA A4 (Corollary 2.14 in Ipsen and Rehman (2008)). *Let A and E be $n \times n$ matrices. If A is nonsingular, then*

$$
\frac{\det(A+E) - \det(E)}{\det(A)} \leq \left(1 + \|A^{-1}\| \times \|E\|\right)^n - 1
$$

LEMMA A5 (Equation (X.2) in Bhatia (1997)). *For $0 \leq r \leq 1$ and positive semidefinite matrices A and B,*

$$
\|A^r - B^r\| \leq \|A - B\|^r
$$

LEMMA A6. *For any $x \in \mathbb{R}$ and $\varepsilon \geq 0$, we have $\Phi_1(x+\varepsilon) - \Phi_1(x) \leq \varepsilon$.*

PROOF. By the mean value theorem, there exists $x^* \in [x, x+\varepsilon]$ such that $\Phi_1(x+\varepsilon) = \Phi_1(x) + \phi_1(x^*)\varepsilon$. The statement follows upon noticing $\phi_1(x^*) \leq \phi_1(0) < 1$. □

B.1.2. Updated Mean and Variance

We first provide two preliminary lemmas necessary to show the asymptotic negligibility of the difference between the truncated and exact updated mean and variance.

LEMMA A7. *Assume Assumptions 1 and 2. For $t \geq r$, define*

$$
\begin{aligned}
&\delta_t^{\Omega}(s_{t-r+2}^t, s_1^{t-r}) \\
= &\max_{i,j \in \{0,1\}} \|\Omega_{t|t}(s_{t-r+2}^t, i, s_1^{t-r}) - \Omega_{t|t}(s_{t-r+2}^t, j, s_1^{t-r})\|
\end{aligned}
$$

There exists positive constants c_δ and c_2 such that

$$\max_{s_t,\cdots,s_{t-r+2},s_{t-r},\cdots,s_1} \delta_t^{\Omega}(s_{t-r+2}^t, s_1^{t-r}) \leq c_\delta \exp(-2c_2(r-1))$$

PROOF. Take any $i, j \in \{0,1\}$, and $s_t, \cdots, s_{t-r+2}, s_{t-r}, \cdots, s_1$. For $\tau = t-r+1, \cdots, t$, let $s_1^\tau\langle i \rangle = (s_\tau, \cdots, s_{t-r+2}, i, s_{t-r}, \cdots, s_1)$. It follows from Lemmas A1 and 1 that

$$\begin{aligned}
&\delta_t^{\Omega}(s_{t-r+2}^t, s_1^{t-r}) \\
=&\Big\|(I - K(s_1^t\langle i \rangle)B_{s_t})A_{s_t} \\
&\times \Big(\Omega_{t-1|t-1}(s_1^t\langle i \rangle) - \Omega_{t-1|t-1}(s_1^t\langle j \rangle)\Big) \\
&\times A'_{s_t}(I - K(s_1^t\langle j \rangle)B_{s_t})'\Big\| \\
\leq& \cdots \\
\leq& \left\|\Psi^{r-1}(s_1^t\langle i \rangle)\right\| \cdot \left\|\Psi^{r-1}(s_1^t\langle j \rangle)\right\| \\
&\times \left\|\Omega_{t-r+1|t-r+1}(s_{t-r+1} = i, s_1^{t-r}) - \Omega_{t-r+1|t-r+1}(s_{t-r+1} = j, s_1^{t-r})\right\| \\
\leq& c_1^2 \exp(-2c_2(r-1)) \left\|\Omega_{t-r+1|t-r+1}(s_{t-r+1} = i, s_1^{t-r}) - \Omega_{t-r+1|t-r+1}(s_{t-r+1} = j, s_1^{t-r})\right\| \\
\leq& c_\delta \exp(-2c_2(r-1))
\end{aligned}$$

where $c_\delta = 2c_1^2 c_\Omega^+$. □

LEMMA A8. *Assume Assumptions 1 and 2. For $t \geq r$, define*

$$\begin{aligned}
&\delta_t^x(s_{t-r+2}^t, s_1^{t-r}) \\
=& \max_{i,j\in\{0,1\}} \left\|x_{t|t}(s_{t-r+2}^t, i, s_1^{t-r}) - x_{t|t}(s_t, \cdots, s_{t-r+2}, j, s_1^{t-r})\right\|
\end{aligned}$$

A stochastically bounded positive random variable m_t and a positive constant c_2 exist such that

$$\max_{s_t,\cdots,s_{t-r+2},s_{t-r},\cdots,s_1} \delta_t^x(s_{t-r+2}^t, s_1^{t-r}) \leq \exp(-c_2(r-1))m_t$$

PROOF. Take any $i, j \in \{0,1\}$ and $s_t, \cdots, s_{t-r+2}, s_{t-r}, \cdots, s_1$. We introduce another updated mean at time t constructed in the following way. At time $t-r+1$, the updated mean and variance are given by $x_{t-r+1|t-r+1}(s_{t-r+1} = j, s_1^{t-r})$ and $\Omega_{t-r+1|t-r+1}(s_{t-r+1} = i, s_1^{t-r})$. Using them, we evaluate the Kalman recursion up to time t where the sequence of regimes is given by $(s_t, \cdots, s_{t-r+2})$. We denote this alternative updated mean by $x^*_{t|t}(s^t_{t-r+2}, s_1^{t-r})$. We can decompose $\delta^x_t(\cdot)$ as

$$\delta^x_t(s^t_{t-r+2}, s_1^{t-r}) \leq \left\| x_{t|t}(s_1^t\langle i\rangle) - x^*_{t|t}(s^t_{t-r+2}, s_1^{t-r}) \right\| + \left\| x^*_{t|t}(s^t_{t-r+2}, s_1^{t-r}) - x_{t|t}(s_1^t\langle j\rangle) \right\| \tag{A12}$$

The first term can be evaluated as

$$\begin{aligned}
&\left\| x_{t|t}(s_1^t\langle i\rangle) - x^*_{t|t}(s^t_{t-r+2}, s_1^{t-r}) \right\| \\
\leq &\left\| \Psi^{r-1}(s_1^t\langle i\rangle) \right\| \times \left\| x_{t-r+1|t-r+1}(s_{t-r+1} = i, s_1^{t-r}) - x_{t-r+1|t-r+1}(s_{t-r+1} = j, s_1^{t-r}) \right\| \\
\leq &\exp(-c_2(r-1)) m_{1,t}
\end{aligned} \tag{A13}$$

where $m_{1,t} = c_1 \| x_{t-r+1|t-r+1}(s_{t-r+1} = i, s_1^{t-r}) - x_{t-r+1|t-r+1}(s_{t-r+1} = j, s_1^{t-r}) \|$. We decompose the second term of (A12) as

$$\begin{aligned}
&\left\| x^*_{t|t}(s^t_{t-r+2}, s_1^{t-r}) - x_{t|t}(s_1^t\langle j\rangle) \right\| \\
= &\Bigg\| \left(\Psi^{r-1}(s_1^t\langle i\rangle) - \Psi^{r-1}(s_1^t\langle j\rangle) \right) x_{t-r+1|t-r+1}(s_{t-r+1} = j, s_1^{t-r}) \\
&+ \sum_{k=1}^{r-1} \left(\Psi^{k-1}(s_1^t\langle i\rangle) K\left(s_1^{t-k+1}\langle i\rangle\right) - \Psi^{k-1}(s_1^t\langle j\rangle) K\left(s_1^{t-k+1}\langle j\rangle\right) \right) y_{t-k+1} \Bigg\| \\
\leq &\left\| \Psi^{r-1}(s_1^t\langle i\rangle) - \Psi^{r-1}(s_1^t\langle j\rangle) \right\| \times \left\| x_{t-r+1|t-r+1}(s_{t-r+1} = j, s_1^{t-r}) \right\| \\
&+ \sum_{k=1}^{r-1} \left\| \left(\Psi^{k-1}(s_1^t\langle i\rangle) K\left(s_1^{t-k+1}\langle i\rangle\right) - \Psi^{k-1}(s_1^t\langle j\rangle) K\left(s_1^{t-k+1}\langle j\rangle\right) \right) \right\| \times \| y_{t-k+1} \|
\end{aligned} \tag{A14}$$

The first term is bounded by $2c_1 \exp(-c_2(r-1)) \cdot \| x_{t-r+1|t-r+1}(s_{t-r+1} = j, s_1^{t-r}) \|$. Taking a further look at the second term,

$$\begin{aligned}
&\Psi^{k-1}(s_1^t\langle i\rangle) K\left(s_1^{t-k+1}\langle i\rangle\right) - \Psi^{k-1}(s_1^t\langle j\rangle) K\left(s_1^{t-k+1}\langle j\rangle\right) \\
= &\Psi^{k-1}(s_1^t\langle i\rangle) \left[K\left(s_1^{t-k+1}\langle i\rangle\right) - K\left(s_1^{t-k+1}\langle j\rangle\right) \right] + \left[\Psi^{k-1}(s_1^t\langle i\rangle) - \Psi^{k-1}(s_1^t\langle j\rangle) \right] K\left(s_1^{t-k+1}\langle j\rangle\right)
\end{aligned} \tag{A15}$$

Let $B_+ = \max_{s=0,1} \|B_s\|$ and $R_- = \max_{s=0,1} \|(R_s R_s')^{-1}\|$. By Lemma A3,

$$\begin{aligned}
&\left\| K\left(s_1^{t-k+1}\langle i\rangle\right) - K\left(s_1^{t-k+1}\langle j\rangle\right)\right\| \\
\leq &\left\|\Omega_{t-k|t-k}\left(s_1^{t-k}\langle i\rangle\right) - \Omega_{t-k|t-k}\left(s_1^{t-k}\langle j\rangle\right)\right\| B_+R_- \\
&\times \left[1 + B_+R_-\left(B_+ + \left\|\Omega_{t-k|t-k}\left(s_1^{t-k}\langle i\rangle\right) - \Omega_{t-k|t-k}\left(s_1^{t-k}\langle j\rangle\right)\right\|\right)\right]
\end{aligned}$$

