University of Glasgow

Adam Smith Business School

WORKING PAPER SERIES

On Bayesian Filtering for Markov Regime Switching Models

Nigar Hashimzade, Oleg Kirsanov, Tatiana Kirsanova, Junior Maih.

Paper No. 2024-01
February 2024

On Bayesian Filtering for Markov Regime Switching Models[*]

Nigar Hashimzade[†] Oleg Kirsanov[‡] Tatiana Kirsanova[§] Junior Maih[¶]

February 1, 2024

Abstract

This paper presents a framework for empirical analysis of dynamic macroeconomic models using Bayesian filtering, with a specific focus on the state-space formulation of New Keynesian Dynamic Stochastic General Equilibrium (NK DSGE) models with multiple regimes. We outline the theoretical foundations of model estimation, provide the details of two families of powerful multiple-regime filters, IMM and GPB, and construct corresponding multiple-regime smoothers. A simulation exercise, based on a prototypical NK DSGE model, is used to demonstrate the computational robustness of the proposed filters and smoothers and evaluate their accuracy and speed. We show that the canonical IMM filter is faster than the commonly used Kim and Nelson (1999) filter and is no less, and often more, accurate. Using it with the matching smoother improves the precision in recovering unobserved variables by about 25%. Furthermore, applying it to the U.S. 1947-2023 macroeconomic time series, we successfully identify significant past policy shifts including those related to the post-Covid-19 period. Our results demonstrate the practical applicability and potential of the proposed routines in macroeconomic analysis.

Keywords: Markov switching models, Filtering, Smoothing
JEL Reference Number: C11, C32, C54, E52

[*]This paper should not be reported as representing the views of Norges Bank. The views expressed are those of the authors and do not necessarily reflect those of Norges Bank.

[†]Department of Economics and Finance, Brunel University, United Kingdom; e-mail: nigar.hashimzade@brunel.ac.uk

[‡]Adam Smith Business School, University of Glasgow, United Kingdom; e-mail: oleg.kirsanov@glasgow.ac.uk

[§]Adam Smith Business School, University of Glasgow, United Kingdom; e-mail: tatiana.kirsanova@glasgow.ac.uk

[¶]Norges Bank; e-mail Junior.Maih@norges-bank.no

1 Introduction

In the evolving landscape of macroeconomic analysis, the empirical examination of dynamic models has become increasingly sophisticated and computationally demanding. This paper contributes to this area by presenting a comprehensive framework for the empirical analysis of dynamic macroeconomic models using Bayesian filtering. We focus on New Keynesian Dynamic Stochastic General Equilibrium (NK DSGE) models with multiple regimes. Our work introduces enhanced filter and smoother algorithms, crucial for accurate macroeconomic modelling and estimation.

Our study is motivated by the increasing popularity of Bayesian methods in macroeconomic time series analysis, particularly within the DSGE framework. These methods have gained traction due to their ability to effectively handle complex models with latent variables and structural changes. Bayesian perspective is invaluable for disentangling convoluted macroeconomic phenomena, such as differentiating between external shocks and policy-driven economic patterns.

Despite significant advancements in the literature, the field continues to face various challenges, especially when estimating macroeconomic dynamic models with multiple regimes. One such challenge is selecting an efficient and accurate filter for likelihood computation. Another challenge is the task of reconstructing latent variables through the smoothing of estimated state variables and regime probabilities.

The prevalent use of the Kim and Nelson (Kim, 1994, Kim and Nelson, 1999) filter in macroeconomic applications (see, *inter alia*, Davig and Doh, 2014, Chang, Maih, and Tan, 2021, Chen, Leeper, and Leith, 2022) suggests limited exploration of alternative methods in this field. Despite its unquestionable power, Kim and Nelson filter is known to have certain flaws. Namely, it is computationally intensive and, when extended to smoothing algorithms, computationally unstable. Perhaps, the latter is the reason for scant use of smoothing for more accurate recovery of latent variables in multiple regime models in the existing economic literature.

Our paper makes both theoretical and empirical contributions in this domain. First, we introduce and extend the Interactive Multiple Model (IMM) filter, originally developed by Bar-Shalom (Blom and Bar-Shalom, 1988). Despite its recognition in the engineering literature, the IMM filter remains underutilised in economic applications. In addition, we extend the Kim and Nelson filter, to accommodate different orders of approximation. Finally, we develop a computationally stable and easily implementable smoothing algorithm that can be conveniently adapted to a wide range of filters in multiple regime setting. Empirically, we apply these methods to a prototypical NK DSGE model and the U.S. macroeconomic time series spanning from 1947 to 2023. This exercise succeeds in identifying significant policy shifts, particularly in the post-Covid-19 era, and thus demonstrates the practical relevance of our methods.

We validate the superiority of the proposed filter-smoother algorithm using rigorous simulation exercises. Our findings indicate that the IMM filter outperforms the Kim and Nelson filter in terms of computational speed while maintaining comparable accuracy. Moreover, the implementation of our proposed smoother significantly enhances the precision in the recovery of latent variables, with an approximate 25% reduction in estimation errors. These empirical insights reveal the importance of smoothing in this framework, overlooked in the existing literature.

