Feedback in Regime Formation

Naoya Nagasaka*

February 26, 2026
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. We show that the truncation step introduced in the filter to circumvent the path dependence problem has an asymptotically negligible impact on the resulting log-likelihood. We then study the monetary-fiscal policy interaction using the regime-switching DSGE model with the proposed regime determination rule to find the presence of feedback from macroeconomic and policy indicators to the monetary-fiscal policy regime.

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

*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.

[T]he [monetary policy] framework needs to evolve with changes in the structure of the economy and our understanding of those changes.—Powell (2025).

1 Introduction

Since the seminal work of Hamilton (1989), regime-switching models have been 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 regime-switching dynamic stochastic general equilibrium (DSGE) models, which are used to investigate the interaction of monetary and fiscal policies, financial frictions, and economic uncertainty.

The traditional regime-switching models feature time-invariant regime transition probabilities. In other words, the probabilities of shifting 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. As stated in the quote above, the policy regime (or, more generally, the policy framework) itself is an endogenous object subject to changes in economic circumstances. Such a feedback channel in regime determination is assumed away by fixing regime transition probabilities.

The primary objective of this paper is to propose a framework for regime-switching that allows for feedback from economic conditions to regime determination. This paper develops the filter to compute the likelihood efficiently. Simulation exercises demonstrate that the proposed model and filter perform well in finite samples. Finally, we apply the framework to a DSGE model with the monetary/fiscal policy mix.

Our 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, we specify a threshold-type regime rule, which produces time-varying transition probabilities. More specifically, the regime rule consists of (i) a constant term governing the tendency to be in one regime rather than another in steady state, (ii) a linear combination of the lagged continuous state variables, and (iii) a random variable independent of fundamentals. The second component represents the feedback channel from economic conditions to regime formation, while we interpret the third component as the forces driving regime shifts for reasons that are not modeled explicitly, such as the political environment and "sentiments". Importantly, the framework nests the case with constant

regime transition probabilities as a special case.

To compute the likelihood associated with these regime-switching state-space models, we extend the usual Kalman filter with regime-switching coefficients by introducing one additional step: computing the transition probabilities for each period. We show that the transition probabilities conditional on the history of observations are functions of the updated mean and variance of the continuous state variables. Under the assumption of Gaussian-distributed errors in the regime-determination rule, the transition probabilities can be written using the normal cumulative distribution function. Due to this property, we can evaluate the likelihood efficiently to make estimation computationally feasible.

It is well known that we need to introduce an 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 whose dimension grows exponentially with the sample size. A common approach employed in the literature, which is adopted in our filter as well, is to truncate the regime history that must be tracked. We keep track of the most recent r periods instead of the entire history and introduce a collapsing step to integrate out the regime at period $r+1$ ago. We 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 regime realizations older than the truncation threshold r. We will see that the order of convergence depends on the VAR(1) coefficient of the state transition equation. As the system becomes more persistent, we have to keep track of the longer history of regimes to ensure convergence.

To evaluate the finite sample performance of the model and filter, we conduct two Monte Carlo simulation exercises. The first simulation design considers a model with a scalar unobserved state variable and a scalar observation. Since the calculation of the exact likelihood is computationally infeasible, we use a large truncation order, $r = 10$, as the baseline. Comparing the log-likelihood from this baseline and from smaller r, we see that the difference 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 $r = 10$ requires approximately 30 seconds to compute the likelihood just once, preventing us from using this case because the estimation requires evaluation of the likelihood at least thousands of times. The second simulation design is the state-space representation

derived from the regime-switching DSGE model. The purpose of this exercise is to investigate the role of truncation in a more realistic model. We confirm that the likelihood with a small r provides a good approximation in this case as well.

As an empirical illustration, the proposed regime determination rule is applied to the DSGE model with monetary-fiscal interaction in 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. These two regimes have the different mechanisms on how inflation is determined and how the transversality condition for public debt 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 influence 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.

We find evidence of feedback from some macroeconomic variables to regime determination. For instance, a higher debt-to-output ratio and higher inflation make the AM/PF regime more likely. This observation can be interpreted from the perspective of optimal policy: 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 an incentive to switch to the AM/PF regime in order to stabilize inflation. Also, the tax rate and the 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 in the context of forecasting. Using the same DSGE model, we demonstrate that we can help predict regime changes because it provides time-varying regime transition probabilities. In particular, since these probabilities respond systematically to evolving macroeconomic conditions, the model can generate early signals of potential regime shifts before they occur, which are important by-products of our framework for policy-making and other decision-making purposes. Furthermore, by taking such time-varying probabilities of regime changes into account, the baseline model improves forecasts of GDP growth and inflation than the exogenous switching model, especially around periods when regime changes are likely. Relative to the model with time-invariant regime transition probabilities, our framework provides

equal or lower average forecast errors in the medium to long run, and reduces the magnitude of the worst-case forecasting errors.

Literature. This paper contributes to two strands of the literature: the asymptotic properties of regime-switching models in general and the structural macroeconomic models with endogenous regime-switching frameworks. Douc et al. (2004) provide asymptotic properties of 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 established by Li and Liu (2023) and Pouzo et al. (2022). Their frameworks cannot be applied to state-space models as they do not accommodate continuous state variables, . As an extension to the models with continuous state variables, Douc and Moulines (2012) show the consistency of the maximum likelihood estimator under 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 models with non-discrete state variables whose transition kernel is time-varying.

Several works investigate regime-switching state-space models in particular, with a particular focus on the collapsing step in the Kalman filter. Regime truncation is introduced by Kim (1994) to avoid the computational issue related to path dependence. Kim and Kang (2019) examine the precision of the Kim filter through simulation exercises. They compare the likelihood computed from the Kim filter with that from the particle filter, the latter of which is expected to be close to the exact likelihood. They numerically show that those two filters produce similar likelihood and parameter estimation. 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 these two papers to the models with endogenous regime-switching.

There is a growing 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 limitations of the regime-switching framework with time-invariant transition probabilities, several works weaken this assumption and specify the transition probabilities as endogenous, forming the literature on so-called "endogenous regime-switching models". Ascari et al. (2022) study the state-space model in which the equation repre-

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

senting 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 ours. Another related work is Benigno et al. (2020), which assumes 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.

Outline. This paper is organized as follows. In Section 2, we introduce the econometric framework incorporating feedback from lagged continuous state variables to regime formation. Section 3 establishes the asymptotic equivalence between exact and truncated likelihood. Two simulation exercises are conducted in Section 4. Section 5 discusses the empirical application using the DSGE model focusing on the monetary/fiscal policy mix. Section 6 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 $t_1 \leqslant t_2$. For any matrix A, let $\|A\|$ denote the spectral norm of A. For any symmetric matrices A and B of the same size, we write $A \geqslant B$ ($A > B$) 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. Occasionally we write $\phi_k(x)$ and $\Phi_k(x)$ to make explicit the dimension of $x \in \mathbb{R}^k$.

2 Baseline Specification

This section introduces the model specification used in this paper, and compare the specification with others in the literature. We will also discuss the filtering algorithm for likelihood evaluation.

2.1 Regime-Switching State-Space Model

We consider a linear-Gaussian state-space model with regime-switching parameters. The linear-Gaussian state-space model is a commonly used empirical framework to estimate dynamic stochastic general equilibrium (DSGE) models. The regime-switching structure allows us to model discrete changes in the coefficients over time, such as shifts in the monetary policy stance (Dovish vs. Hawkish) and macroeconomic volatility.

The model consists of an observed variable $y_t \in \mathbb{R}^{d_y}$ and an unobserved continuous 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 a standard Gaussian distribution. The first equation is the transition equation specifying the dynamics of the latent variable x_t. The second equation is the observation equation which relates the observed y_t to the unobserved x_t. The coefficients $(A_{s_t}, B_{s_t}, Q_{s_t}, R_{s_t})$ depend on the discrete latent regime $s_t \in \mathcal{S} \equiv \{0, 1\}$.[2] The model (1) does not include constant terms for expositional simplicity, although it is fairly straightforward to incorporate them.

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

Motivated by this consideration, we 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 rule.

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

where $\tau \in \mathbb{R}$, $\lambda \in \mathbb{R}^{d_x}$, and (η_t) is independent of u_t and ε_t, with cumulative distribution function F. For the theoretical discussion in Section 3, we will assume F to be the standard normal which simplifies the expression of the regime transition probabilities. Meanwhile,

[2] The regime s_t is assumed to be binary throughout this paper, while one can easily extend the framework to allow for more than two regimes.

other choices for F, such as the logistic function, are also possible. For example, η_t can follow an AR(1) process so that we allow the persistence of regimes orthogonal to the economic model.

The regime specification in 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, it allows the policy stance to be influenced by various economic indicators. The constant term τ governs the baseline tendency toward regime 1. The specification (2) also includes the random component η_t independent of all lags and leads of (x_t, y_t). One way to interpret this term is as a regime shifter that is exogenous to the state dynamics, such as political forces. Alternatively, this term might capture sentiments, which do not directly drive the business cycle but have implications for macroeconomic outcomes 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 regime to adopt. In the context of macroeconomic policy analysis, she can be the policy authority. In other applications where no actual decision maker is present (e.g., time-varying volatility), we may regard her as "nature". This decision maker chooses the regime that yields a higher payoff. In this context, the left-hand side of the inequality in (2) can be interpreted as a net payoff from choosing regime 1 over regime 0.

2.2 Comparison with Related Regime-Switching Rules

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 into an otherwise standard state-space model. Their long-run Phillips curve has a kink that depends on trend inflation: the slope of the long-run Phillips curve changes when the current trend inflation crosses a certain threshold. This specification has similarities with equation (2), except that they incorporate feedback from the current state variables to the regime rather than the lagged states as in our specification, and they do not incorporate η_t, the stochastic component orthogonal to the structural shocks ε_t and

[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 (e.g., Angeletos et al. 2007). Conceptually, this idea is close to the regime-switching state-space models like ours.

measurement errors u_t. In terms of the estimation strategy, while they rely on a particle filter, this paper develops the algorithm based on the Kalman filter.[4] Although we do not compare those two filters, ours is expected to be computationally more efficient in linear-Gaussian settings.

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

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

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 w_t depends on the past structural shock ε_{t-1}, which introduces feedback from the structural shock as a fundamental driver of economic fluctuations to regime determination. By recursive substitution, 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

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

Both equations involve the summation of initial conditions, the sequence of structural shocks ε_t, and stochastic terms ν_t or η_t. Although ε_0 matters for the former, it does not for the latter, its contribution is expected to be negligible under $|\alpha| < 1$. Hence, our framework can be regarded as an extension of the CCP-type regime-switching model in the sense that the weight in the linear combination is regime-dependent.

As a special case, 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. 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, those two frameworks are identical to regime-switching models with time-invariant transition probabilities.

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

The specification of the regime-switching rule is an application-dependent question. An advantage of our approach is that, as discussed earlier, we directly model the net payoff of taking one regime over another as a function of state variable x. In practice, structural shocks are not directly observed, and regime determination is based on macroeconomic and financial indicators available in real time. By specifying the regime indicator as a function of x_{t-1}, our approach directly reflects the information available at the time regime choices are made. This formulation therefore provides a more realistic description of regime determination, in which regime changes are triggered by the systematic assessment of economic conditions rather than by latent disturbances.

2.3 Filtering Algorithm

In order to estimate the parameters, we need to derive the likelihood implied by equations (1). We extend the Kalman filter to allow for the regime rule (2). Here, we present only the key steps in the algorithm. A detailed description is given in Appendix A.