Applying Lemma 5, we have $\|\Omega_{t-k|t-k}(s_1^{t-k}\langle i\rangle) - \Omega_{t-k|t-k}(s_1^{t-k}\langle j\rangle)\| \leq c_\delta \exp(-2c_2(r-k-1))$, and thus

$$\left\| K\left(s_1^{t-k+1}\langle i\rangle\right) - K\left(s_1^{t-k+1}\langle j\rangle\right)\right\| \leq c_{\delta K} \exp(-2c_2(r-k-1))$$

where $c_{\delta K} = c_\delta B_+R_-[1 + B_+R_-(B_+ + 2c_\Omega)]$. Together with Lemma 1, the whole first term of equation (A15) is bounded by

$$\begin{aligned}
&\left\|\Psi^{k-1}(s_1^t\langle i\rangle)\left[K\left(s_1^{t-k+1}\langle i\rangle\right) - K\left(s_1^{t-k+1}\langle j\rangle\right)\right]\right\| \\
\leq &c_1 \exp(-c_2(k-1)) \times c_K \exp(-2c_2(r-k-1)) \\
\leq &c_3 \exp(-c_2(2r-k-3))
\end{aligned}$$

where $c_3 = \max\{c_1, c_K\}$ is a positive constant. For the second term of (A15), note that

$$\begin{aligned}
\Psi(s_1^t\langle i\rangle) - \Psi(s_1^t\langle j\rangle) &= \left(I - K(s_1^t\langle i\rangle)B_{s_t}\right)A_{s_t} - \left(I - K(s_1^t\langle j\rangle)B_{s_t}\right)A_{s_t} \\
&= \left(-K(s_1^t\langle i\rangle) + K(s_1^t\langle j\rangle)\right)B_{s_t}A_{s_t}
\end{aligned} \tag{A16}$$

Then,

$$\begin{aligned}
&\Psi^{k-1}(s_1^t\langle i\rangle) - \Psi^{k-1}(s_1^t\langle j\rangle) = \Psi(s_1^t\langle i\rangle)\Psi^{k-2}(s_1^{t-1}\langle i\rangle) - \Psi(s_1^t\langle j\rangle)\Psi^{k-2}(s_1^{t-1}\langle j\rangle) \\
=&\Psi(s_1^t\langle i\rangle)\left(\Psi^{k-2}(s_1^{t-1}\langle i\rangle) - \Psi^{k-2}(s_1^{t-1}\langle j\rangle)\right) + \left(\Psi(s_1^t\langle i\rangle) - \Psi(s_1^t\langle j\rangle)\right)\Psi^{k-2}(s_1^{t-1}\langle j\rangle) \\
=&\cdots \\
=&\sum_{l=1}^{k-1}\Psi^{l-1}(s_1^t\langle i\rangle)\left(\Psi(s_1^{t-l}\langle i\rangle) - \Psi(s_1^{t-l}\langle j\rangle)\right)\Psi^{k-1-l}(s_1^{t-l}\langle j\rangle)
\end{aligned}$$

Consequently, we have

$$
\begin{aligned}
&\left\|\Psi^{k-1}(s_1^t\langle i\rangle) - \Psi^{k-1}(s_1^t\langle j\rangle)\right\| \\
\leq &\sum_{l=1}^{k-1} \left\|\Psi^{l-1}(s_1^t\langle i\rangle)\right\| \times \left\|\Psi(s_1^{t-l}\langle i\rangle) - \Psi(s_1^{t-l}\langle j\rangle)\right\| \times \left\|\Psi^{k-1-l}(s_1^{t-l}\langle j\rangle)\right\| \\
\leq &\sum_{l=1}^{k-1} c_1^2 c_{\delta K} \max_s\{\|B_s\| \cdot \|A_s\|\} \exp\left(-c_2(2r+k-2l-2)\right) \\
\leq &c_4 \exp\left(-c_2(2r+k-2)\right) \frac{\exp(2c_2)\left[1-\exp(2c_2(k-1))\right]}{1-\exp(2c_2)} \\
= &\frac{c_4}{1-\exp(2c_2)} \exp(-2c_2(r-1)) \left[\exp(-c_2(k-2)) - \exp(c_2 k)\right]
\end{aligned}
$$

where $c_4 = c_1^2 c_{\delta K} \max_s\{\|B_s\| \cdot \|A_s\|\}$. Then, the second term of (A15) is bounded by

$$
\begin{aligned}
&\left\|\left(\Psi^{k-1}(s_1^t\langle i\rangle) - \Psi^{k-1}(s_1^t\langle j\rangle)\right) K\left(s_1^{t-k+1}\langle j\rangle\right)\right\| \\
\leq &\frac{c_4 c_K}{1-\exp(2c_2)} \exp(-2c_2(r-1)) \left[\exp(-c_2(k-2)) - \exp(c_2 k)\right]
\end{aligned}
$$

Substituting them all together in equation (A14) yields

$$
\begin{aligned}
&\left\|x_{t|t}^* - x_{t|t}(s_1^t\langle j\rangle)\right\| \\
\leq &2c_1 \exp(-c_2(r-1)) \cdot \left\|x_{t-r+1|t-r+1}(s_1^{t-r}\langle j\rangle)\right\| + \sum_{k=1}^{r-1} c_3 \exp(-c_2(2r-k-3))\|y_{t-k+1}\| \\
&+ \sum_{k=1}^{r-1} \frac{c_4 c_K}{1-\exp(2c_2)} \exp(-2c_2(r-1)) \left[\exp(-c_2(k-2)) - \exp(c_2 k)\right] \times \|y_{t-k+1}\| \\
\leq &\exp(-c_2(r-1)) \left[2c_1 \left\|x_{t-r+1|t-r+1}(s_1^{t-r+1}\langle j\rangle)\right\| + \sum_{k=1}^{r-1} c_3 \exp(-c_2(r-k-2)) \|y_{t-r+1}\|\right. \\
&\left.+ \sum_{k=1}^{r-1} \frac{c_4 c_K}{1-\exp(2c_2)} \exp(-c_2(r-1)) \left[\exp(-c_2(k-2)) - \exp(c_2 k)\right] \times \|y_{t-r+1}\|\right] \\
\equiv &\exp(-c_2(r-1)) m_{2,t}
\end{aligned}
$$

Using this and (A13), equation (A12) becomes

$$
\delta_t^x(s_{t-r+2}^t, s_1^{t-r}) \leq \exp(-c_2(r-1)) m_t
$$

where $m_t = m_{1,t} + m_{2,t}$. It remains to see the stochastic boundedness of m_t. By Lemma

A2, $\mathbb{E}_0|m_{1,t}| \leq c_1(\|x_{t-r+1|t-r+1}(i, s_1^{t-r})\| + \|x_{t-r+1|t-r+1}(j, s_1^{t-r})\| \leq 2c_1c_x < \infty$. For $m_{2,t}$,

$$\begin{aligned}
&\mathbb{E}_0|m_{2,t}| \\
\leq& 2c_1\mathbb{E}_0\left\|x_{t-r+1|t-r+1}(s_1^{t-r+1}\langle j\rangle)\right\| + \|\mu_y\|\sum_{k=1}^{r-1} c_3\exp(-c_2(r-k-2)) \\
&+ \|\mu_y\|\sum_{k=1}^{r-1}\frac{c_4c_K\exp(-c_2(r-1))}{1-\exp(2c_2)}\left[\exp(-c_2(k-2)) - \exp(c_2k)\right] \\
\leq& 2c_1c_x + \|\mu_y\|c_3\frac{\exp(-c_2(r-3)) - \exp(-2c_2)}{1-\exp(c_2)} \\
&+ \|\mu_y\|\frac{c_4c_K\exp(-c_2(r-1))}{1-\exp(2c_2)}\left[\frac{\exp(c_2)[1-\exp(-c_2(r-1))]}{1-\exp(-c_2)} - \frac{\exp(c_2)[1-\exp(c_2(r-1))]}{1-\exp(c_2)}\right] \\
\leq& 2c_1c_x + \|\mu_y\|c_3\frac{-\exp(-2c_2)}{1-\exp(c_2)} + \|\mu_y\|\frac{c_4c_K}{1-\exp(2c_2)}\frac{\exp(c_2)}{1-\exp(c_2)} \\
<& \infty
\end{aligned}$$

By Markov inequality, $\mathbb{E}_0|m_t| = \mathbb{E}_0|m_{1,t} + m_{2,t}| < \infty$ implies $m_t = O_p(1)$. □

Using Lemmas A7 and A8, we can establish the proposition claiming that the updated mean and variance from the truncated and exact filters are asymptotically equivalent.