One should note that while the Kim and Nelson filter has dominated the analysis of multiple-regime macroeconomic models, there have been a few exceptions. Liu, Wang, and Zha (2013) apparently applied IMM to study the role of land-price dynamics in macroeconomy. Binning and Maih (2015) showed how sigma-point filters can be adapted to the multiple regime setting based on the IMM, using a prototypical NK DSGE model. Bjørnland, Larsen, and Maih (2018) applied it to study the interplay between oil price shocks and macroeconomic instability. More recently, Leith, Kirsanova, Machado, and Ribeiro (2024) used IMM in a study of monetary and fiscal policy changes in the United States. We are unaware of other IMM applications in macroeconomics to date.

All computations presented in this paper were implemented in the RISE© toolbox (Maih, 2015).[1]

The paper is organised as follows. The next section presents theoretical foundations. We derive two families of filters, one of which encompasses the Kim and Nelson filter and the other one encompasses the canonical IMM. We derive a Markov-switching smoother adapted to the appropriate filter family. Section 3 tests the efficacy of the proposed filter and smoother algorithms on artificial data. An empirical investigation is presented in Section 4. Section 5 concludes.

[1] RISE stands for 'Rationality in Switching Environments'. The codes and documentation are available at https://github.com/jmaih/RISE_toolbox

2 Switching Filters and Smoothers

2.1 The Filtering Problem

We start with a general multiple-regime state-space representation of a linear discrete-time dynamic model consisting of a measurement equation (1) and a transition equation (2),

$$
\begin{aligned}
y_t &= c_{y,s_t} + Z_{s_t}\alpha_t + g_{s_t}\varepsilon_t, \qquad (1)\\
\alpha_t &= c_{\alpha,s_t} + T_{s_t}\alpha_{t-1} + R_{s_t}\eta_t, \qquad (2)
\end{aligned}
$$

where y_t a $p \times 1$ vector of observations, α_t is a $m \times 1$ vector of unobserved state variables, and ε_t and η_t are independent standard Gaussian random variables, $t = 1, \ldots, n$. All model parameters, $\{c_{y,s_t}, c_{\alpha,s_t}; Z_{s_t}, T_{s_t}; g_{s_t}, R_{s_t}\}$, depend on regime s_t, which is an outcome of a Markov process with $h \geq 1$ discrete regimes. This process is described by the transition probability matrix with the generic element $Q(s_{t-1}, s_t) = \Pr[s_t \mid s_{t-1}]$, so that $\sum_{s_t=1}^{h} Q(s_{t-1}, s_t) = 1$ for every regime s_{t-1} and every time t.

The information available at time t is fully contained in the vector of observations $Y_t := \{y_1, ..., y_t\}$. The object of interest is an estimate of the unobserved state vector α_t, for which three estimators, $\alpha_{t|t-1}$, $\alpha_{t|t}$ and $\alpha_{t|n}$ are available in Bayesian framework. The first estimator is the forecast of α_t based on information Y_{t-1},

$$\alpha_{t|t-1} := \mathbb{E}\left[\alpha_t \mid Y_{t-1}\right].$$

Its mean square error (MSE) is defined as

$$P_{t|t-1} := \mathbb{E}\left[\left(\alpha_t - \alpha_{t|t-1}\right)\left(\alpha_t - \alpha_{t|t-1}\right)' \mid Y_{t-1}\right].$$

In the linear single-regime setting with Gaussian shocks these objects and the associated likelihood $f(y_t \mid Y_{t-1})$ are computed by the well-established technique of the standard Kalman filter (KF), which in this case is exact and optimal (Kalman, 1960). Working in a multiple-regime environment is more challenging because of the explosive dimensionality of the problem.

Specifically, in a multiple-regime environment, exact estimation is infeasible because the number of histories that a Kalman-type filter needs to take into account increases exponentially with every time period. At any given time t, a multiple-regime dynamic system can be in one of h possible regimes, each corresponding to a realisation of h mutually exclusive and exhaustive random events. Denote the sequence of realised regimes from the beginning of observations up to time t by $\mathcal{J}_t$:

$$\mathcal{J}_t = \{s_1, s_2, ..., s_{t-1}, s_t\} \in \mathbb{H}_{t,t}$$

where $\mathbb{H}_{N,t}$ is the set of all possible histories of length N that end at period t. There are h^t possible mutually exclusive and exhaustive histories up to time t. Using the total probability theorem, the conditional pdf at time t is obtained as a Gaussian mixture with the number of terms equal to h^t:

$$f(y_{t+1} \mid Y_t) = \sum_{\mathcal{J}_t} f(y_{t+1} \mid \mathcal{J}_t, Y_t) \Pr[\mathcal{J}_t \mid Y_t].$$