The structure of the filter is similar to the Kalman filter with regime-switching coefficients proposed by Kim (1994), which iterates through the forecasting and updating steps for each $t = 1, \cdots, T$. Kim (1994)'s algorithm assumes that the regime transition probabilities are time-invariant and are treated as parameters. As our transition probabilities depend on the lagged state variables x_{t-1}, an extra step is required to calculate the filtered regime transition probabilities for each period.

Truncation. When the parameters are subject to regime-switching, the likelihood evaluation through the Kalman filter is subject to path dependence: the likelihood of y_t depends on s_1^t, the entire history of the regime up to period t. The possible combinations of regime histories grow exponentially with T, making the exact filter—i.e., the filter without any approximation scheme—computationally infeasible. To circumvent this issue, we propose an algorithm that tracks only the most recent $r(T)$ periods. The truncation order $r(T)$ increases with the sample size T.[5] Henceforth, we simply write r instead of $r(T)$ unless we need to stress the dependence of the truncation number on T.

More specifically, let $\mathcal{F}_{t-1} := \sigma(y_1^{t-1})$, the σ-field generated by the history of observa-

[5]Hashimzade et al. (2025) propose an alternative filtering and smoothing algorithm for state-space models with regime-switching coefficients based on the Interactive Multiple Model filter, and numerically show that its approximation accuracy is close to that of Kim (1994)'s filter while being more computationally efficient. Their assessment of approximation accuracy is based on a particular DSGE model, and a theoretical investigation of whether these findings extend to more general models remains an open question.

tions up to time $t-1$. We denote by $\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2})$ and $\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})$ the approximations of conditional mean and variance of the state vector given s_{t-r+1}^{t-2} and $\mathcal{F}_{t-1}$, computed using the filter with regime truncation. Then, at the end of the iteration at time t, we introduce the truncation step to compress the $2^{r(T)}$ conditional means and variances into $2^{r(T)-1}$ objects:

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

where $p_r(\cdot|\mathcal{F}_t)$ is the regime probability conditional on the information up to time t. Note that $\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2})$ and $\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})$ do not coincide with the true conditional mean and variance, $x_{t-1|t-1}(s_{t-r+1}^{t-2})$ and $\Omega(s_{t-r+1}^{t-2})$. We carry the truncated regime history s_{t-r+2}^{t} forward to time $t+1$ by using $\bar{x}_{t|t}(s_{t-r+2}^{t})$ and $\bar{\Omega}_{t|t}(s_{t-r+2}^{t})$ as the inputs of the iteration at $t+1$. Section 3 proves that this truncation strategy yields an asymptotically equivalent likelihood to the exactly (but computationally infeasible) one.

Time-Varying Transition Probabilities. As discussed above, the key departure from the model with time-invariant regime transition probabilities is that the computation of regime transition probabilities should be built into the filter. In order to illustrate this step in more detail, we consider the lag-augmented state-space model without loss of generality.

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

This augmentation is necessary because, in order to compute the transition probability from s_{t-1} to s_t conditional on $\mathcal{F}_{t-1}$, we require information about the probability of s_{t-1}. Since s_{t-1} depends on x_{t-2}, we require the conditional mean and variance of x_{t-2} given the truncated regime history and $\mathcal{F}_{t-1}$. The lag-augmentation facilitates the computation of this object.

We redefine $x_t = [x_t', x_{t-1}']'$. Under the assumption that $\eta_t \sim N(0,1)$, the transition

probability from s_{t-1} to s_t is approximated as

$$
\begin{aligned}
&p_r(s_t = 0 | s_{t-1} = 0, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}) \\
&\approx \frac{\Phi_2\left(-\tau\iota_2; \Lambda\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2}), I + \Lambda\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\Lambda'\right)}{\Phi_1\left(-\tau; \lambda'\left(\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x)}, 1 + \lambda'\left(\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x, d_x+1:2d_x)}\lambda\right)} \\
&p_r(s_t = 1 | s_{t-1} = 1, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}) \\
&\approx \frac{\Phi_2\left(\tau\iota; -\Lambda\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2}), I + \Lambda\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\Lambda'\right)}{\Phi_1\left(\tau; -\lambda'\left(\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x)}, 1 + \lambda'\left(\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x, d_x+1:2d_x)}\lambda\right)}
\end{aligned}
$$

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

where ι_n ($n \in \mathbb{N}$) is a $n \times 1$ vector whose elements are all unity, and $\Lambda = \begin{bmatrix} \lambda' & 0 \\ 0 & \lambda' \end{bmatrix}$ is a $2 \times 2d_x$ matrix. In the denominator, we extract the $(d_x + 1)$-st through $2d_x$-th elements from $\bar{x}_{t-1|t-1}(\cdot)$ and the corresponding rows and columns from $\bar{\Omega}_{t-1|t-1}(\cdot)$ to take into account the information about x_{t-2} given $\mathcal{F}_{t-1}$ and truncated regime history. See Appendix A for the derivation.

Augmenting these approximate transition probabilities in the forecasting step of the filter enables us to derive the likelihood, which is a key ingredient of maximum likelihood estimation and Bayesian inference.

3 Asymptotic Equivalence between Exact and Truncated Likelihood

This section establishes that the approximated likelihood computed from our filter is asymptotically identical to the exact likelihood. Let $\theta := ((A_s, B_s, Q_s, R_s)_{s=0,1}, \tau, \lambda)$ denote

the collection of parameters, which lies in a compact parameter space Θ. Let

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

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

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

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

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

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

Intuitively, the UCO implies that the available macro data y_t contains enough information to infer the latent economic states x_t over a fixed-length window at any date t. The UCC ensures that the structural shocks ε_t can influence the state x_t over any fixed horizon at all dates.

To see how these two assumptions operate in our framework, we consider the Kalman filter in the absence of regime truncation, with different initial variance–covariance matrices Ω^1 and Ω^2 but the same regime path $(s_1, \ldots, s_t)$. Denote by $\Omega_{t|t}^1(s_1^t)$ and $\Omega_{t|t}^2(s_1^t)$ the conditional variance matrices corresponding to initial conditions 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

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

Therefore, it suffices to study the properties of $\Psi^k(\cdot)$ to establish filter convergence. As shown in Jazwinski (1970), $\Psi^k(\cdot)$ converges exponentially with respect to k under UCO and UCC.

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

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

Now we provide the main proposition in this section. We set the initial state variables arbitrarily: $x_1 = \tilde{x}$ and $s_1 = \tilde{s}$. Although these two initial variables are fixed here, we can allow x_1 and s_1 to follow known distributions with a minor change in the proof. Let $p_{r(T),\theta}(y_1^T | x_1 = \tilde{x}, s_1 = \tilde{s})$ be the likelihood under regime truncation 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 log-likelihoods is asymptotically negligible.

Proposition 1. Assume Assumptions 1 and 2 and let F be a standard normal distribution. For any $\theta \in \Theta$, we have

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

where c_2 is given in Lemma 1.

This result reflects that the influence of distant past information decays over time, due to the exponential stability of the Kalman filter in Lemma 1. Hence, regime realizations in the distant past can be asymptotically neglected. This convergence occurs at an exponential rate in r.

Here we provide a sketch of the proof. The details can be found in Appendix B. Hence-

forth we simply denote $r = r(T)$. 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

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

After taking logs on both sides, the difference between the two log-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 in the regime transition probabilities (the third line). Since both of them are functions of the conditional mean and variance implied by 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 2 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

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

Utilizing these inequalities, one can show the boundedness of the second and third lines of (4), establishing our claim.

Remark 1. Li (2023) shows an 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.

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 of the transition probabilities 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 Simulations

This section presents two simulations to evaluate the finite-sample performance of the filter studied in the previous section. We are particularly interested in how different assumptions on the truncation order r affect the log-likelihood and parameter inference. The first exercise examines maximum likelihood estimates of parameters of a simple univariate model. The second simulation considers a larger and more realistic framework, building on a New Keynesian model following our empirical specification.

4.1 Simulation 1: Univariate Model

The first simulation considers a univariate state-space model with $d_x = d_y = d_u = d_\varepsilon = 1$. We generate a simulated sample $\{y_t\}_{t=1}^T$ based on the parameters given in Table 1 and $\eta_t \sim N(0,1)$. The persistence of x_t captured by A_{s_t} varies across simulation designs while

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	

maintaining $A_0 < A_1$. The standard deviations Q_{s_t} and R_{s_t} are larger in regime 1. Hence, regime 1 can be interpreted as the more persistent and volatile regime. We allow feedback from x_{t-1} into regime determination by setting nonzero λ. Regime 1 is more likely to occur since τ is positive. The sample size is set to $T = 400$.

We first evaluate the difference between the log-likelihood computed using a small truncation number r (plausible for empirical applications) and that computed with a relatively large truncation order $r = 10$ over 500 simulations. Both are evaluated at the true parameter values. Given that it is almost infeasible to calculate the exact log-likelihood (i.e., $r = 400$), we use $r = 10$ as the benchmark.[6]

Figure 1 reports the kernel densities of the differences between the log-likelihood under the benchmark ($r = 10$) and that obtained with smaller truncation order $r = 1, 2, 3, 4$, under $(A_0, A_1) = (0.5, 0.7)$. The difference shrinks as we enlarge r and the rate of shrinkage appears to be exponential, as expected from the theoretical analysis in the previous section.

Figure 2 investigates implications of the persistence parameters for the kernel densities of the differences in the log-likelihood. We fix $r = 2$, and consider $(A_0, A_1) = (0.7, 0.95)$ shown as a solid line, $(0.5, 0.95)$ as a dashed line, and $(0.5, 0.7)$ as a dotted line. The model exhibiting less persistence gives a smaller difference, as can be seen from more concentrated densities. The persistence parameters at both more and less persistent regimes matter for the shrinkage.

Next, in order to examine the finite-sample performance of the filter, we estimate the parameters $(A_0, A_1, Q_1, R_1, \tau, \lambda)$ using maximum likelihood estimation for $(A_0, A_1) = (0.5, 0.7)$ for the 500 series of simulated data. For each simulation, the asymptotic confidence intervals are constructed using the inverse Hessian. The coverage probabilities of

[6] Another possible strategy to approximate the exact log-likelihood is to use the filtering algorithm that allows for nonlinearity. For example, Kim and Kang (2019) use the particle filter as the benchmark to examine the accuracy of Kim (1994) filter.