PROPOSITION A1. *Let $\Delta^{\Omega}_{r,t}(s_1^t) = \bar{\Omega}_{t|t}(s_{t-r+2}^t) - \Omega_{t|t}(s_1^t)$ and $\Delta^{x}_{r,t}(s_1^t) = \bar{x}_{t|t}(s_{t-r+2}^t) - x_{t|t}(s_1^t)$. Then, there exist positive constants $c_{\Omega,\Delta}$ and c_2 as well as a positive stochastically bounded random variable $M_{x,t}$ such that*

$$\max_{s_1^t}\|\Delta^{\Omega}_{r,t}(s_1^t)\| \leq c_{\Omega,\Delta}\exp(-2c_2(r-1))$$
$$\max_{s_1^t}\|\Delta^{x}_{r,t}(s_1^t)\| \leq M_{x,t}\exp(-c_2(r-1))$$

PROOF. Take any s_1^t. By applying the truncation step to $\bar{\Omega}_{t|t}(s_{t-r+2}^t)$ and $\bar{x}_{t|t}(s_{t-r+2}^t)$,

$$\begin{aligned}
\Delta_{r,t}^{\Omega}(s_1^t) =& p_r(s_{t-r+1}|s_{t-r+2}^t, \mathcal{F}_t)\left(\bar{\Omega}_{t|t}(s_{t-r+1}^t) - \Omega_{t|t}(s_1^t)\right) \\
&+ \sum_{s^* \neq s_{t-r+1}} p_r(s_{t-r+1} = s^*|s_{t-r+2}^t, \mathcal{F}_t)\left(\bar{\Omega}(s_{t-r+1} = s^*, s_{t-r+2}^t) - \Omega(s_1^t)\right) \\
=& p_r(s_{t-r+1}|s_{t-r+2}^t, \mathcal{F}_t)\bar{\Psi}(s_{t-r+1}^t)\left(\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-1}) - \Omega_{t-1|t}(s_1^{t-1})\right)\Psi(s_1^t)' \\
&+ \sum_{s^* \neq s_{t-r+1}} p_r(s_{t-r+1} = s^*|s_{t-r+2}^t, \mathcal{F}_t)\left(\bar{\Omega}(s_{t-r+1} = s^*, s_{t-r+2}^t) - \Omega(s_1^t)\right) \\
=& p_r(s_{t-r+1}|s_{t-r+2}^t, \mathcal{F}_t)\bar{\Psi}(s_{t-r+1}^t)\Delta_{r,t-1}^{\Omega}(s_1^{t-1})\Psi(s_1^t)' \\
&+ \sum_{s^* \neq s_{t-r+1}} p_r(s_{t-r+1} = s^*|s_{t-r+2}^t, \mathcal{F}_t)\left(\bar{\Omega}(s_{t-r+1} = s^*, s_{t-r+2}^t) - \Omega(s_1^t)\right)
\end{aligned}$$

and

$$\begin{aligned}
\Delta_{r,t}^{x}(s_1^t) =& p_r(s_{t-r+1}|s_{t-r+2}^t, \mathcal{F}_t)\left(\bar{x}_{t|t}(s_{t-r+1}^t) - x_{t|t}(s_1^t)\right) \\
&+ \sum_{s^* \neq s_{t-r+1}} p_r(s_{t-r+1} = s^*|s_{t-r+2}^t, \mathcal{F}_t)\left(\bar{x}(s_{t-r+1} = s^*, s_{t-r+2}^t) - x(s_1^t)\right) \\
=& p_r(s_{t-r+1} = s^*|s_{t-r+2}^t, \mathcal{F}_t)[\Psi(s_1^t)\underbrace{\left(\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-1}) - x_{t-1|t-1}(s_1^{t-1})\right)}_{=\Delta_{r,t-1}^{x}(s_1^{t-1})} \\
&+ \left(\bar{\Psi}(s_{t-r+1}^t) - \Psi(s_1^t)\right)\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-1})\Big] \\
&+ \sum_{s^* \neq s_{t-r+1}} p_r(s_{t-r+1} = s^*|s_{t-r+2}^t, \mathcal{F}_t)\left(\bar{x}(s_{t-r+1} = s^*, s_{t-r+2}^t) - x(s_1^t)\right)
\end{aligned}$$

Sequentially applying these expressions and using Lemmas A7 and A8 yield

$$\begin{aligned}
&\|\Delta_{r,t}^{\Omega}(s_1^t)\| \\
\leq& \|\bar{\Psi}^{r-1}(s_{t-r+1}^t)\| \cdot \|\Psi^{r-1}(s_1^t)\| \cdot \|\Delta_{r,t-r+1}^{\Omega}(s_1^{t-r+1})\| \\
&+ \sum_{k=1}^{r-1} \|\bar{\Psi}^{k-1}(s_{t-r+1}^t)\| \cdot \|\Psi^{k-1}(s_1^t)\| \cdot \max_{s^*} \|\Omega(s_{t-r+3-k}^{t-k+1}, s_{t-r+2-k} = s^*, s_1^{t-r+1-k}) - \Omega(s_1^t)\| \\
\leq& 2c_{\Omega}^{+}c_1^2\exp(-2c_2(r-1)) + c_\delta\exp(-2c_2(r-1))\left(1 + \frac{c_1^2\exp(-2c_2)}{1-\exp(-2c_2)}\right) \\
=& c_\Delta\exp(-2c_2(r-1))
\end{aligned}$$

where $c_{\Omega,\Delta} = 2c_\Omega^+ c_1^2 + c_\delta(1 + \frac{c_1^2 \exp(-2c_2)}{1-\exp(-2c_2)})$ and

$$
\begin{aligned}
&\|\Delta^x_{r,t}(s_1^t)\| \\
\leq &\|\Psi^{r-1}(s_1^t)\| \cdot \|\Delta^x_{1,t-r+1}(s_1^{t-r+1})\| + \sum_{k=1}^{r-1} \|\Psi^{k-1}(s_1^t)\| \cdot \|\bar{\Psi}(s_{t-r+1}^{t-k+1}) - \Psi(s_1^{t-k+1})\| \cdot \|\bar{x}_{t-k|t-k}(s_{t-r+1}^{t-k})\| \\
&+ \sum_{k=1}^{r-1} \|\Psi^{k-1}(s_1^t)\| \cdot \max_{s^*} \|x(s_{t-r+2-k} = s^*, s_{t-r+3-k}^{t-k+1}, s_1^{t-r+1-k}) - x(s_1^{t-k+1})\| \\
\leq &2c_x c_1 \exp(-c_2(r-1)) \\
&+ \sum_{k=1}^{r-1} \frac{c_1 c_4}{1-\exp(2c_2)} \exp(-c_2(2r+k-3)) \left[\exp(-c_2(k-2)) - \exp(c_2 k)\right] \left\|\bar{x}_{t-k|t-k}(s_{t-r+1}^{t-k})\right\| \\
&+ \sum_{k=1}^{r-1} c_1 \exp(-c_2(k-1)) \times \exp(-c_2(r-1)) m_{t-k+1} \\
&\leq \exp(-c_2(r-1)) M_{x,t}
\end{aligned}
$$

where

$$
\begin{aligned}
M_{x,t} = &2c_x c_1 + \sum_{k=1}^{r-1} \frac{c_1 c_4}{1-\exp(2c_2)} \left[\exp(-c_2(r+2k-4) - \exp(-c_2(r-2))\right] \|\bar{x}_{t-k|t-k}(s_{t-r+1}^{t-k})\| \\
&+ \sum_{k=1}^{r-1} c_2 \exp(-c_2(k-1)) m_{t-k+1}
\end{aligned}
$$

where $\mathbb{E}_0|M_{x,t}|$ is finite, which implies $M_{x,t} = O_p(1)$ by the Markov's inequality. □

Applying the forecasting step at period $t+1$, we obtain the following corollary.

COROLLARY A1. *Let* $\Delta^\Sigma_{r,t+1}(s_1^{t+1}) = \bar{\Sigma}_{t+1|t}(s_{t-r+2}^{t+1}) - \Sigma_{t+1|t}(s_1^{t+1})$ *and* $\Delta^y_{r,t+1}(s_1^{t+1}) = \bar{y}_{t+1|t}(s_{t-r+2}^{t+1}) - y_{t+1|t}(s_1^{t+1})$. *Then, there exist positive constants* $c_{\Sigma,\Delta}$ *and* c_2 *as well as a positive stochastically bounded random variable* $M_{y,t}$ *such that*

$$
\begin{aligned}
\max_{s_1^{t+1}} \|\Delta^\Sigma_{r,t+1}(s_1^{t+1})\| &\leq c_{\Sigma,\Delta} \exp(-2c_2(r-1)) \\
\max_{s_1^{t+1}} \|\Delta^y_{r,t+1}(s_1^{t+1})\| &\leq M_{y,t} \exp(-c_2(r-1))
\end{aligned}
$$

B.1.3. Per-Period Likelihood

We are now ready to evaluate the difference of the per-period likelihood in the second line of (4), $|\log p_r(y_t|s_{t-r+1}^t, \mathcal{F}_{t-1}) - \log p(y_t|s_1^t, \mathcal{F}_{t-1})|$.

PROPOSITION A2. *There exists a positive stochastically bounded random variable N_t such that*

$$\max_{s_1^t} \left| \log p_r(y_t | s_{t-r+2}^t, \mathcal{F}_{t-1}) - \log p(y_t | s_1^t, \mathcal{F}_{t-1}) \right| \leq \exp(-c_2(r-1)) N_t$$

PROOF. Take any s_1^t. Note that $p_r(y_t|s_{t-r+2}^t, \mathcal{F}_{t-1}) = \phi(y_t; \bar{y}_{t|t-1}(s_{t-r+2}^t), \bar{\Sigma}_{t|t-1}(s_{t-r+2}^t))$ and $p(y_t|s_1^t, \mathcal{F}_{t-1}) = \phi(y_t; y_{t|t-1}(s_1^t), \Sigma_{t|t-1}(s_1^t))$. Then we have

$$\begin{aligned}
&\left|\log p_r(y_t | s_{t-r+2}^t, \mathcal{F}_{t-1}) - \log p(y_t | s_1^t, \mathcal{F}_{t-1})\right| \\
\leq & \frac{1}{2} \left| \log \left(\det \bar{\Sigma}_{t|t-1}(s_{t-r+2}^t) \right) - \log \left(\det \Sigma_{t|t-1}(s_1^t) \right) \right| \\
&+ \frac{1}{2} \left| \left(y_t - \bar{y}_{t|t-1}(s_{t-r+2}^t) \right)' \bar{\Sigma}_{t|t-1}(s_{t-r+2}^t)^{-1} \left(y_t - \bar{y}_{t|t-1}(s_{t-r+2}^t) \right) \right. \\
&\left. - \left(y_t - y_{t|t-1}(s_1^t) \right)' \Sigma_{t|t-1}(s_1^t)^{-1} \left(y_t - y_{t|t-1}(s_1^t) \right) \right|
\end{aligned}$$