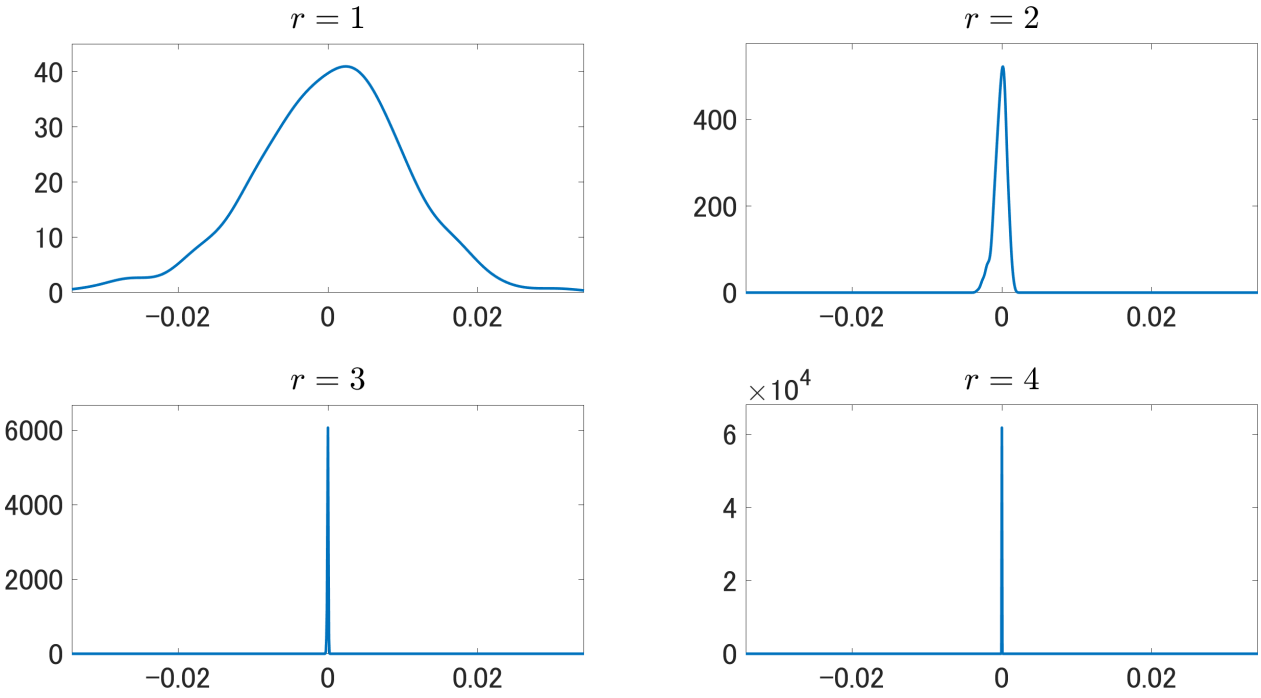

Figure 1: Kernel Density of Likelihood Differences under $(A_0, A_1) = (0.7, 0.95)$

Note: The figure plots the kernel density of the difference between log-likelihood with $r = 10$ and the one with $r = 1, 2, 3, 4$ over 500 simulations under $(A_0, A_1) = (0.5, 0.7)$.

the asymptotic 95% confidence intervals are reported in Table 2. Except for A_0, the coverage probability with $r = 1$ is close to 95% compared to that of $r = 2, 3$, implying that the filter with small r exhibits satisfactory finite-sample performance.

Table 2 also reports the median run time for a single likelihood evaluation. As we have to keep track of all 2^r possible regime histories, the run time almost doubles as we increase r by one. When $r = 10$, the filter spends almost half a minute to compute the likelihood once. It is unrealistic to estimate the model in such a setting since the estimation algorithm generally requires at least thousands of likelihood evaluations.

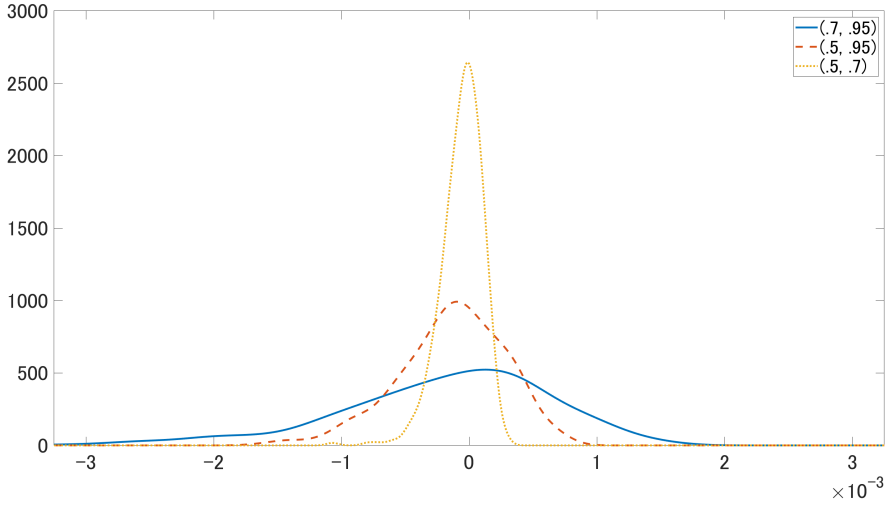

Figure 2: Kernel Density of Likelihood Differences under Different Persistence Parameters

Note: The figure plots the kernel density of the difference between log-likelihood with $r = 10$ and the one with $r = 2$ over 500 simulations under different combinations of (A_0, A_1): $(0.7, 0.95)$ by solid line, $(0.5, 0.95)$ by dashed line, and $(0.5, 0.7)$ by dotted line.

Table 2: Coverage Probability of the 95% Confidence Interval and Computational Costs

r	1	2	3	4	Truth
Panel A. Coverage Probability					
A_0	0.918	0.910	0.920	0.916	0.5
A_1	0.936	0.906	0.932	0.928	0.7
Q_1	0.864	0.838	0.848	0.866	2
R_1	0.808	0.760	0.786	0.778	1.5
λ	0.694	0.662	0.684	0.720	0.8
τ	0.700	0.684	0.680	0.704	0.2
Panel B. Median Implementation Time (ss)					
	0.073	0.135	0.258	0.501	30.567 ($r = 10$)

Note: Panel A reports the coverage probability of the 95% confidence interval under different assumptions on r. The rightmost column shows the true parameter values. Panel B reports the median time (in seconds) for single likelihood evaluation, with the rightmost column showing the computation cost for $r = 10$. Both results are under $(A_0, A_1) = (0.5, 0.7)$.

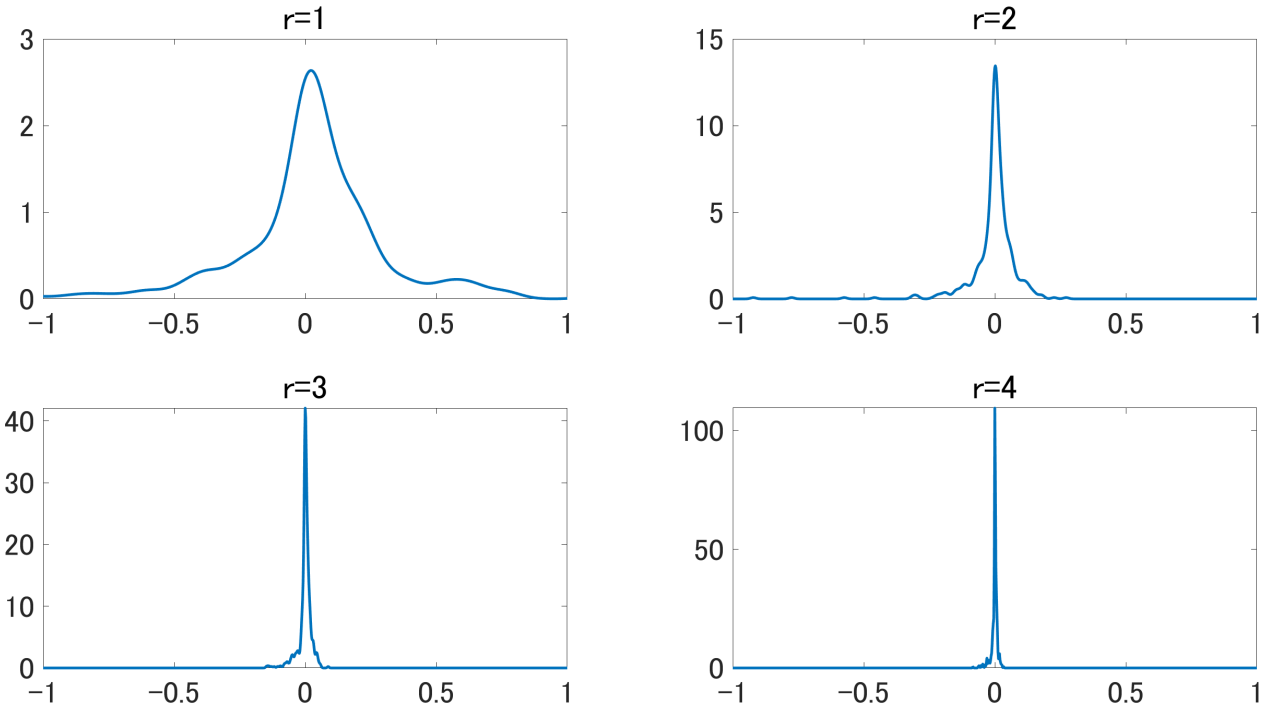

Figure 3: Kernel Density of Likelihood Differences from DSGE

Note: The figure plots the kernel density of the difference between the log-likelihood with $r = 10$ and the one with $r = 1, 2, 3, 4$ over 500 simulations from the DSGE model.

Table 3: Median Time for Single Likelihood Evaluation for DSGE Model

r	1	2	3	4	10
Time (ss)	0.364	0.724	1.460	2.910	173.473

Note: This table reports the median time for single likelihood evaluation with different assumptions on r for the DSGE model.

4.2 Simulation 2: Log-Likelihood from DSGE

The purpose of the second simulation exercise is to examine whether the regime truncation matters in a more realistic model. Given that DSGE models are rarely estimated with maximum likelihood, we only provide the log-likelihood for different assumptions on r. It is instructive to check whether the log-likelihood is precisely calculated with relatively small r as it is a central input of Bayesian inference, which is typically employed to estimate DSGE models.

As the data-generating process, we use our empirical model discussed in Section 5—the New Keynesian model featuring switches in (i) monetary-fiscal policy regimes with time-varying regime transition probabilities, and (ii) regimes in macroeconomic volatility

with time-invariant transition probabilities—with two modifications. First, we assume that the economy stays in the low volatility regime by shutting down volatility switching. Second, we employ the posterior mean given in Tables 4 and 5, but set the constant term in the policy regime rule to be zero—i.e., $\tau^{pol} = 0$. This treatment makes both regimes equally likely ex ante.

We solve the DSGE model at the posterior mean and $\tau^{pol} = 0$ to obtain the state-space representation. The sample length is set to $T = 200$, aligning well with the sample size of most DSGE exercises using the post-WWII quarterly data.

Figure 3 plots the kernel densities of the absolute value of the difference between the log-likelihood with $r = 1, 2, 3, 4$ and that with $r = 10$ over 500 Monte Carlo replications. The absolute differences are larger than in the simple univariate application (Figure 1), as reflected in the wider range of the horizontal axis, but they are still concentrated at zero. We also observe the shrinkage of the difference for as r increases, consistent with both theory and the previous simulation results. In terms of the run time displayed in Table 3, we can calculate the likelihood in a reasonable amount of time even for the relatively large state-space model studied here. However, the computational burden grows exponentially as r increases, and the case with $r = 10$ requires nearly three minutes to a single likelihood evaluation.

5 Empirical Application

In this section, we study the implications of the feedback channel in regime determination using the New Keynesian model with a monetary-fiscal policy mix. After providing an overview of the model, we discuss the prior distributions of the parameters as well as the estimation strategy. Then we provide the posterior distributions and show that the proposed model is useful for forecasting. We set $r = 2$ throughout this section, as this choice provides a sufficiently accurate approximation as shown in the simulation exercise above.[7]

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

5.1 Model

5.1.1 Model Environment

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 to the early 1980s. The representative household consumes the final good and supplies labor in each period. In addition to the one-period government bond, the household is allowed to hold long-term government securities. There is a continuum of firms producing differentiated goods, which are aggregated into a final consumption good. These firms are subject to monopolistic competition and quadratic price adjustment costs.

The government conducts fiscal and monetary policies. The one-period government bond is assumed to have zero net supply. Total government expenditure consists of direct lump-sum transfers to the household as well as government consumption. The total government expenditure and interest payments on government bonds will be financed by the lump-sum tax. The nominal interest rate is determined following a Taylor rule that depends on current inflation rate and output gap. The tax rule and Taylor rule are subject to regime-switching as discussed below.

5.1.2 Regime-Switching

We introduce two types of regimes to the model: $s_t^{vol} \in \{0, 1\}$ related to volatility and $s_t^{pol} \in \{AM/PF, PM/AF\}$—active monetary/passive fiscal policies and passive monetary/active fiscal policies—capturing policy stances. The volatility of the structural shocks changes over time depending on s_t^{vol}. Regime $s_t^{vol} = 1$ can be regarded as a high-volatility regime. The volatility regime follows a standard Markov process with a time-invariant transition probability matrix P^{vol}.

$$P^{vol} = \begin{bmatrix} 1 - p_{0,1}^{vol} & p_{0,1}^{vol} \\ p_{1,0}^{vol} & 1 - p_{1,0}^{vol} \end{bmatrix}$$

Switching in s_t^{vol} is assumed to be exogenous. This assumption might be reasonable if we regard the changes in economic volatility as events driven by factors outside the model. For example, the estimation results in Bianchi and Ilut (2017) suggest that the oil crisis in the mid-1970s is classified as a high-volatility regime, and this episode is mainly due to geopolitical forces. In addition, the elevated volatility during the Great Recession results

from financial market instability, while financial frictions are not modeled explicitly here.

The other regime $s_t^{pol} \in \{AM/PF, PM/AF\}$ is related to the monetary-fiscal policy mix framework pioneered by Leeper (1991).[8] 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 consider the linearized tax rule and the Taylor rule.[9]

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

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

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 $\tilde{b}_{t-1}^m$, total government expenditure $\tilde{e}_t$, and output gap $\hat{y}_t - \hat{y}_t^*$. The policy interest rate $\tilde{R}_t$ is influenced by the inflation rate $\tilde{\pi}_t$ as well as the 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 keep inflation close to its target. On the other hand, fiscal policy plays a role in determining 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 sufficient inflation to inflate away the public debt. The central bank gives up stabilizing the price level and responds weakly to inflation. 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-lived relative to the other two regimes and our regime determination rule does not allow for more than two regimes with not natural ordering (e.g., high/middle/low parameter values).[10]

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

[9] For a variable X_t, we let X denote its value at the steady state. We denote $\hat{x}_t = \log((X_t/A_t)/(X/A))$, where A_t is the total factor productivity, to represent the percentage deviation of the detrended variable from the steady state. In addition, if X_t is expressed as a ratio to nominal GDP, we denote $\tilde{x}_t = X_t - X$ to avoid taking the percentage twice. If not, we denote $\tilde{x}_t = \log(X_t/X)$.

[10] One may consider introducing two regime factors representing the monetary and fiscal policy stance respectively. This would enable us to investigate two other policy combinations not considered here: AM/AF and PM/PF. Instead, we 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 relative to other regimes, possibly because they do not guarantee a unique rational expectations equilibrium (AM/AF gives an explosive equilibrium, and PM/PF is subject to the

The regime indicator s_t^{pol} is specified as in equation (5). We allow key macro and fiscal variables (namely output gap, inflation, nominal interest rate, and tax rate) to influence the policy regime, while abstracting away feedback from other variables to reduce the number of estimated parameters.

$$s_t^{pol} = \begin{cases} AM/PF & \tau^{pol} + \lambda_y(\hat{y}_{t-1} - \hat{y}^*_{t-1}) + \lambda_\pi \tilde{\pi}_{t-1} + \lambda_R \tilde{R}_{t-1} + \lambda_b \tilde{b}^m_{t-1} + \lambda_\tau \tilde{\tau}_{t-1} + \eta_t \geqslant 0 \\ PM/AF & \text{otherwise} \end{cases} \tag{5}$$

where $\eta_t = \rho_\eta \eta_{t-1} + \varepsilon_{\eta,t}$, $\varepsilon_{\eta,t} \sim N(0,1)$. Allowing for serial correlation in the error term η_t is crucial for the estimation results because the persistence of s_t^{pol} depends solely on that of economic variables otherwise.

5.2 Data, Solution, and Estimation

We construct the quarterly dataset of macroeconomic and fiscal variables, consisting 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. The dataset is constructed following the description in Bianchi and Melosi (2022) because Bianchi and Ilut (2017) relies on the old definition in NIPA, which does not perfectly align with the currently available NIPA tables. The sample spans 1954Q4–2009Q3.

The model needs to be solved for each draw of the parameters to obtain a state-space representation. We employ the perturbation method proposed by Maih and Waggoner (2018) which accommodates regime-switching DSGE models with time-varying transition probabilities as in our application. Maih and Waggoner (2018)'s perturbation method is available in the Rationality In Switching Environments (RISE) Toolbox developed by Junior Maih.[11] We use the built-in functions in RISE to solve the model and then plug the resulting state-space representation into our filtering algorithm to compute the likelihood. The details of the perturbation method are available in Appendix C.

Given the nonlinearity of our model, the posterior distribution might be irregular or multimodal. For this reason, we employ the Sequential Monte Carlo (SMC) approach, a variant of Markov Chain Monte Carlo methods which has been shown in practice to be robust to such nonstandard posterior distributions.[12] Appendix C provides details of

indeterminacy of equilibria).

[11] `https://github.com/jmaih/RISE_toolbox`

[12] Bognanni and Herbst (2018) discuss issues in the estimation of Markov-switching vector autoregressive

the sampler. Another attractive feature of SMC is that we can parallelize the likelihood evaluations across particles, which is the most computationally demanding part.

5.3 Priors and Calibration

The right columns in Tables 4 and 5 show the prior distributions of the estimated parameters. Other than the parameters governing the transition of the policy regime, we employ exactly the same prior as in Bianchi and Ilut (2017). The persistence 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 at $\beta = 0.9965$, 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 rule, the prior means of τ^{pol} and ρ_η imply that the probability of remaining in the same regime is 0.85 in the absence of the feedback channel. Each element of λ follows the uniform distribution centered at zero with sufficiently wide support so as not to impose prior information on the signs and sizes of the feedback effects.

5.4 Posteriors

The left columns in Tables 4 and 5 show the posterior distributions of the parameters for the baseline model and the exogenous switching model where the transition probabilities of the policy regimes are constant. We specify the exogenous switching model by restricting all elements in λ to be zero, which gives time-invariant regime transition probabilities. For both the baseline and exogenous switching models, $\phi_\pi(PM)$—the coefficient on the inflation rate in the Taylor rule under passive monetary policy—is below unity, while the corresponding parameter under active monetary policy, $\phi_\pi(AM)$, is above unity. In addition, the sensitivity of tax rate to outstanding debt under the passive fiscal policy, $\delta_b(PF)$ is above the gross real interest rate, $1/\beta - 1 = 0.004$. These observations justify the labeling of the policy regimes—AM/PF on the one hand and PM/AF on the other hand. The baseline model and exogenous switching model yield similar posterior distributions, while we find discrepancies in τ^{pol} and ρ_η which are estimated to be smaller in the baseline model.

Feedback Coefficients λ. The posterior mean of λ_b is positive and its 90% credible does not include zero. This estimate implies that the policy regime is more likely to be AM/PF

models, and propose the SMC-based estimation algorithm.

Table 4: Prior and Posterior Distributions: Part1

	Posterior (Baseline)			Posterior (Exog. Switch)			Prior		
	Mean	5%	95%	Mean	5%	95%	Type	Para (1)	Para (2)
δ_y	0.297	0.244	0.351	0.292	0.237	0.347	N	0.200	0.200
δ_e	0.309	0.197	0.425	0.283	0.173	0.392	N	0.500	0.250
ι_y	0.092	-0.116	0.290	0.099	-0.114	0.307	N	0.100	0.200
ϕ_y	-0.623	-0.686	-0.563	-0.605	-0.669	-0.542	N	0.100	0.200
ς	0.492	0.434	0.549	0.506	0.451	0.559	B	0.500	0.250
Φ	0.396	0.356	0.435	0.383	0.339	0.427	B	0.500	0.250
κ	0.002	0.001	0.002	0.001	0.001	0.002	G	0.300	0.150
ρ_χ	0.996	0.993	0.998	0.995	0.993	0.998	B	0.500	0.200
ρ_a	0.649	0.586	0.712	0.704	0.639	0.767	B	0.500	0.200
ρ_d	0.969	0.965	0.973	0.972	0.968	0.976	B	0.500	0.200
ρ_{eS}	0.180	0.137	0.223	0.182	0.137	0.229	B	0.200	0.050
ρ_μ	0.051	0.023	0.082	0.046	0.020	0.075	B	0.500	0.200
ρ_{tp}	0.219	0.151	0.285	0.215	0.143	0.286	B	0.500	0.200
100π	0.579	0.546	0.613	0.587	0.552	0.622	N	0.500	0.050
100γ	0.501	0.463	0.538	0.499	0.459	0.539	N	0.420	0.050
b^m	0.716	0.663	0.770	0.735	0.675	0.794	N	1.000	0.100
g	1.080	1.076	1.084	1.079	1.075	1.084	N	1.080	0.040
τ	0.171	0.170	0.173	0.171	0.169	0.172	N	0.180	0.005

Note: This table describes the posterior mean as well as the 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 the uniform distribution, Para (1) and (2) are the prior mean and standard deviation respectively. For the uniform distribution, those two parameters are the lower and upper bounds.

after observing an increase in the debt-to-output ratio in the previous quarter. This feedback channel can be justified from an optimal policy perspective. Suppose that the policy authority observes a high debt-to-output ratio when operating under the PM/AF regime. Under the PM/AF regime, public debt will be inflated away to satisfy the transversality condition. Under the standard New Keynesian mechanism, high inflation leads to large price dispersion, and it causes welfare losses. To stabilize inflation, the policy authority has an incentive to switch to the AM/PF regime. The point estimate of λ_π is negative although the credible interval includes zero. Noting that the intertemporal government budget constraint depends not only on current inflation but also on series of inflation expectations, the estimates of λ_b and λ_π suggest the role inflation expectations play in the policy regime determination through public debt dynamics.