Without loss of generality, let $\det \bar{\Sigma}_{t|t-1}(s_{t-r+2}^t) \geq \det \Sigma_{t|t-1}(s_1^t)$. Due to $|\log x - \log y| \leq \frac{|x-y|}{x \wedge y}$ and Lemma A4, the first term can be written as

$$\begin{aligned}
&\left| \log \left(\det \bar{\Sigma}_{t|t-1}(s_{t-r+2}^t) \right) - \log \left(\det \Sigma_{t|t-1}(s_1^t) \right) \right| \\
\leq & \frac{\det \bar{\Sigma}_{t|t-1}(s_{t-r+2}^t) - \det \Sigma_{t|t-1}(s_1^t)}{\det \Sigma_{t|t-1}(s_1^t)} \\
\leq & \left(1 + \left\| \Sigma_{t|t-1}(s_1^t)^{-1} \right\| \cdot \left\| \bar{\Sigma}_{t|t-1}(s_{t-r+2}^t) - \Sigma_{t|t-1}(s_1^t) \right\| \right)^{d_y} - 1 \\
\leq & \left(1 + R_- \left\| \Delta_{r,t}^{\Omega}(s_1^t) \right\| \right)^{d_y} - 1 \\
= & \exp(-c_2(r-1)) \sum_{d=1}^{d_y} \binom{d_y}{d} \left(R_- c_{\Sigma,\Delta} \exp(-c_2(r-1)) \right)^d \\
\leq & \exp(-c_2(r-1)) \sum_{d=1}^{d_y} \binom{d_y}{d} \left(R_- c_{\Sigma,\Delta} \exp(c_2) \right)^d \equiv \exp(-c_2(r-1)) c_5
\end{aligned}$$

We decompose the expression in the second and third lines as

$$
\begin{aligned}
&\left(y_t - \bar{y}_{t|t-1}(s_{t-r+2}^t)\right)' \left(\bar{\Sigma}_{t|t-1}(s_{t-r+2}^t)^{-1} - \Sigma_{t|t-1}(s_1^t)^{-1}\right) \left(y_t - \bar{y}_{t|t-1}(s_{t-r+2}^t)\right) \\
&+ y_t'\Sigma_{t|t-1}(s_1^t)^{-1}\left(-\bar{y}_{t|t-1}(s_{t-r+2}^t) + y_{t|t-1}(s_1^t)\right) \\
&+ \left(-\bar{y}_{t|t-1}(s_{t-r+2}^t) + y_{t|t-1}(s_1^t)\right)' \Sigma_{t|t-1}(s_1^t)^{-1} y_t \\
&+ \left(\bar{y}_{t|t-1}(s_{t-r+2}^t) - y_{t|t-1}(s_1^t)\right)' \Sigma_{t|t-1}(s_1^t)^{-1} y_{t|t-1}(s_1^t) \\
&+ \bar{y}_{t|t-1}(s_{t-r+2}^t)'\Sigma_{t|t-1}(s_1^t)^{-1}\left(\bar{y}_{t|t-1}(s_{t-r+2}^t) - y_{t|t-1}(s_1^t)\right)
\end{aligned}
$$

The second and third terms are bounded by $\exp(-c_2(r-1))M_{x,t}\|y_t\|$ where $M_{x,t}\|y_t\| = O_p(1)$ as the product of two $O_p(1)$ terms. Similarly, the fourth and fifth terms are bounded by $\exp(-c_2(r-1))R_- M_{x,t}\|y_{t|t-1}\|$ where $M_{x,t}\|y_{t|t-1}\| = O_p(1)$ due to Lemma A2. Regarding the first term, note that

$$
\bar{\Sigma}_{t|t-1}(s_{t-r+2}^t)^{-1} - \Sigma_{t|t-1}(s_1^t)^{-1} = \bar{\Sigma}_{t|t-1}(s_{t-r+2}^t)^{-1}\left(\Sigma_{t|t-1}(s_1^t) - \bar{\Sigma}_{t|t-1}(s_{t-r+2}^t)\right)\Sigma_{t|t-1}(s_1^t)^{-1}
$$

and then

$$
\begin{aligned}
&\left|\left(y_t - \bar{y}_{t|t-1}(s_{t-r+2}^t)\right)' \left(\bar{\Sigma}_{t|t-1}(s_{t-r+2}^t)^{-1} - \Sigma_{t|t-1}(s_1^t)^{-1}\right) \left(y_t - \bar{y}_{t|t-1}(s_{t-r+2}^t)\right)\right| \\
&\leq \exp(-c_2(r-1)) c_{\Sigma,\Delta} R_-^2 \|y_t - \bar{y}_{t|t-1}(s_{t-r+2}^t)\|^2
\end{aligned}
$$

where $\|y_t - \bar{y}_{t|t-1}(s_{t-r+2}^t)\|^2 = O_p(1)$. Hence, we have

$$
\left|\log \bar{p}(y_t|y_1^{t-1}, s_{t-r+2}^t) - \log p(y_t|y_1^{t-1}, s_1^t)\right| \leq \exp(-c_2(r-1))N_t
$$

where $N_t = O_p(1)$. □

B.1.4. Transition Probability

We consider the difference between the log transition probabilities obtained from the exact and approximated filters.

$$p_r(s_t=0|s_{t-1}=0,s_{t-r+1}^{t-2},F_{t-1})$$
$$=\frac{\Phi\left(-\tau\iota;\Lambda\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2}),I+\Lambda\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\Lambda'\right)}{\Phi\left(-\tau;\lambda'\left(\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x)},1+\lambda'\left(\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x,d_x+1:2d_x)}\lambda\right)}$$

$$p(s_t=0|s_{t-1}=0,s_1^{t-2},F_{t-1})$$
$$=\frac{\Phi\left(-\tau\iota;\Lambda x_{t-1|t-1}(s_1^{t-2}),I+\Lambda\Omega_{t-1|t-1}(s_1^{t-2})\Lambda'\right)}{\Phi\left(-\tau;\lambda'\left(x_{t-1|t-1}(s_1^{t-2})\right)_{(d_x+1:2d_x)},1+\lambda'\left(\Omega_{t-1|t-1}(s_1^{t-2})\right)_{(d_x+1:2d_x,d_x+1:2d_x)}\lambda\right)}$$

PROPOSITION A3. *There exists a positive stochastically bounded random variable $N_{TP,t}$ such that*

$$\max_{s_1^t}|p_r(s_t|s_{t-r+1}^{t-1},\mathcal{F}_{t-1})-p(s_t|s_1^{t-1},\mathcal{F}_{t-1})|\leq N_{TP,t}\exp(-c_2(r-1))$$

PROOF. Consider $s_t=0$ and $s_{t-1}=0$. The probabilities for other combinations of (s_t,s_{t-1}) follow similarly. Define $f:\mathbb{R}^2\times\mathbb{R}^3\to[0,1]$ as

$$\begin{aligned}f(x,\text{vech}(L))=&\frac{\Phi_2\left(-\tau\iota;x,LL'\right)}{\Phi_1\left(-\tau;x_2,(LL')_{(2,2)}\right)}\\=&\frac{\Phi_2\left(L^{-1}(-\tau\iota-x)\right)}{\Phi_1\left(((LL')_{(2,2)})^{-1/2}(-\tau-x_2)\right)}\end{aligned}$$

where $x=[x_1,x_2]'$ and L is a 2×2 lower triangular matrix. It obviously follows that

$$f\left(\Lambda\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2}),\text{vech}\left(\left(I+\Lambda\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\Lambda'\right)^{1/2}\right)\right)=p_r(s_t=0|s_{t-1}=0,s_{t-r+1}^{t-2},\mathcal{F}_{t-1})$$

and

$$f\left(\Lambda\tilde{x}_{t-1|t-1}(s_1^{t-2}),\text{vech}\left(\left(I+\Lambda\tilde{\Omega}_{t-1|t-1}(s_1^{t-2})\Lambda'\right)^{1/2}\right)\right)=p(s_t=0|s_{t-1}=0,s_1^{t-2},\mathcal{F}_{t-1})$$

Let $\mu_{t-1|t-1}(s_1^{t-2})=\Lambda x_{t-1|t-1}(s_1^{t-2})$ and $V_{t-1|t-1}(s_1^{t-2})=I+\Lambda\Omega(s_1^{t-2})\Lambda'$, and define $\bar{\mu}_{t-1|t-1}(s_{t-r+1}^{t-2})$ and $\bar{V}_{t-1|t-1}(s_{t-r+1}^{t-2})$ similarly. We apply the mean value theorem to f around $(\mu_{t-1|t-1}(s_1^{t-2}), V_{t-1|t-1}(s_1^{t-2}))$. There exists a constant $\alpha \in (0,1)$ such that

$$
\begin{aligned}
& p_r(s_t = 0|s_{t-1} = 0, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}) = p(s_t = 0|s_{t-1} = 0, s_1^{t-2}, \mathcal{F}_{t-1}) \\
& + \left(\nabla f(\mu^\dagger_{t-1|t-1}(s_1^{t-2}), \mathrm{vech}(V^\dagger(s_1^{t-2})^{1/2}))\right)' \begin{bmatrix} \bar{\mu}_{t-1|t-1}(s_{t-r+1}^{t-2}) - \mu_{t-1|t-1}(s_1^{t-2}) \\ \mathrm{vech}(\bar{V}_{t-1|t-1}(s_{t-r+1}^{t-2})^{1/2}) - \mathrm{vech}(V_{t-1|t-1}(s_1^{t-2})^{1/2}) \end{bmatrix}
\end{aligned} \tag{A17}
$$

where $\mu^\dagger_{t-1|t-1}(\cdot) = \alpha\bar{\mu}_{t-1|t-1}(\cdot) + (1-\alpha)\mu_{t-1|t-1}(\cdot)$ and $V^\dagger_{t-1|t-1}(\cdot)^{1/2} = \alpha\bar{V}_{t-1|t-1}(\cdot)^{1/2} + (1-\alpha)V_{t-1|t-1}(\cdot)^{1/2}$.

Let $L = [\ell_{ij}]_{i,j=1,2}$ with $\ell_{12} = 0$. Note that

$$
\begin{aligned}
\Phi_2\left(L^{-1}(-\tau\iota - x)\right) =& \Phi_1\left(-\frac{1}{\ell_{11}}(\tau + x_1)\right)\Phi_1\left(\frac{\ell_{21}}{\ell_{11}\ell_{22}}(\tau + x_1) - \frac{1}{\ell_{22}}(\tau + x_2)\right) \\
:=& \Phi_1(z_1)\,\Phi_1(z_2)
\end{aligned}
$$

and

$$
\begin{aligned}
\Phi_1\left(((LL')_{(2,2)})^{-1/2}(-\tau - x_2)\right) =& \Phi_1\left(-\frac{1}{\ell_{21}^2 + \ell_{22}^2}(\tau + x_2)\right) \\
:=& \Phi_1(z_3)
\end{aligned}
$$

The derivative of f with respect to x_1 is given by

$$
\frac{\partial}{\partial x_1} f(x, \mathrm{vech}(L)) = \frac{1}{(\Phi_1(z_3))^2}\left(\Phi_1(z_2)\,\phi_1(z_1) \times \frac{d}{dx_1}z_1 + \Phi_1(z_1)\,\phi_1(z_2) \times \frac{d}{dx_1}z_2\right)
$$

Since z_3 is $O_p(1)$ for any α, so is $(\Phi_1(z_3))^2$. The derivatives $\frac{d}{dx_1}z_1$ and $\frac{d}{dx_1}z_2$ are also $O_p(1)$. Hence the whole expression is $O_p(1)$. The remaining derivatives can be shown to be $O_p(1)$ similarly. Therefore, the gradient in (A17) is $O_p(1)$.

Using Proposition A1, it follows that

$$
\begin{aligned}
\max_{s_1^{t-2}} \left\|\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2}) - \Omega_{t-1|t-1}(s_1^{t-2})\right\| &\le c_{\Omega,\Delta,2}\exp(-2c_2(r-1)) \\
\max_{s_1^{t-2}} \left\|\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2}) - x_{t-1|t-1}(s_1^{t-2})\right\| &\le M_{x,t,2}\exp(-c_2(r-1))
\end{aligned}
$$

where $c_{\Omega,\Delta,2}$ is a positive constant and $M_{x,t,2}$ is a positive and stocastically bounded random variable.

We can deduce

$$
\begin{aligned}
&\left\|\bar{\mu}_{t-1|t-1}(s_{t-r+1}^{t-2}) - \mu_{t-1|t-1}(s_1^{t-2})\right\|_1 \\
\leq &\sqrt{2}\left\|\bar{\mu}_{t-1|t-1}(s_{t-r+1}^{t-2}) - \mu_{t-1|t-1}(s_1^{t-2})\right\|_2 \\
\leq &\sqrt{2}\|\Lambda\|\left\|\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2}) - x_{t-1|t-1}(s_1^{t-2})\right\|_2
\end{aligned}
$$

where we have used $\|a\|_1 \leq \sqrt{n}\|a\|_2$ holding for an $n \times 1$ vector a. Furthermore,

$$
\begin{aligned}
&\|\text{vech}\left(\bar{V}_{t-1|t-1}(s_{t-r+1}^{t-2})^{1/2}\right) - \text{vech}\left(V_{t-1|t-1}(s_1^{t-2})^{1/2}\right)\|_1 \\
=&\|\bar{V}_{t-1|t-1}(s_{t-r+1}^{t-2})^{1/2} - V_{t-1|t-1}(s_1^{t-2})^{1/2}\|_1 \\
\leq &2\|\bar{V}_{t-1|t-1}(s_{t-r+1}^{t-2})^{1/2} - V_{t-1|t-1}(s_1^{t-2})^{1/2}\|_F \\
\leq &2\|\bar{V}_{t-1|t-1}(s_{t-r+1}^{t-2})^{1/2} - V_{t-1|t-1}(s_1^{t-2})^{1/2}\| \\
\leq &2\|\bar{V}_{t-1|t-1}(s_{t-r+1}^{t-2}) - V_{t-1|t-1}(s_1^{t-2})\|^{1/2} \\
\leq &2\|\Lambda\|\left\|\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2}) - \Omega_{t-1|t-1}(s_1^{t-2})\right\|^{1/2}
\end{aligned}
$$

where $\|\cdot\|_1$, $\|\cdot\|_F$, and $\|\cdot\|$ stand for the element-wise 1-norm, the Frobenius norm, and the spectral norm of matrix respectively.

By $\log(1+x) \leq x$ for $x > 0$,

$$
\begin{aligned}
&|\log p_r(s_t = 0|s_{t-1} = 0, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}) - \log p(s_t = 0|s_{t-1} = 0, s_1^{t-2}, \mathcal{F}_{t-1})| \\
\leq &\left|\frac{\left(\nabla f(\mu_{t-1|t-1}^{\dagger}(s_1^{t-2}), \text{vech}(V^{\dagger}(s_1^{t-2})^{1/2}))\right)' \begin{bmatrix} \bar{\mu}_{t-1|t-1}(s_{t-r+1}^{t-2}) - \mu_{t-1|t-1}(s_1^{t-2}) \\ \text{vech}(\bar{V}_{t-1|t-1}(s_{t-r+1}^{t-2})^{1/2}) - \text{vech}(V_{t-1|t-1}(s_1^{t-2})^{1/2}) \end{bmatrix}}{p(s_t = 0|s_{t-1} = 0, s_1^{t-2}, \mathcal{F}_{t-1})}\right. \\
=&O_p(1)\left[\left\|\bar{\mu}_{t-1|t-1}(s_{t-r+1}^{t-2}) - \mu_{t-1|t-1}(s_1^{t-2})\right\|_1 + \|\text{vech}\left(\bar{V}_{t-1|t-1}(s_{t-r+1}^{t-2})^{1/2}\right) - \text{vech}\left(V_{t-1|t-1}(s_1^{t-2})^{1/2}\right.\right. \\
\leq &O_p(1)\exp\left(-c_2(r-1)\right)
\end{aligned}
$$

□

B.1.5. Proof of Proposition 1

By (4), we have

$$
\begin{aligned}
&|\log p_{r,\theta}(y_1^T|x_1=\tilde{x},s_1=\tilde{s})-\log p_\theta(y_1^T|x_1=\tilde{x},s_1=\tilde{s})| \\
&\le \sum_{t=r+1}^{T}\left(\max_{s_2^t}\left|\log p_r(y_t|y_1^{t-1},s_{t-r+1}^{t})-\log p(y_t|y_1^{t-1},s_1^t)\right|\right) \\
&\quad+\sum_{t=r+1}^{T}\left(\max_{s_2^t}\left|\log p_r(s_t|s_{t-r+1}^{t-1},y_1^{t-1})-\log p(s_t|s_1^{t-1},y_1^{t-1})\right|\right)
\end{aligned}
$$

By Proposition A2, the term inside the first summation is bounded by $\exp(-c_2(r-1))N_t$. By Proposition A3, the term inside the second summation is bounded by $\exp(-c_2(r-1))N_{TP,t}$.

B.2. Proofs for Section 4

We define a k-local Doeblin set following Douc and Moulines (2012).

DEFINITION A1. *Take $k \in \mathbb{Z}$. A set $C \in \mathcal{X}$ is a k-local Doeblin set if there exist two positive functions $\varepsilon_C^- : (\mathbb{R}^{d_y})^k \to \mathbb{R}^+$ and $\varepsilon_C^+ : (\mathbb{R}^{d_y})^k \to \mathbb{R}^+$, a family of probability measures $\{\lambda_C^\theta\langle z\rangle\}_{\theta\in\Theta, z\in(\mathbb{R}^{d_y})^k}$, and a family of positive functions $\{\varphi_C^\theta\langle z\rangle\}_{\theta\in\Theta, z\in(\mathbb{R}^{d_y})^k}$ such that for any $\theta \in \Theta$ and $z \in (\mathbb{R}^{d_y})^k$, we have $\lambda_C^\theta\langle z\rangle(C) = 1$ and for any $A \in \mathcal{X}$ and $\xi \in C$, we have*

$$
\varepsilon_C^-(z)\varphi_C^\theta\langle z\rangle(\xi)\lambda_C^\theta\langle z\rangle(A) \le \mathbf{L}\langle z\rangle(\xi, A\cap C) \le \varepsilon_C^+(z)\varphi_C^\theta\langle z\rangle(\xi)\lambda_C^\theta\langle z\rangle(A) \tag{A18}
$$

where $\mathbf{L}\langle y_1^k\rangle(\xi_1, A) = \int\cdots\int \left[\prod_{i=1}^k \phi\left(y_i; B_{s_i}R_iR_i'\right) Q^\theta(d\xi_{i+1}|\xi_i, y_i)\right] \mathbf{1}_A(\xi_{k+1})$ is the transition kernel of (ξ_t) given the observations y_1^k.

The following lemma verifies Assumption (A1) by Douc and Moulines (2012).