Another parameter related to fiscal policy, λ_τ has a positive posterior mean, consistent with the view that the policy authority cares about fiscal discipline in the passive fiscal

Table 5: Prior and Posterior Distributions: Part2

	Posterior (Baseline)			Posterior (Exog. Switch)			Prior		
	Mean	5%	95%	Mean	5%	95%	Type	Para (1)	Para (2)
$\psi_\pi(PM)$	0.226	0.162	0.294	0.232	0.168	0.298	G	0.800	0.300
$\psi_\pi(AM)$	1.696	1.298	2.105	1.774	1.381	2.187	N	2.500	0.500
$\psi_y(PM)$	0.208	0.186	0.230	0.187	0.168	0.207	G	0.150	0.100
$\psi_y(AM)$	1.071	0.885	1.268	1.093	0.898	1.297	G	0.400	0.200
$\rho_R(PM)$	0.700	0.658	0.743	0.675	0.639	0.713	B	0.500	0.200
$\rho_R(AM)$	0.861	0.837	0.884	0.868	0.846	0.889	B	0.500	0.200
$\delta_b(PF)$	0.048	0.036	0.061	0.047	0.034	0.061	G	0.070	0.020
$\rho_\tau(AF)$	0.793	0.743	0.841	0.790	0.735	0.840	B	0.500	0.200
$\rho_\tau(PF)$	0.979	0.969	0.988	0.978	0.968	0.987	B	0.500	0.200
$100\sigma_R(0)$	0.072	0.066	0.079	0.073	0.066	0.079	IG	0.500	0.500
$100\sigma_R(1)$	0.380	0.340	0.421	0.389	0.347	0.434	IG	0.500	0.500
$100\sigma_\chi(0)$	1.977	1.835	2.123	1.986	1.835	2.147	IG	1.000	1.000
$100\sigma_\chi(1)$	4.671	4.190	5.174	4.715	4.248	5.218	IG	1.000	1.000
$100\sigma_a(0)$	0.396	0.342	0.450	0.346	0.296	0.398	IG	1.000	1.000
$100\sigma_a(1)$	0.689	0.556	0.831	0.612	0.490	0.736	IG	1.000	1.000
$100\sigma_\tau(0)$	0.257	0.239	0.276	0.254	0.236	0.274	IG	2.000	2.000
$100\sigma_\tau(1)$	0.732	0.664	0.802	0.731	0.656	0.810	IG	2.000	2.000
$100\sigma_d(0)$	6.263	5.645	6.887	6.591	5.896	7.325	IG	10.000	2.000
$100\sigma_d(1)$	10.668	9.290	12.039	10.943	9.448	12.542	IG	10.000	2.000
$100\sigma_{eS}(0)$	0.226	0.192	0.261	0.226	0.190	0.263	IG	2.000	2.000
$100\sigma_{eS}(1)$	0.387	0.310	0.469	0.387	0.308	0.470	IG	2.000	2.000
$100\sigma_{tp}(0)$	2.588	2.433	2.748	2.582	2.418	2.749	IG	1.000	1.000
$100\sigma_{tp}(1)$	3.323	2.955	3.705	3.318	2.926	3.710	IG	1.000	1.000
$100\sigma_\mu(0)$	0.140	0.129	0.152	0.137	0.126	0.148	IG	1.000	1.000
$100\sigma_\mu(1)$	0.272	0.243	0.303	0.273	0.243	0.305	IG	1.000	1.000
$p_{0,1}^{vol}$	0.094	0.072	0.117	0.101	0.076	0.128	B	0.170	0.100
$p_{1,0}^{vol}$	0.242	0.180	0.306	0.247	0.183	0.312	B	0.170	0.100
τ^{pol}	-6.690	-8.734	-4.759	-1.098	-3.781	1.485	N	0.000	5.000
$0.01\lambda_y$	-0.358	-0.661	-0.074	—	—	—	U	-10.000	10.000
$0.01\lambda_\tau$	0.605	0.085	1.154	—	—	—	U	-10.000	10.000
$0.01\lambda_\pi$	-0.264	-1.441	0.940	—	—	—	U	-10.000	10.000
$0.01\lambda_R$	0.890	0.273	1.530	—	—	—	U	-10.000	10.000
$0.01\lambda_b$	0.025	0.015	0.036	—	—	—	U	-10.000	10.000
ρ_η	0.960	0.938	0.981	0.993	0.989	0.997	B	0.900	0.050

Note: See the footnote for Table 4.

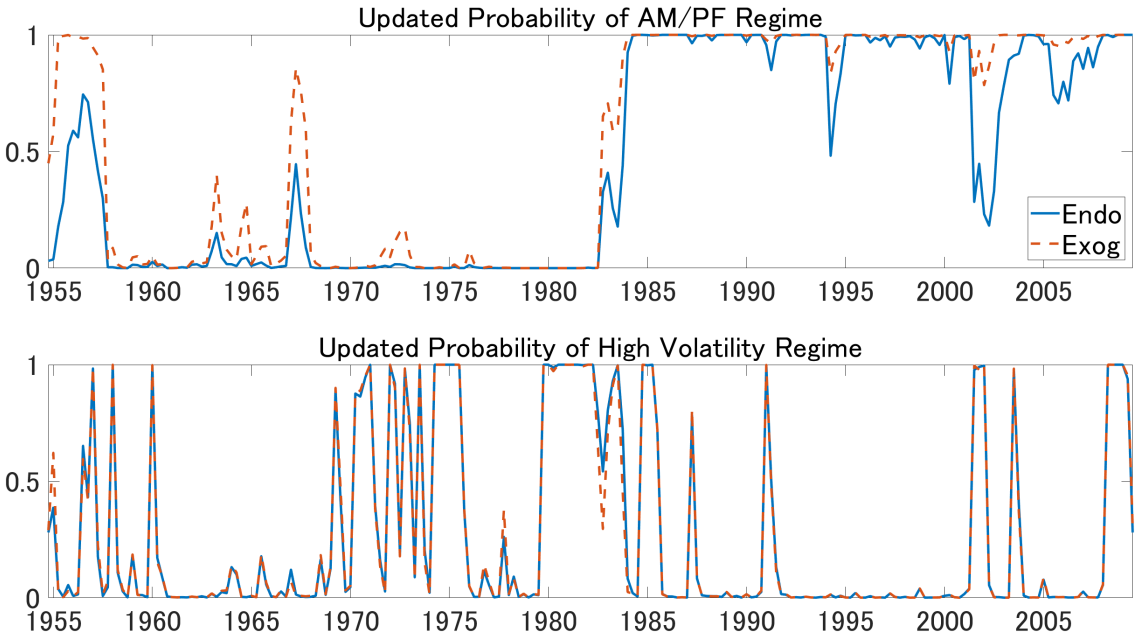

Figure 4: Updated Regime Probabilities from Baseline Model and Exogenous Switching Model

Note: The first panel shows the updated regime probability for AM/PF, and the second panel shows the probability for the high volatility regime, from our baseline model (blue solid line) and the exogenous switching model (orange dashed line).

policy regime. The positive point estimate of λ_R reflects that fiscal discipline is more respected in a high-interest-rate environment.

The relationship between policy regime and the business cycle is captured by λ_y, which is point-estimated to be negative. This implies that the policy authority becomes less concerned with stabilizing debt when output gap is positive–i.e., during expansions. This estimate is not a reflection of pursuing expansionary stimulus in booms, but of allowing debt dynamics to drift during expansions rather than maintaining fiscal discipline.

5.5 Regime Probability

Figure 4 reports the updated regime probabilities at the posterior mean from our baseline model (blue solid line) and the exogenous switching model (orange dashed line). The first panel plots the probability of the AM/PF regime, while the second panel plots the probability of the high-volatility regime. The two models deliver similar volatility regime probabilities. Part of the reason lies in our specification: in our baseline model, policy regimes switch endogenously while volatility regimes are exogenous. We observe a rise in

volatility around the first oil crisis in the early 1970s, the appointment of Volcker as Chair of the Federal Reserve (around 1979-1980), the collapse of the dot-com bubble (around 2000), and the Great Recession (toward the end of the sample).

Turning to the policy regime probabilities in the second panel, the two figures share some characteristics. We observe the PM/AF regime in most of the 1960s and 1970s. In the early 1980s, macroeconomic policy switches to AM/PF and stays mostly toward the end of the sample. The notable differences appear in the late 1950s and late 1960s when the baseline model does not assign high probability to AM/PF, whereas the exogenous switching model does, 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.

We interpret our findings based on historical narratives, with particular attention to the episodes when the two models disagree. In the late 1950s, fiscal policy returned to normalcy after the Korean War by reducing defense expenditures. 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 in Davig and Leeper (2006). These narratives suggest categorizing this period as the combination of passive monetary and passive fiscal policies. The baseline (exogenous switching) model captures the passive fiscal (passive monetary) nature in the regime probability. The policy stance in the late 1960s is also mixed. The increase in the Federal funds rate was related to the concern about surging inflation due mainly to increased fiscal spending related to the Vietnam War. There was a conflict between hawkish monetary policy and hawkish fiscal policy (Blinder, 2022), which might have caused instability in the estimated regime probability in the exogenous switching model.

The narrative episodes favors the PM/AF regime in early 2000s. The policy rate was kept low to support the recovery from the dot-com bubble collapse in the early 2000s despite stable inflation.[13] The dovish monetary policy is accompanied by President George W. Bush's in 2002 and 2004. The high probability of the PM/AF regime from the baseline model aligns more closely with these historical accounts than does the probability implied by the exogenous switching model.

[13]Davig and Leeper (2006) provides the narrative evidence to support categorizing the monetary policy during this period as passive.

5.6 Forecasting

The endogenous regime-switching model is especially useful 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 regime transition probabilities, improving forecasting performance especially around periods of regime change. In particular, because the transition probabilities evolve with economic conditions, the model can provide early signals of potential regime shifts by adjusting the estimated probability of switching in advance of an actual regime change, thereby providing useful information for policy-making and other decision-making purposes.

5.6.1 Transition Probability and Real-Time Forecast

To illustrate this mechanism more closely, we focus on 2001Q2, the period at which the baseline model and the exogenous switching model yield different inferences about the underlying policy regime.: the baseline model detects a switch from AM/PF to PM/AF, while the exogenous switching model indicates no regime change (Figure 4). The top panel of Figure 5 shows the transition probabilities and the updated regime probabilities for 2001-2002. The left panel plots $p(s_t = AM/PF \mid s_{t-1} = AM/PF, \mathcal{F}_{t-1})$, the probability of staying in the AM/PF regime, from the baseline model (solid line labeled "Endo") and the exogenous switching model (dashed line labeled "Exog"), and $p(s_t = AM/PF \mid \mathcal{F}_t)$, the updated AM/PF regime probability from the baseline model (dotted line) for 2001Q1-2002Q1. By construction, the exogenous switching model exhibits a flat transition probability. In contrast, the transition probability under the baseline model declines in 2001Q3, predicting that the regime would be likely to switch from AM/PF to PM/AF given the information up to 2001Q2. Indeed, this is the quarter in which the policy stance changes from AM/PF to PM/AF according to the updated regime probability. Our baseline model successfully predicts the policy regime change at this period.

The ability to predict regime transition probabilities in turn improves forecasting performance for economic variables. To illustrate this, we implement the following forecasting exercise. Using the information available at a selected date, we use the models to forecast real GDP growth and inflation. The bottom two panels of Figure 5 show the four-quarter-ahead forecasts of GDP growth and inflation at 2001Q2, the period just before the regime switch. The forecast generated by the baseline model (solid line labeled "Endo") produces more accurate forecasts than those generated by the exogenous switch-

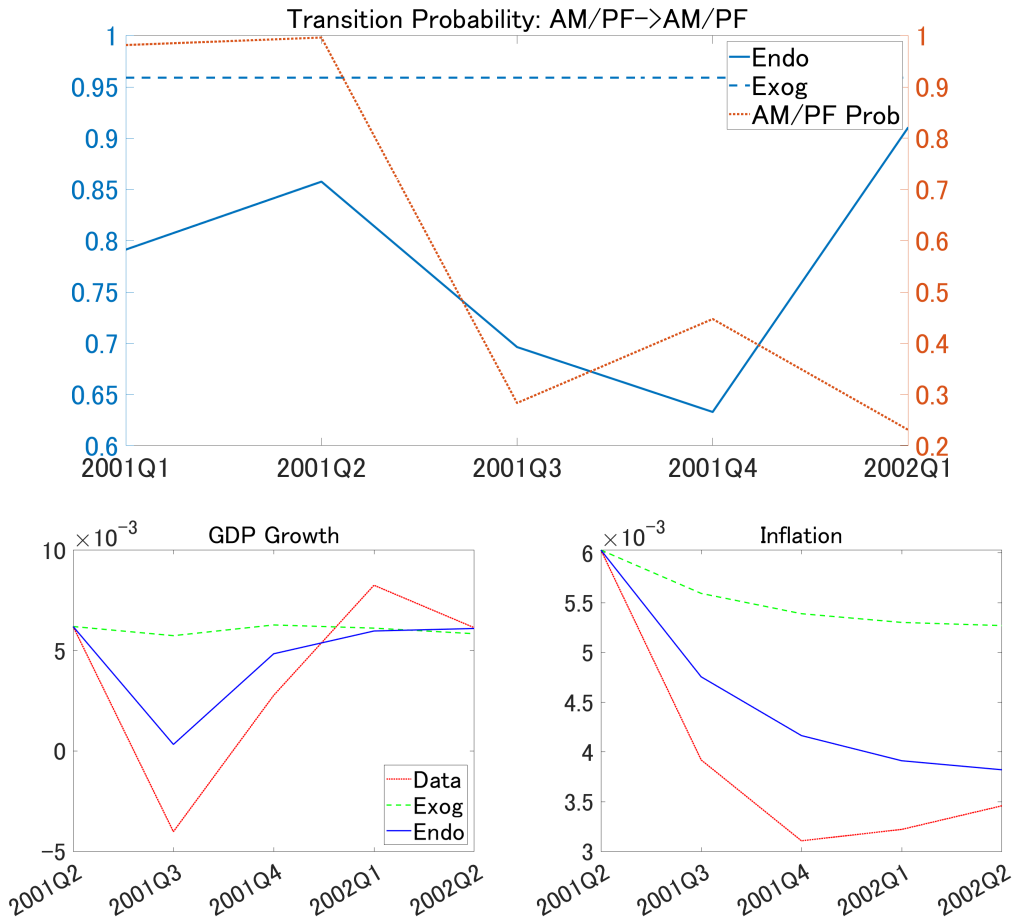

Figure 5: Transition Probability and Forecast at 2001Q2

Note: The top 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 bottom two panels plot the 4-quarter-ahead forecast of GDP growth and inflation rate given the information at 2001Q2, along with the realizations of them.

ing model (dashed line labeled "Exog"), especially at one- to two-quarter horizons. The superior performance of the baseline model reflects its ability to capture time variation in the transition probabilities, a feature absent in the exogenous switching specification.