LEMMA A9. *Assume Assumptions A1 and 4. There exists $k \in \mathbb{Z}$ and a set $K \in (\mathcal{B}(\mathbb{R}^{d_y}))^k$ with the following properties.*

(i) $\mathbb{P}\left[Y_1^k \in K\right] > 2/3$

(ii) Take any $\eta > 0$. There exists a r-local Doeblin set $C \in \mathcal{B}(\mathbb{R}^{d_x}) \times \sigma(\mathcal{S})$ such that for any $\theta \in \Theta$ and $y_1^k \in K$, we have

$$
\sup_{\xi_1\in C^C} p^\theta_{\delta_{\xi_1}}(y_1^k) \le \eta \sup_{\xi_1\in\mathbb{R}^{d_x}\times\mathcal{S}} p^\theta_{\delta_{\xi_0}}(y_1^k) < \infty \tag{A19}
$$

where δ_{ξ_1} is the Dirac measure concentrated at ξ_1, and

$$\inf_{y_1^k \in K} \frac{\varepsilon_C^-(y_1^k)}{\varepsilon_C^+(y_1^k)} > 0 \tag{A20}$$

(iii) There exists a set $D \in \mathcal{B}(\mathbb{R}^{d_x}) \times \sigma(\mathcal{S})$ such that

$$\mathbb{E}\left[\log^- \inf_{\theta \in \Theta} \inf_{\xi \in D} \mathbf{L}\langle Y_1^k \rangle(\xi, D)\right] < \infty$$

PROOF. Take a set K so that $\mathbb{P}\left[Y_1^k \in K\right] > 2/3$. For any compact set $C_1 \in \mathbb{R}^{d_x}$ and $C_2 \subset \mathcal{S}$, we will show that a compact set $C = C_1 \times C_2$ is k-local Doeblin. Take any $A \in \mathcal{X}$. Define the measure $\lambda_C(A) = \frac{\mu(A \cap C)}{\mu(C)}$ where μ denotes the Lebesgue measure. Let $\varphi^\theta \langle y_1^k \rangle(\xi) = 1$. Let

$$\varepsilon_C^- \langle y_1^k \rangle = \prod_{i=1}^k \inf_{\theta \in \Theta} \min_{(x',s)' \in C} \phi(y_i; B_s x, R_s R_s') \times \inf_{\theta \in \Theta} \min_{\xi_1, \cdots, \xi_{k+1} \in C} \prod_{i=1}^k Q^\theta(\xi_{i+1} | \xi_i, y_i)$$
$$\varepsilon_C^+ \langle y_1^k \rangle = \prod_{i=1}^k \sup_{\theta \in \Theta} \sup_{(x',s)' \in \mathbf{X}} \phi(y_i; B_s x, R_s R_s') \times \sup_{\theta \in \Theta} \sup_{\xi_1, \cdots, \xi_{r+1} \in \mathbf{X}} \prod_{i=1}^k Q^\theta(\xi_{i+1} | \xi_i, y_i)$$

By Assumption 4 and the fact that the normal density is positive and bounded, $\varepsilon_C^- \langle y_1^k \rangle$ and $\varepsilon_C^+ \langle y_1^k \rangle$ are positive and bounded for any y_1^k. Under these settings, (A18) is satisfied for any $\xi \in C$ and $A \in \mathcal{X}$. Hence, C is a r-local Doeblin set. Note that

$$\begin{aligned} \lim_{|\xi| \to \infty} \sup_{y_1^k \in K} p_{\delta_\xi}^\theta(y_1^k) &= \lim_{|\xi| \to \infty} \sup_{y_1^k \in K} \mathbf{L}\langle y_1^k \rangle(\xi, \mathbf{X}) \\ &= 0 \end{aligned}$$

because the normal density is bounded and q^θ is bounded by 1. Since the choice of a compact set C_1 above was arbitrary, (A19) holds by taking C_1 to be sufficiently large. Since ε^- and ε^+ are positive and bounded, (A20) also holds.

To establish claim (iii), take any compact set $D \in \mathcal{X}$. Define

$$F \equiv \{y_1^k \in \left(\mathbb{R}^{d_y}\right)^k \mid -\log \inf_{\theta \in \Theta} \inf_{\xi \in D} \mathbf{L}\langle y_1^k \rangle(\xi, D) \geq 0\}$$

Then we have

$$\begin{aligned}
\mathbb{E}\left[\log^- \inf_{\theta\in\Theta}\inf_{\xi\in D}\mathbf{L}\langle Y_1^k\rangle(\xi,D)\right] &= \mathbb{E}\left[\max\left\{-\log\inf_{\theta\in\Theta}\inf_{\xi\in D}\mathbf{L}\langle Y_1^k\rangle(\xi,D),0\right\}\right] \\
&= -\int_F \log\inf_{\theta\in\Theta}\inf_{\xi\in D}\mathbf{L}\langle Y_1^k\rangle(\xi,D)\mathbb{P}(dY_1^k) \\
&\leq -\int_F \log\left(\varepsilon_D^-\langle Y_1^k\rangle\lambda_D(D)\right)\mathbb{P}(dY_1^k) \\
&< \infty
\end{aligned}$$

where the inequality on the third line comes from (A18) and the last inequality follows because the terms inside the bracket on the third line are all positive and bounded for any Y_1^k. $\square$

LEMMA A10. *Assume Assumption 5. For any $x \in \mathbf{X}$, the function $\theta \mapsto p^\theta_{\delta_x}(Y_1^T)$ is continuous on Θ.*

PROOF. It follows from the continuity of the function $\theta \mapsto Q^\theta$ and of $\phi(y_t; B_{s_t}x_t, Q_{s_t}Q'_{s_t})$ with respect to θ. $\square$

Proof of Proposition 2

PROOF. The proposition follows from Proposition 1 and Theorem 2 in Douc and Moulines (2012). Note that their propositions continue to hold when the transition kernel of x_t depends on y_t, i.e., replacing $Q^\theta(x_i, dx_{i+1})$ with $Q^\theta(d\xi_{i+1}|\xi_i, y_i)$ in their equation (3). Thus, what we have to check is whether assumptions (A1)–(A3) hold.

(A1) and (A3) follow from Lemmas A9 and A10 respectively. (A2) (i) holds because the normal density is positive. (A2) (ii) also holds since the normal density is bounded above. $\square$

Appendix C. Empirical Application

C.1. Model Description

This section describes the model used in the empirical application. The model is borrowed from Bianchi and Ilut (2017) which studies the monetary/fiscal policy mix in the post-WWII U.S. The economy consists of the infinitely lived representative household, firms subject to monopolistic competition and nominal price rigidity, and the government operating monetary and fiscal policies. There are two binary variables s_t^{pol} and s_t^{vol}

which govern the policy stance and economic volatility respectively. The model here is slightly different from Bianchi and Ilut (2017) since we do not include the AM/AF regime because Bianchi and Ilut (2017) found the periods in which the government took this policy stance very short. The shocks ε_t^x ($x \in \{d, \mu, a, tp, e^L, e^S, \chi, \tau, R\}$) are the standard Gaussian random variables.

C.1.1. Household

The representative household chooses the stream of consumption, labor supply, and bond holdings to maximize

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \exp(d_t) \left[\log(C_t - \Phi C_{t-1}^A) - h_t\right]$$

subject to

$$P_t C_t + P_t^m B_t^m + R_t^{-1} B_t^s = P_t W_t h_t + B_{t-1}^s (1 + \rho P_t^m) B_{t-1}^m + P_t D_t - T_t + TR_t$$

where C_t is consumption, h_t is labor supply, P_t is aggregate price level, W_t is real wage, D_t is dividend income from the firms, and T_t and TR_t are lump-sum tax and transfer respectively. The preference includes external habit formation where C_{t-1}^a is the average level of consumption at the last period. The term d_t represents the preference shock following

$$d_t = \rho_d d_{t-1} + \sigma_d(s_t^{vol})\varepsilon_t^d$$

The household can hold two types of assets, short-term government bond B_t^s with return R_t and long-term government debt B_t^m. The parameter ρ governs the average maturity of the government debt.

C.1.2. Firms

There is a continuum of firms indexed by $j \in [0, 1]$. Facing monopolistic competition as well as the Rotemberg-type nominal price rigidity, they choose price $P_t(j)$ to maximize the present value of profits

$$\mathbb{E}_0 \sum_{t=0}^{\infty} Q_t \left[\left(\frac{P_t(j)}{P_t} \right) Y_t(j) - W_t h_t(j) - AC_t(j) \right]$$

subject to the demand curve

$$Y_t(j) = \left(\frac{P_t(j)}{P_t}\right)^{-1/\nu_t} Y_t$$

and the quadratic price adjustment cost proportional to the real output:

$$AC_t(j) = 0.5\varphi \left(\frac{P_t(j)}{P_{t-1}(j)} - \Pi_{t-1}^{\varsigma}\Pi^{1-\varsigma}\right)^2 \frac{Y_t(j)P_t(j)}{P_t}$$

where $\Pi_t = P_t/P_{t-1}$ and Π is its steady state value. The term ν_t is the inverse of the elasticity of substitution connected with the markup shock $\aleph_t = 1/(1-\nu_t)$. The rescaled markup shock $\mu_t = \frac{\kappa}{1+\varsigma\beta}\log(\aleph_t/\aleph)$ with $\kappa = \frac{1-\nu}{\nu\varphi\Pi^2}$ follows the exogenous process $\mu_t = \rho_\mu\mu_{t-1} + \sigma_\mu(s_t^{vol})\varepsilon_t^\mu$. The sum of profits is discounted by the stochastic discount factor Q_t. The production function is given by $Y_t(j) = A_t h_t(j)^{1-\alpha}$ with $\alpha \in [0,1]$. The total factor productivity (TFP) A_t evolves as $\log(A_t/A_{t-1}) = \gamma + a_t$ where $a_t = \rho_a a_{t-1} + \sigma_a(s_t^{vol})\varepsilon_t^a$.