5.6.2 Overall Forecasting Performance

The discussion above was based on a single period in the sample. To investigate the overall forecasting performance of the model, rows (a)–(c) of Table 6 compare root mean square errors (RMSEs) using the ratio of the RMSE from the baseline model to that from the exogenous switching model for three variables: GDP growth, inflation, and Federal Funds rate. At the one-quarter-horizon, the exogenous switching model provides better

Table 6: Ratio of RMSE and Maximum Absolute Error

		GDP Growth	Inflation	Fed Funds Rate
(a)	RMSE (1Q)	1.034	1.020	0.999
(b)	RMSE (4Q)	1.004	1.023	0.985
(c)	RMSE (8Q)	0.999	1.016	0.979
(d)	Maximum Absolute Error (1Q)	0.990	0.987	0.999

Note: This table shows the ratios of RMSEs and maximum absolute error from the baseline model to these from the exogenous switching model for GDP growth, inflation, and Federal Funds rate.

forecasts except for the Fed Funds rate, as is reflected in the RMSE ratios larger than one.[14] As the forecasting horizon gets longer, the performance of the baseline model relative to the exogenous switching model improves (except for the inflation from one-quarter-ahead to four-quarter-ahead), and the baseline model exhibits the better forecast for GDP and Fed Funds rate at eight quarter horizon. This implies that the endogenous switching feature of the model can deliver equal or lower average forecast errors at the medium to long run.

Although the baseline model is associated with larger average forecast error in the short run, it produces smaller extreme (worst-case) forecast errors than the exogenous switching model. Indeed, row (d) of Table 6 shows that ratios of maximum absolute errors from the baseline and exogenous switching models are all below one. The baseline model is more robust in terms of tail errors, reducing the magnitude of the worst-case forecasting errors across GDP growth, inflation, and the Federal Funds rate.

6 Conclusion

This paper develops state-space models with regime-switching coefficients allowing for feedback from lagged continuous state variables into regime determination. We 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. Our framework generalizes conventional models with time-invariant regime transition probabilities if we shut out the feedback channel.

[14] As an alternative measure of the model fitness, the log marginal likelihood from these models is computed to be 6060.54 for the baseline model and 6063.74 for the exogenous switching model. The observation that the exogenous switching model provides the larger marginal likelihood is consistent with the one-quarter ahead RMSEs, because the marginal likelihood accumulates the sequential one-step-ahead predictive densities. However, the larger marginal likelihood—i.e., better one-period ahead forecasting performance—does not necessarily imply better forecasting performance at the medium to long run.

To circumvent the path dependence problem, we truncate the history of regimes that must be tracked. We prove that the likelihood from the filter with such a truncation step is asymptotically equivalent to the one from the exact filter if the truncation order increases with the sample size. The approximation error decreases exponentially with the truncation length, with the speed of convergence governed by the persistence of the state dynamics. This result provides theoretical guidance for practical implementation: in moderately persistent systems, small truncation orders deliver accurate likelihood evaluations while keeping the computational burden manageable. The theoretical findings are supported by simulation exercises based on both a simple univariate model and a state-space representation derived from an empirical New Keynesian DSGE framework.

Empirically, we study the monetary–fiscal policy mix in the post-war United States using a regime-switching DSGE model. We uncover evidence of systematic feedback from key macroeconomic and policy indicators to policy regime transitions. In particular, we find higher debt increases the likelihood of the active monetary/passive fiscal regime, consistent with a stabilization motive of the consolidated government. Another important implication of endogenous regime determination is its relevance for forecasting. Because transition probabilities vary over time as economic conditions evolve, the model generates early signals of potential regime changes. The time-varying transition probabilities also provide equal or superior average forecasting performance to that from the exogenous switching model in the medium to long run, and reduces the size of worst-case forecasting errors.

While we focus on monetary-fiscal interaction, the methodology is readily applicable to other structural parameters that may change discretely over time. For example, one could study regime-dependent financial frictions, shifts in macroeconomic volatility, or changes in the conduct of macroprudential policy. These promising applications are left for future work.

References

Angeletos, George-Marios, Christian Hellwig, and Alessandro Pavan (2007) "Dynamic global games of regime change: Learning, multiplicity, and the timing of attacks," *Econometrica*, 75 (3), 711–756.

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."

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.

Bognanni, Mark and Edward Herbst (2018) "A sequential Monte Carlo approach to inference in multiple-equation Markov-switching models," *Journal of Applied Econometrics*, 33 (1), 126–140.

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) "Fluctuating Macro Policies and the Fiscal Theory," *NBER Macroeconomics Annual*, 21, 247–315.

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.

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

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

Hashimzade, Nigar, Oleg Kirsanov, Tatiana Kirsanova, and Junior Maih (2025) "Filtering and Smoothing in Regime-Switching DSGE Models."

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.

Powell, Jerome H. (2025) "Monetary Policy and the Fed's Framework Review," `https://www.federalreserve.gov/newsevents/speech/powell20250822a.htm`, Speech at "Labor Markets in Transition: Demographics, Productivity, and Macroeconomic Policy," Jackson Hole Economic Policy Symposium.

A Filter for Regime-Switching State-Space Model

The regime-switching state-space model is given by

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

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

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

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.

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

We re-define $x_t = [x_t', x_{t-1}']'$, and so forth. 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 $\bar{x}_{0|0}(i)$ and $\bar{\Omega}_{0|0}(i)$ for $i = 0, 1$. We also set the initial regime probability $p_r(s_0 = i)$ for $i = 0, 1$.
- For $t = 1, \cdots, T$,

  (i) Forecasting Step

$$\bar{x}_{t|t-1}(s_{t-r+1}^t) = A_{s_t} \bar{x}_{t-1|t-1}(s_{t-r+1}^{t-1}) \tag{6}$$
$$\bar{\Omega}_{t|t-1}(s_{t-r+1}^t) = A_{s_t} \bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-1}) A_{s_t}' + Q_{s_t} Q_{s_t}' \tag{7}$$
$$\bar{y}_{t|t-1}(s_{t-r+1}^t) = B_{s_t} \bar{x}_{t|t-1}(s_{t-r+1}^t) \tag{8}$$
$$\bar{\Sigma}_{t|t-1}(s_{t-r+1}^t) = B_{s_t} \bar{\Omega}_{t|t-1}(s_{t-r+1}^t) B_{s_t}' + R_{s_t} R_{s_t}' \tag{9}$$

(ii) Calculating Likelihood

$$p_r(y_t|\mathcal{F}_{t-1}) = \sum_{s^t_{t-r+1}} p_r(y_t|s^t_{t-r+1}, \mathcal{F}_{t-1}) p_r(s_t|s^{t-1}_{t-r+1}, \mathcal{F}_{t-1}) p_r(s^{t-1}_{t-r+1}|\mathcal{F}_{t-1}) \quad (10)$$

The first part is given by $y_t|s^t_{t-r+1}, \mathcal{F}_{t-1} \sim N\left(\bar{y}_{t|t-1}(s^t_{t-r+1}), \bar{\Sigma}_{t|t-1}(s^t_{t-r+1})\right)$. The third part is obtained from equation (13) at the previous iteration. The second part is the transition probability from s^{t-1}_{t-r+1} to s_t conditional on the information set $\mathcal{F}_{t-1}$, which will be elaborated on later.

(iii) Updating step

$$\begin{aligned} &\bar{x}_{t|t}(s^t_{t-r+1}) \\ =&\bar{x}_{t|t-1}(s^t_{t-r+1}) + \bar{\Omega}_{t|t-1}(s^t_{t-r+1}) B'_{s_t} \left[\bar{\Sigma}_{t|t-1}(s^t_{t-r+1})\right]^{-1} \left(y_t - \bar{y}_{t|t-1}(s^t_{t-r+1})\right) \end{aligned} \quad (11)$$

$$\begin{aligned} &\bar{\Omega}_{t|t}(s^t_{t-r+1}) \\ =&\bar{\Omega}_{t|t-1}(s^t_{t-r+1}) - \bar{\Omega}_{t|t-1}(s^t_{t-r+1}) B'_{s_t} \left[\bar{\Sigma}_{t|t-1}(s^t_{t-r+1})\right]^{-1} B_{s_t} \bar{\Omega}_{t|t-1}(s^t_{t-r+1}) \end{aligned} \quad (12)$$

and

$$p_r(s^t_{t-r+2}|\mathcal{F}_t) = \sum_{s_{t-r+1}} p_r(s^t_{t-r+1}|\mathcal{F}_t) \quad (13)$$

where

$$p_r(s^t_{t-r+1}|\mathcal{F}_t) = \frac{p_r(y_t|s^t_{t-r+1}, \mathcal{F}_{t-1}) p_r(s^t_{t-r+1}|\mathcal{F}_{t-1})}{p_r(y_t|\mathcal{F}_{t-1})} \quad (14)$$

(iv) Truncation

$$\bar{x}_{t|t}(s^t_{t-r+2}) = \sum_{s_{t-r+1}} \frac{p_r(s^t_{t-r+1}|\mathcal{F}_t)}{p_r(s^t_{t-r+2}|\mathcal{F}_t)} \bar{x}_{t|t}(s^t_{t-r+1}) \quad (15)$$

$$\bar{\Omega}_{t|t}(s^t_{t-r+2}) = \sum_{s_{t-r+1}} \frac{p_r(s^t_{t-r+1}|\mathcal{F}_t)}{p_r(s^t_{t-r+2}|\mathcal{F}_t)} \bar{\Omega}_{t|t}(s^t_{t-r+1}) \quad (16)$$

Remark 2. When $r = 1$, s^t_{t-r+2} 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 3. Kim (1994) includes the second order adjustment term in (16). As in Li (2023), we ignore that term because it is not obvious whether this additional term improves the approximation, 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.

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 (10).

$$\begin{aligned}
& p_r(s_t = 0|s_{t-1} = 0, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}) \\
=& \frac{p_r\left(s_t = 0, s_{t-1} = 0 \mid s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}\right)}{p_r\left(s_{t-1} = 0 \mid s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}\right)} \\
=& \frac{\int p_r(s_t = 0, s_{t-1} = 0|x_{t-1}, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}) p_r\left(x_{t-1}|s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}\right) dx_{t-1}}{\int p_r(s_{t-1} = 0|x_{t-1}, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}) p_r\left(x_{t-1}|s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}\right) dx_{t-1}} \\
\approx& \frac{\Phi_2\left(-\tau\iota; \Lambda\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2}), I + \Lambda\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\Lambda'\right)}{\Phi_1\left(-\tau; \lambda'\left(\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x)}, 1 + \lambda'\left(\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x, d_x+1:2d_x)}\lambda\right)}
\end{aligned}$$

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

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

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

for $r \geqslant 3$. If $r \leqslant 2$, we use $\bar{x}_{t-1|t-1}$ and $\bar{\Omega}_{t-1|t-1}$ defined in Remark A1.

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

$$\begin{aligned} &p_r(s_t = 1|s_{t-1} = 1, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}) \\ \approx &\frac{\Phi_2\left(\tau\iota; -\Lambda\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2}), I + \Lambda\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\Lambda'\right)}{\Phi_1\left(\tau; -\lambda'\left(\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x)}, 1 + \lambda'\left(\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x, d_x+1:2d_x)}\lambda\right)} \end{aligned}$$

The remaining probabilities are

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

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

$$p_r(s_t = 0|s_{t-1} = 0, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}) \approx \frac{\Phi_2\left(-\tau\iota; \lambda'\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2}), \Sigma_\eta + \Lambda\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\Lambda'\right)}{\Phi_1\left(-\tau; \lambda'\left(\bar{x}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x)}, \frac{1}{1-\rho^2} + \lambda'\left(\bar{\Omega}_{t-1|t-1}(s_{t-r+1}^{t-2})\right)_{(d_x+1:2d_x, d_x+1:2d_x)}\lambda\right)}$$

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

$$\begin{aligned} p(s_t = 0|s_{t-1} = 0, x_{t-1}) &= \frac{p(s_t = 0, s_{t-1} = 0|x_{t-1})}{p(s_{t-1} = 0|x_{t-1})} \\ &= \frac{p([\eta_t, \eta_{t-1}]' \leqslant -\tau\iota - \Lambda x_{t-1}|x_{t-1})}{p(\eta_{t-1} \leqslant -\tau - \lambda' x_{t-2}|x_{t-1})} \\ &= \frac{\Phi_2\left(-\tau\iota - \Lambda x_{t-1}; 0, \Sigma_\eta\right)}{\Phi_1(-\tau - \lambda' x_{t-2}; 0, (1-\rho^2)^{-1})} \end{aligned}$$

and

$$p(s_t = 1|s_{t-1} = 1, x_{t-1}) = \frac{\Phi_2\left(\tau\iota + \Lambda x_{t-1}; 0, \Sigma_\eta\right)}{\Phi_1\left(\tau + \lambda' x_{t-2}; 0, (1-\rho^2)^{-1}\right)}$$

B Proofs

B.1 Proofs for Section 3

B.1.1 Auxiliary Lemmas

Lemma 2. Assume Assumptions 1 and 2. There exist positive constants c_Ω^+ and c_Ω^- such that $c_\Omega^- \leqslant \|\Omega_{t|t}(s_1^t)\| \leqslant c_\Omega^+$ 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 $\frac{\alpha}{1+\alpha\beta} I \leqslant \Omega_{t|t}(s_1^t) \leqslant \frac{1+\alpha\beta}{\alpha} I$. Taking the matrix norm for both hand sides establishes our claim. □

Lemma 3. Assume Assumptions 1 and 2. There exist a positive constant c_x such that for any $t = 1, 2, \cdots$,

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

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

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

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

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

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

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 4. 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,

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

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

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

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 $\lambda_{min}(A+B) \geqslant \lambda_{min}(A)$. Then, $\|(A+B)^{-1}\| \leqslant \|A^{-1}\|$. Therefore, we have

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

□

Lemma 5 (Corollary 2.14 in Ipsen and Rehman (2008))**.** Let A and E be $n \times n$ matrices. If A is nonsingular, then

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

Lemma 6 (Equation (X.2) in Bhatia (1997))**.** For $0 \leqslant r \leqslant 1$ and positive semidefinite matrices A and B,

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

Lemma 7. For any $x \in \mathbb{R}$ and $\varepsilon \geqslant 0$, we have $\Phi_1(x+\varepsilon) - \Phi_1(x) \leqslant \varepsilon$.

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 $\phi_1(x^*) \leqslant \phi_1(0) < 1$. □

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 8. Assume Assumptions 1 and 2. For $t \geqslant r$, 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

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

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 2 and 1 that

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

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

Lemma 9. Assume Assumptions 1 and 2. For $t \geqslant r$, define

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

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

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

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 $x^*_{t|t}(s_{t-r+2}^t, s_1^{t-r})$. We can decompose $\delta_t^x(\cdot)$ as

$$\delta_t^x(s_{t-r+2}^t, s_1^{t-r}) \leqslant \left\| x_{t|t}(s_1^t\langle i\rangle) - x^*_{t|t}(s_{t-r+2}^t, s_1^{t-r}) \right\| + \left\| x^*_{t|t}(s_{t-r+2}^t, s_1^{t-r}) - x_{t|t}(s_1^t\langle j\rangle) \right\| \tag{17}$$

The first term can be evaluated as

$$\begin{aligned}
&\left\| x_{t|t}(s_1^t\langle i\rangle) - x^*_{t|t}(s_{t-r+2}^t, s_1^{t-r}) \right\| \\
\leqslant &\left\| \Psi^{r-1}(s_1^t\langle i\rangle) \right\| \times \left\| x_{t-r+1|t-r+1}(s_{t-r+1} = i, s_1^{t-r}) - x_{t-r+1|t-r+1}(s_{t-r+1} = j, s_1^{t-r}) \right\| \\
\leqslant &\exp(-c_2(r-1)) m_{1,t}
\end{aligned} \tag{18}$$

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 (17) as

$$\begin{aligned}
&\left\| x^*_{t|t}(s_{t-r+2}^t, s_1^{t-r}) - x_{t|t}(s_1^t\langle j\rangle) \right\| \\
= &\Bigg\| \left(\Psi^{r-1}(s_1^t\langle i\rangle) - \Psi^{r-1}(s_1^t\langle j\rangle) \right) x_{t-r+1|t-r+1}(s_{t-r+1} = j, s_1^{t-r}) \\
&+ \sum_{k=1}^{r-1} \left(\Psi^{k-1}(s_1^t\langle i\rangle) K\left(s_1^{t-k+1}\langle i\rangle\right) - \Psi^{k-1}(s_1^t\langle j\rangle) K\left(s_1^{t-k+1}\langle j\rangle\right) \right) y_{t-k+1} \Bigg\| \\
\leqslant &\left\| \Psi^{r-1}(s_1^t\langle i\rangle) - \Psi^{r-1}(s_1^t\langle j\rangle) \right\| \times \left\| x_{t-r+1|t-r+1}(s_{t-r+1} = j, s_1^{t-r}) \right\| \\
&+ \sum_{k=1}^{r-1} \left\| \left(\Psi^{k-1}(s_1^t\langle i\rangle) K\left(s_1^{t-k+1}\langle i\rangle\right) - \Psi^{k-1}(s_1^t\langle j\rangle) K\left(s_1^{t-k+1}\langle j\rangle\right) \right) \right\| \times \| y_{t-k+1} \|
\end{aligned} \tag{19}$$

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,

$$\begin{aligned}
&\Psi^{k-1}(s_1^t\langle i\rangle) K\left(s_1^{t-k+1}\langle i\rangle\right) - \Psi^{k-1}(s_1^t\langle j\rangle) K\left(s_1^{t-k+1}\langle j\rangle\right) \\
= &\Psi^{k-1}(s_1^t\langle i\rangle) \left[K\left(s_1^{t-k+1}\langle i\rangle\right) - K\left(s_1^{t-k+1}\langle j\rangle\right) \right] + \left[\Psi^{k-1}(s_1^t\langle i\rangle) - \Psi^{k-1}(s_1^t\langle j\rangle) \right] K\left(s_1^{t-k+1}\langle j\rangle\right)
\end{aligned} \tag{20}$$

Let $B_+ = \max_{s=0,1} \|B_s\|$ and $R_- = \max_{s=0,1} \|(R_s R_s')^{-1}\|$. By Lemma 4,

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

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

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

where $c_{\delta K} = c_\delta B_+ R_- [1 + B_+ R_- (B_+ + 2c_\Omega)]$. Together with Lemma 1, the whole first term of equation (20) is bounded by

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

where $c_3 = \max\{c_1, c_K\}$ is a positive constant. For the second term of (20), note that

$$\begin{aligned}
\Psi(s_1^t\langle i\rangle) - \Psi(s_1^t\langle j\rangle) &= \left(I - K(s_1^t\langle i\rangle)B_{s_t}\right) A_{s_t} - \left(I - K(s_1^t\langle j\rangle)B_{s_t}\right) A_{s_t} \\
&= \left(-K(s_1^t\langle i\rangle) + K(s_1^t\langle j\rangle)\right) B_{s_t} A_{s_t}
\end{aligned} \tag{21}$$

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

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

where $c_4 = c_1^2 c_{\delta K} \max_s\{\|B_s\| \cdot \|A_s\|\}$. Then, the second term of (20) is bounded by

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

Substituting them all together in equation (19) yields

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

Using this and (18), equation (17) becomes

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

where $m_t = m_{1,t} + m_{2,t}$. It remains to see the stochastic boundedness of m_t. By Lemma 3, $\mathbb{E}_0|m_{1,t}| \leqslant c_1(\|x_{t-r+1|t-r+1}(i, s_1^{t-r})\| + \|x_{t-r+1|t-r+1}(j, s_1^{t-r})\| \leqslant 2c_1c_x < \infty$. For $m_{2,t}$,

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

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 8 and 9, we can establish the proposition claiming that the updated mean and variance from the truncated and exact filters are asymptotically equivalent.

Proposition 2. Let $\Delta_{r,t}^{\Omega}(s_1^t) = \bar{\Omega}_{t|t}(s_{t-r+2}^t) - \Omega_{t|t}(s_1^t)$ and $\Delta_{r,t}^{x}(s_1^t) = \bar{x}_{t|t}(s_{t-r+2}^t) - x_{t|t}(s_1^t)$. 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

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

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 8 and 9 yield

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

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

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

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 1. 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

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

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 3. There exists a positive stochastically bounded random variable N_t such that

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

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

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

Without loss of generality, let $\det\bar{\Sigma}_{t|t-1}(s_{t-r+2}^t) \geqslant \det\Sigma_{t|t-1}(s_1^t)$. Due to $|\log x - \log y| \leqslant \frac{|x-y|}{x\wedge y}$ and Lemma 5, the first term can be written as

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

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 3. 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

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

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

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

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.