C.1.3. Fiscal Policy

The short-term government debt is assumed to have zero net supply. The intratemporal government budget constraint is written as

$$P_t^m B_t^m = B_{t-1}^m(1+\rho P_t^m) - T_t + E_t + TP_t$$

where $E_t = P_t G_t + TR_t$ is the total government expenditure which is the sum of nominal government spending and transfer payment. The last term on the right hand side, TP_t, is the residual term which is necessary to avoid computational issues due to the fact that we are using observations to characterize debt, tax, and expenditure. We divide the government budget constraint by the nominal output $P_t Y_t$ to have

$$b_t^m = \frac{b_{t-1}^m R_{t-1,t}^m}{\Pi_t Y_t/Y_{t-1}} - \tau_t + e_t + tp_t$$

where $x_t = X_t/P_t Y_t$ for any variable X_t and $R_{t-1}^m, t = (1+\rho P_t^m)/P_{t-1}^m$. We assume $tp_t = \rho_{tp} tp_{t-1} + \sigma_{tp}(s_t^{vol})\varepsilon_t^{tp}$. The expenditure is assumed to be decomposed by short-term and long-term components: $\tilde{e}_t = \tilde{e}_t^L + \tilde{e}_t^S$[19] where

$$\tilde{e}_t^L = \rho_{e^L}\tilde{e}_{t-1}^L + \sigma_{e^L}(s_t^{vol})\varepsilon_t^{e^L}$$
$$\tilde{e}_t^S = \rho_{e^S}\tilde{e}_{t-1}^S + (1-\rho_{e^S})\phi_y\left(\hat{y}_t - \hat{y}_t^*\right) + \sigma_{e^S}(s_t^{vol})\varepsilon_t^{e^S}$$

where y_t^* is (detrended) potential output in the absence of price rigidity. The fraction of government spending to total expenditure, $\chi_t = P_t G_t / E_t$, is assumed to follow

$$\tilde{\chi}_t = \rho_\chi \tilde{\chi}_{t-1} + (1-\rho_\chi)\iota_y\left(\hat{y}_t - \hat{y}_t^*\right) + \sigma_\chi(s_t^{vol})\varepsilon_t^\chi$$

We specify the tax rule using the regime-switching coefficients.

$$\tilde{\tau}_t = \rho_\tau(s_t^{pol})\tilde{\tau}_{t-1} + \left(1-\rho_\tau(s_t^{pol})\right)\left[\delta_b(s_t^{pol})\tilde{b}_{t-1}^m + \delta_e\tilde{e}_t + \delta_y\left(\hat{y}_t - \hat{y}_t^*\right)\right] + \sigma_\tau(s_t^{vol})\varepsilon_t^\tau$$

The coefficient on government debt, $\delta_b(s_t^{pol})$ is one of the most important parameters in the model which governs the responsiveness of fiscal policy to the increase of government debt.

C.1.4. Monetary Policy

The monetary authority sets a nominal interest rate R_t based on the Taylor rule with the regime-switching parameters.

$$\frac{R_t}{R} = \left(\frac{R_{t-1}}{R}\right)^{\rho_R(s_t^{pol})}\left[\left(\frac{\Pi_t}{\Pi}\right)^{\psi_\pi(s_t^{pol})}\left(\frac{Y_t}{Y_t^*}\right)^{\psi_y(s_t^{pol})}\right]^{1-\rho_R(s_t^{pol})}\exp\left(\sigma_R(s_t^{vol})\varepsilon_t^R\right)$$

The parameter $\psi_\pi(s_t^{pol})$ determines the strength of nominal interest rate adjustment when observing a rise in the inflation rate.

C.1.5. Market Clearing

The final good market clearing requires

$$Y_t = C_t + G_t$$

[19]We denote $\hat{x}_t = \log((X_t/A_t)/(X/A))$, percentage deviation of the detrended variable from the steady state. For variables normalized by nominal GDP, we denote $\tilde{x}_t = x_t - x$. For the other variables, let $\tilde{x}_t = \log(X_t/X)$.

C.1.6. Regime Switching

As described in the main text, regime shifts for volatility captured by s_t^{pol} occur following a time-invariant transition probability matrix P^{vol}.

$$P^{vol} = \begin{bmatrix} 1 - p_{1,2}^{vol} & p_{1,2}^{vol} \\ p_{2,1}^{vol} & 1 - p_{2,1}^{vol} \end{bmatrix}$$

The policy regime indicator evolves based on our baseline regime rule.

$$s_t^{pol} = \begin{cases} AM/PF & \text{if } \tau^{pol} + \lambda_y \left(\hat{y}_{t-1} - \hat{y}_{t-1}^{*}\right) + \lambda_\pi \tilde{\pi}_{t-1} + \lambda_R \tilde{R}_{t-1} + \lambda_b \tilde{b}_{t-1} + \eta_t \geq 0 \\ PM/AF & \text{otherwise} \end{cases}$$

where $\eta_t = \rho_\eta \eta_{t-1} + \varepsilon_{\eta,t}$, $\varepsilon_{\eta,t} \sim N(0,1)$. As shown by Chang et al. (2017), restricting all λ to be zero implies the traditional Hamilton (1989) regime-switching structure with time-invariant transition probabilities. I estimate the model with this restriction as well, and label it "the exogenous switching model".

C.2. Solving Model

The equilibrium conditions from the model above are summarized as

$$\mathbb{E}_t F_{s_t}\left(x_{t-1}, x_t, x_{t+1}\right) = 0$$

where x_t is a collection of the variables in the model and ε_t is a collection of the structural shocks. To find the solution of the form

$$x_t = g_{s_t}(x_{t-1})$$

the perturbation method proposed by Maih and Waggoner (2018) is employed. Adding a perturbation parameter χ into the system, we seek to find

$$x_t = g_{s_t}(x_{t-1}; \chi)$$

which satisfies

$$\mathbb{E}_t \left[\sum_{j=1}^{J} p_{ij}(x_t; \chi) F_i \left(g_j \left(h_i(x_{t-1}; \chi); \chi \right), g_i(x_{t-1}; \chi), x_{t-1} \right) \right] = 0$$

where $p_{i,j}(\cdot)$ is the transition probability from $s_t = i$ to $s_{t+1} = j$, and $h_i(\cdot)$ is the perturbed policy function. This framework reduces to the original system when $\chi = 1$, and the steady state when $\chi = 0$. Let $x_i = g_i(x_i; 0)$. Maih and Waggoner (2018) choose those functions to be

$$p_{ij}(x_t; \chi) = \begin{cases} \chi p_{ij}(x_t) & \text{if } i \neq j \\ \chi \left(p_{ij}(x_t) - 1 \right) + 1 & \text{otherwise} \end{cases}$$

and

$$h_i(x_{t-1}; \chi) = g_i\left(x_{t-1}; \chi\right) + (1 - \chi)\left(x_j - x_i\right)$$

This choice implies $F_i(x_i, x_i, x_i) = 0$, and thus x_i can be interpreted as the deterministic steady state. Having these two functions, the standard perturbation method applies and we can find the approximated solution. The policy function from the first order perturbation does not exhibit feedback coefficients, while they appear in the regime transition probabilities.

C.3. Sequential Monte Carlo Algorithm

This section lays out the sequential Monte Carlo algorithm to infer posterior distributions used in Section 6. See Herbst and Schorfheide (2014) and Herbst and Schorfheide (2016) for detailed description.

C.3.1. Algorithm Overview

(1) Initialization: Draw the particles $\theta_0^i \sim p(\theta)$ where $p(\theta)$ is the prior distribution, and set $W_0^i = 1$ for $i = 1, \cdots, N$. Alternatively, the initial particles can be drawn from a proposal distribution $g(\theta)$. In this case, incremental weights should be adjusted by $p(\theta)/g(\theta)$.

(2) For $n = 1, \cdots, N_\phi$,

  (a) Correction step: Define

$$\tilde{w}_n^i = \left[p(Y|\theta_{n-1}^i) \right]^{\phi_n - \phi_{n-1}}, \quad i = 1, \cdots, N$$

and we normalize this weight by

$$\tilde{W}_n^i = \frac{\tilde{w}_n^i W_{n-1}^i}{N^{-1}\sum_{j=1}^N \tilde{w}_n^j W_{n-1}^j}, \quad i = 1, \cdots, N$$

(b) Selection step: Calculate

$$ESS_n = \frac{N}{N^{-1}\sum_{i=1}^N \left(\tilde{W}_i^n\right)^2}$$

- When $ESS_n < N/2$, resample the particles from the multinomial distribution characterized by the particles $\{\theta_{n-1}^i\}_{i=1}^N$ with the associated weights $\{\tilde{W}_n^i\}_{i=1}^N$. We define $\{\hat{\theta}_n^i\}_{i=1}^N$ to be N draws of particles from the multinomial distribution described above. Let $W_n^i = 1$ for any $i = 1, \cdots, N$.
- Otherwise, let $\hat{\theta}_n^i = \theta_{n-1}^i$ and $W_n^i = \tilde{W}_n^i$, $i = 1, \cdots, N$.

(c) Mutation step: Compute mean θ_n^* and variance Σ_n^* of the distribution characterized by the particles $\{\theta_{n-1}^i, W_n^i\}_{i=1}^N$. Let

$$c_n = c_{n-1} f\left(1 - R_{n-1}\right), \quad f(x) = 0.95 + 0.10\frac{\exp(16(x-0.25))}{1+\exp(16(x-0.25))}$$

where R_{n-1} is the rejection rate at the previous stage. Generate the random partition of the parameters $\{\theta_{n,b}\}_{b=1}^{N_{blocks}}$. For any $i = 1, \cdots, N$, run the block Metropolis-Hastings algorithm for N_{MH} times using the proposal distribution

$$\begin{aligned}\theta_b|\theta_{n,b,m-1}^i, \theta_{n,-b,m}^i, \theta_{n,b}^*, \Sigma_{n,b}^* \sim & \omega N\left(\theta_{n,b,m-1}^i, c_n^2\Sigma_{n,b}^*\right) \\ & + \frac{1-\omega}{2} N\left(\theta_{n,b,m-1}^i, c_n^2 \text{diag}(\Sigma_{n,b}^*)\right) \\ & + \frac{1-\omega}{2} N\left(\theta_{n,b}^*, c_n^2\Sigma_{n,b}^*\right)\end{aligned}$$

where $\theta_{n,b,m-1}^i$ is the parameter from the previous iteration of the MH, $\theta_{n,-b,m}^i$ is the parameter outside the block b, and $\theta_{n,b}^*$ and $\Sigma_{n,b}^*$ are the partition of θ_n^* and Σ_n^* based on the block b respectively. This gives us the new particle θ_n^i.

C.3.2. Hyperparameters

The hyperparameters we have to choose a priori are $(N, N_\phi, N_{blocks}, N_{MH}, \omega, \{\phi_n\}_{n=0}^{N_\phi})$. We set $N = 6,000$, $N_\phi = 250$, $N_{blocks} = 3$, $N_{MH} = 1$, $\omega = 0.1$, and $\phi_n = \left(n/N_\phi\right)^\lambda$ where $\lambda = 2$.