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

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

Proposition 4. There exists a positive stochastically bounded random variable $N_{TP,t}$ such that

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

Proof. Consider $s_t = 0$ and $s_{t-1} = 0$. The probabilities for other combinations of (s_t, s_{t-1}) can be considered analogously. 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

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

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

$$\begin{aligned} & p_r(s_t = 0|s_{t-1} = 0, s_{t-r+1}^{t-2}, \mathcal{F}_{t-1}) = p(s_t = 0|s_{t-1} = 0, s_1^{t-2}, \mathcal{F}_{t-1}) \\ & + \left(\nabla f(\mu_{t-1|t-1}^{\dagger}(s_1^{t-2}), \text{vech}(V^{\dagger}(s_1^{t-2})^{1/2}))\right)' \\ & \begin{bmatrix} \bar{\mu}_{t-1|t-1}(s_{t-r+1}^{t-2}) - \mu_{t-1|t-1}(s_1^{t-2}) \\ \text{vech}(\bar{V}_{t-1|t-1}(s_{t-r+1}^{t-2})^{1/2}) - \text{vech}(V_{t-1|t-1}(s_1^{t-2})^{1/2}) \end{bmatrix} \end{aligned} \tag{22}$$

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

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

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

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 (22) is $O_p(1)$.

Using Proposition 2, it follows that

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

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

dom variable.

Let $\|\cdot\|_1$, $\|\cdot\|_F$, and $\|\cdot\|$ be the 1-norm. We can deduce

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

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

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

where the third inequality follows from Lemma A5.

We simply denote $\mu^\dagger = \mu^\dagger_{t-1|t-1}(s_1^{t-2})$ and $V^\dagger = V^\dagger(s_1^{t-2})^{1/2})$. By $\log(1+x) \leqslant x$ for $x > 0$, we have

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

□

B.1.5 Proof of Proposition 1

By (4), we have

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

By Proposition 3, the term inside the first summation is bounded by $\exp(-c_2(r-1))N_t$. By Proposition 4, the term inside the second summation is bounded by $\exp(-c_2(r-1))N_{TP,t}$.

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: $\tilde{e}_t = \tilde{e}_t^L + \tilde{e}_t^S$ where

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

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

$$Y_t = C_t + G_t$$

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.

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

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. We 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

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

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

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

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

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

which satisfies

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

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

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

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 5. See Herbst and Schorfheide (2014) and Herbst and Schorfheide (2016) for more details.

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 charac-

terized by the particles $\{\theta^i_{n-1}, W^i_n\}^N_{i=1}$. 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 $\{\theta_{n,b}\}^{N_{blocks}}_{b=1}$. For any $i = 1, \cdots, N$, run the block Metropolis-Hastings algorithm for N_{MH} times using the proposal distribution

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

where $\theta^i_{n,b,m-1}$ is the parameter from the previous iteration of the MH, $\theta^i_{n,-b,m}$ 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^i_n.

C.3.2 Hyperparameters

The hyperparameters we have to choose a priori are $(N, N_\phi, N_{blocks}, N_{MH}, \omega, \{\phi_n\}^{N_\phi}_{n=0})$. 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 $\lambda = 2$.

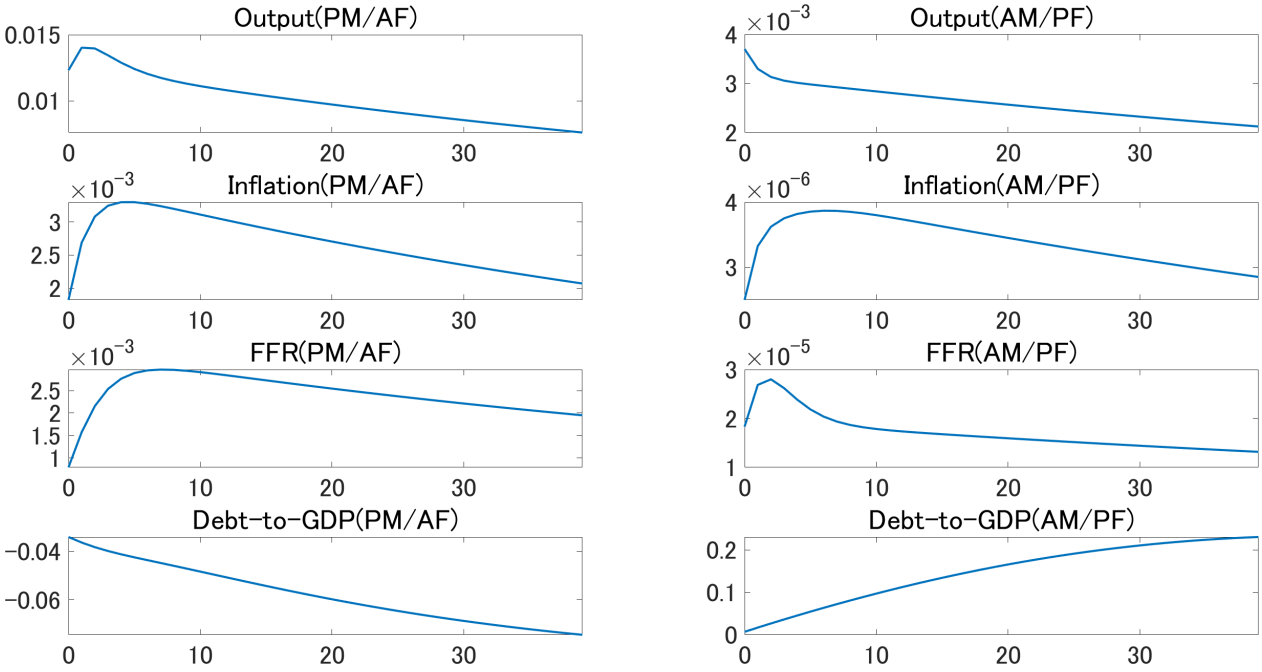

Figure 6: Impulse Response Functions to a Long-Term Expenditure Shock

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. A reader may find the discussion overlapping with Bianchi and Ilut (2017) because the economic intuition remains similar.

C.4.1 Long-Term Expenditure Shock

Figure 6 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. 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 govern-

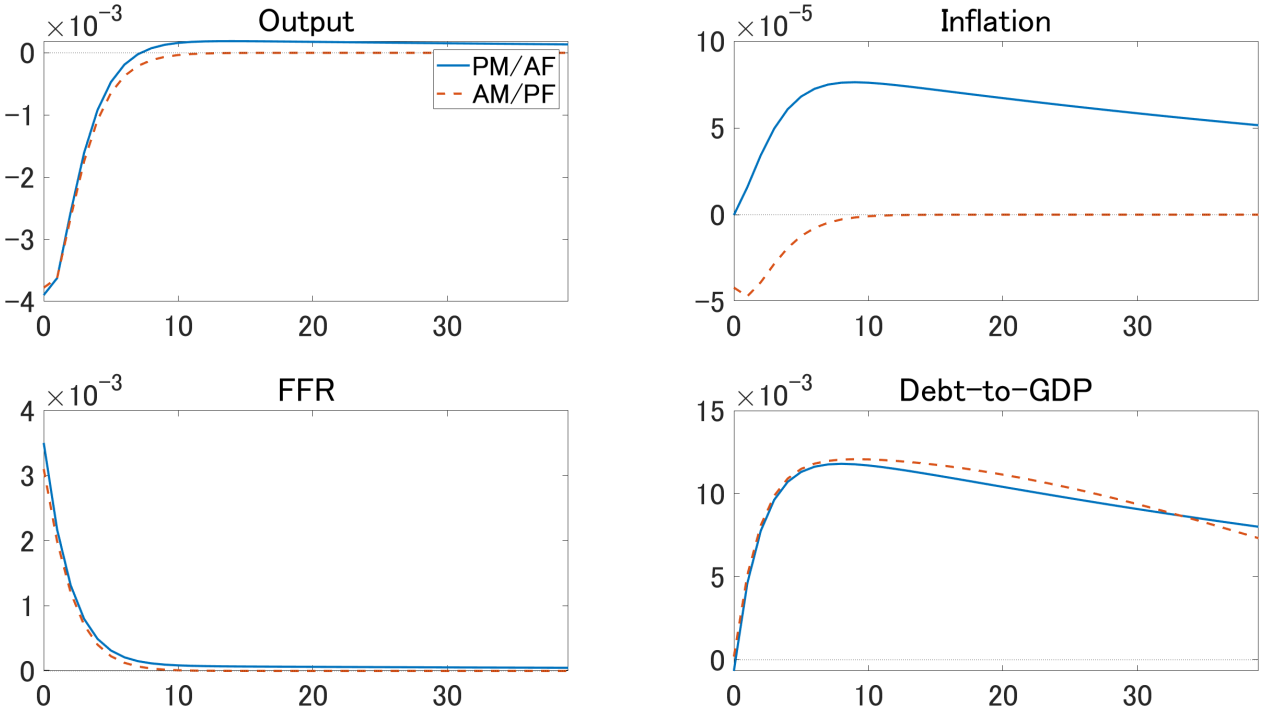

Figure 7: Impulse Response Functions to A Monetary Policy Shock

ment 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-output ratio.

C.4.2 Monetary Policy Shock

Figure 7 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 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.

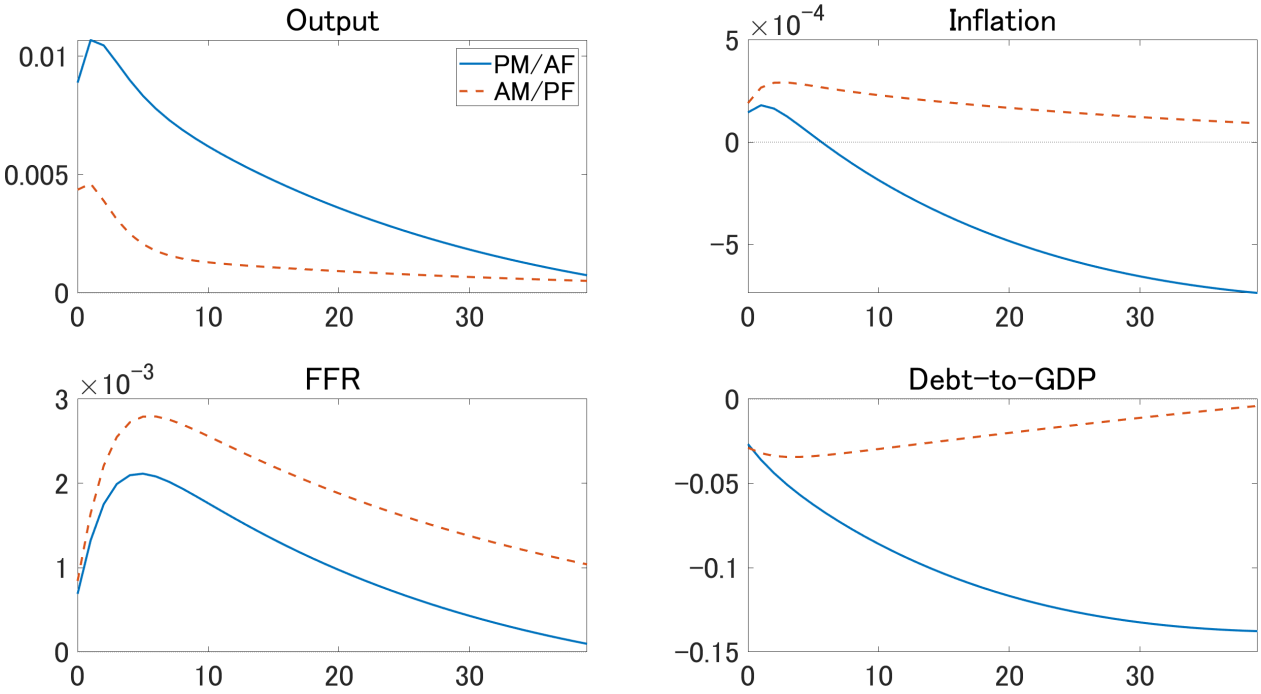

Figure 8: Impulse Response Functions to A Preference Shock

C.4.3 Preference Shock

The impulse responses to a positive preference shock are shown in Figure 8. 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.