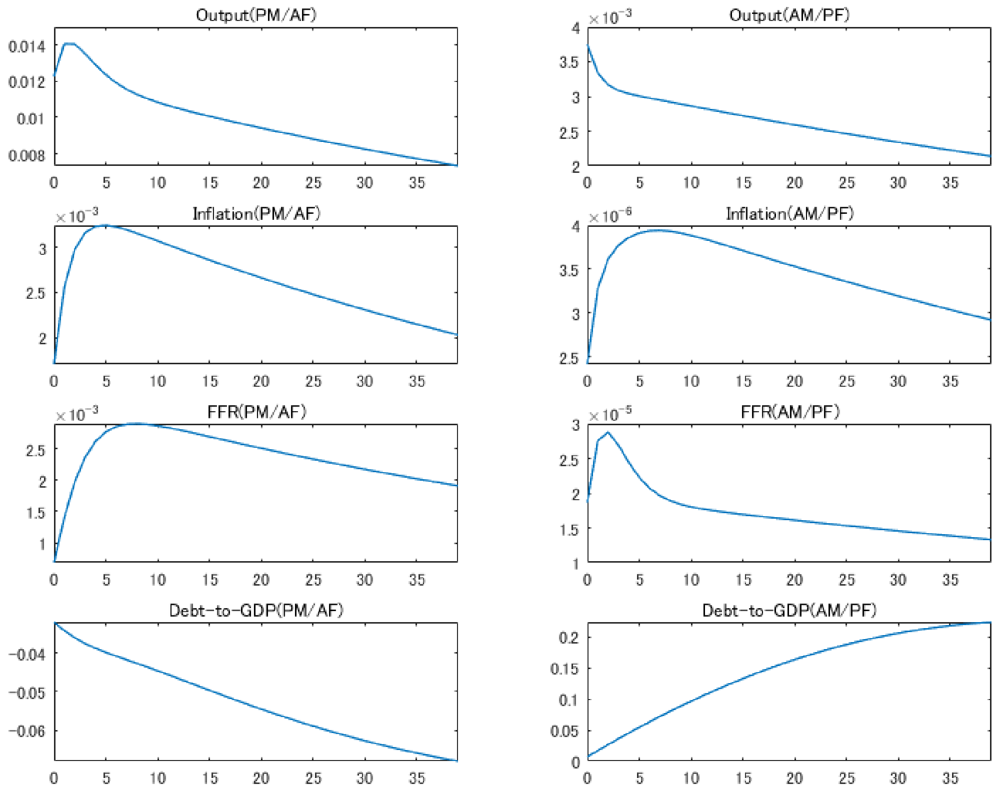

FIGURE A1. Impulse Response Functions to a Long-Term Expenditure Shock

C.4. Impulse Responses

This subsection investigates the impulse response functions to the three structural shocks: long-term expenditure shock, monetary policy shock, and preference shock. The responses are drawn under the condition that the policy regime stays the same for 40 quarters after the shock, but the agents take into account the possibility of regime shifts as well as the feedback channel in the regime rule. You may find the discussion overlapping with Bianchi and Ilut (2017) because the economic intuition remains similar.

C.4.1. Long-Term Expenditure Shock

Figure A1 reports the responses to a positive long-term expenditure shock. If the agents do not take into account the possibility of regime shifts, we observe an increase in output followed by rising inflation under the PM/AF regime. Since the Taylor principle is violated, the nominal interest rate does not increase as much as the rise in inflation.

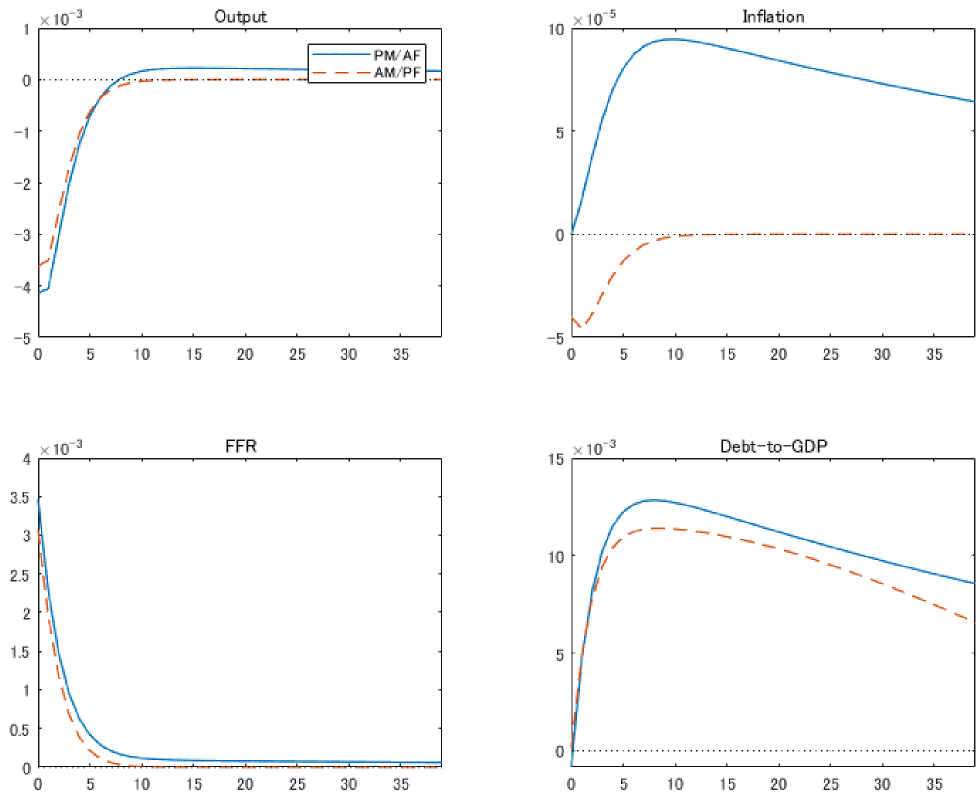

FIGURE A2. Impulse Response Functions to A Monetary Policy Shock

An expansion of the output and a decline in the real interest rate imply a lower debt burden.

Under the AM/PF regime, on the contrary, the effect of the expenditure shock on output and inflation is much smaller because of the Ricardian equivalence: The agents expect increases in the tax rate in the future as the fiscal authority is responsible for the government budget constraint. Despite the fiscal policy being disciplined under the PF policy, the tax rate does not increase enough to keep the debt level at the original level. Combined with the modest inflation, this leads to a hike in the debt-to-GDP ratio.

C.4.2. Monetary Policy Shock

Figure A2 displays the impulse response functions to a contractionary monetary policy shock. The effect of the monetary policy shock on the inflation rate is qualitatively different between the policy regime. In the AM/PF regime, the traditional channel in

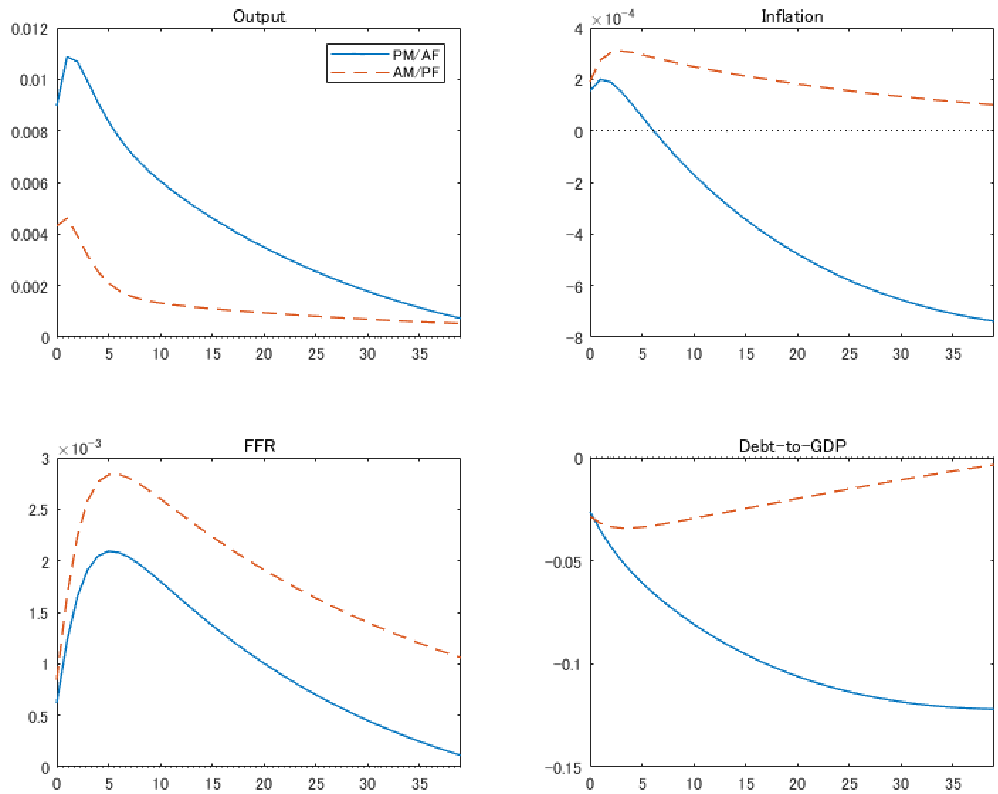

FIGURE A3. Impulse Response Functions to A Preference Shock

the New Keynesian model is at work: Since the Taylor principle holds, the real interest rate goes up after the positive monetary policy shock, leading to the contraction of consumption through intertemporal substitution, which finally causes the decline in the inflation rate. On the contrary, the inflation rate goes up after an increase in the nominal interest rate in the PM/AF regime. The debt burden increases because of the contraction of output and the increase in the interest rate. This in turn implies a surge of inflation to let the intertemporal government budget constraint holds.

C.4.3. Preference Shock

The impulse responses to a positive preference shock are shown in Figure A3. An expansion in the output leads to inflation at impact. Under the PM/AF regime, the inflation rate starts to decline after a while since the expansion makes the fiscal burden smaller.