The probability of a given regime history is computed using Bayes formula:

$$\begin{aligned}
\Pr[\mathcal{J}_t \mid Y_t] &= \Pr[\mathcal{J}_t \mid y_t, Y_{t-1}] = \frac{f(y_t \mid \mathcal{J}_t, Y_{t-1}) \Pr[\mathcal{J}_t \mid Y_{t-1}]}{f(y_t \mid Y_{t-1})} \\
&= \frac{f(y_t \mid \mathcal{J}_t, Y_{t-1}) \Pr[s_t, \mathcal{J}_{t-1} \mid Y_{t-1}]}{f(y_t \mid Y_{t-1})} \\
&= \frac{f(y_t \mid \mathcal{J}_t, Y_{t-1}) \Pr[s_t \mid \mathcal{J}_{t-1}, Y_{t-1}]}{f(y_t \mid Y_{t-1})} \Pr[\mathcal{J}_{t-1} \mid Y_{t-1}]
\end{aligned}$$

When the regime switches have Markov property, $\Pr[s_t \mid \mathcal{J}_{t-1}, Y_{t-1}] = \Pr[s_t \mid s_{t-1}] = Q(s_{t-1}, s_t)$, which simplifies the second term in the numerator. However, conditioning on the entire past history is still needed for the last term even if the regimes follow a Markov process.

In practice, one has to resort to some approximation. In this sense, all practical multiple-regime filters are approximate and, therefore, suboptimal. One popular approach involves merging two or more histories into one. A version of this approach is well known in economic applications as the Kim and Nelson filter. We focus on this framework and study two families of filters with different mechanisms of approximation.

In the next section, we present the Generalised Pseudo-Bayesian (GPB) filters, and Interacting Multiple Models (IMM) filters of arbitrary order (length of tracked histories) N.

2.2 Two Practical Families of Filters

We begin with the GPB(N) family, which includes the Kim and Nelson filter as a special case of GPB(2); see Kim (1994) and Bar-Shalom et al. (2001) for an exposition. Following commonly used notations, the GPB(N) filter uses information from the previous N periods, including the current one. Thus, GPB(1) ignores past history and uses current period information only, GPB(2) incorporates information from the current period and one immediately preceding period, and so on. The IMM algorithm is conceptually different from the GPB in the way it combines past histories. The version of IMM developed in Blom and Bar-Shalom (1988) corresponds to IMM(1); we refer to it as canonical IMM.

One would expect that a higher N leads to increased accuracy at the cost of a larger amount of computations. We investigate relative accuracy and speed for different N within each family.

Another interesting question is whether the canonical IMM outperforms the KN filter in accuracy and speed in a prototypical macroeconomic application.

In this section, we present the GPB(N) and the IMM(N) algorithms in turn, using uniform notations for the entire GPB family. Where relevant, we will remind the reader that KN is the same as GPB(2).

2.2.1 Preliminaries

Let $\mathcal{H}_t$ denote the history of regimes in N consecutive periods ending with period t,

$$\mathcal{H}_t := \{s_{t-N+1}, ..., s_{t-1}, s_t\} \in \mathbb{H}_{N,t},$$

and let $\mathcal{C}_t$ be the 'collapsed' history, defined as

$$\mathcal{C}_t := \{s_{t-N+2}, ..., s_t\} \in \mathbb{H}_{N-1,t}.$$

Hence

$$\mathcal{H}_t = \{\mathcal{C}_{t-1}, s_t\} = \{s_{t-N+1}, \mathcal{C}_t\},$$

and

$$\begin{aligned} \mathcal{H}_{t-1} &= \{s_{t-N}, ..., s_{t-2}, s_{t-1}\} \in \mathbb{H}_{N,t-1}, \\ \mathcal{H}_{t-1} \cup \mathcal{H}_t &= \{s_{t-N}, ..., s_{t-2}, s_{t-1}, s_t\} \in \mathbb{H}_{N+1,t}. \end{aligned}$$

Let

$$\mu_{t|t}^{(\mathcal{H}_t)} := \Pr\left[\mathcal{H}_t \mid Y_t\right]$$

be the probability of realisation of a particular history $\mathcal{H}_t$ conditional on information at time t.

2.2.2 Family of GPB Filters

The GPB algorithm of order N, denoted GPB(N), takes into account all h^N possible histories of the fixed length N, finishing at the current time period. It is implemented as follows.

Define

$$\mu_{t|t}^{(\mathcal{C}_t)} := \Pr\left[\mathcal{C}_t \mid Y_t\right] = \sum_{s_{t-N+1}=1}^{h} \mu_{t|t}^{(\mathcal{H}_t)}$$

as the probability of the collapsed history, $\mathcal{C}_t$, conditional on information at time t.

Algorithm 1 *GPB(N) Algorithm*

Step 0. Start at $t=1$. Initialise h^{N-1} versions of the state vector $\alpha_{t-1|t-1}^{(\mathcal{C}_{t-1})}$, the MSE matrix $P_{t-1|t-1}^{(\mathcal{C}_{t-1})}$, and probabilities $\mu_{t-1|t-1}^{(\mathcal{C}_{t-1})}$, $\mathcal{C}_{t-1} \in \mathbb{H}_{N-1,t-1}$.

Step 1. Compute h^N standard KF forecasts to obtain

GPB(N) Forecast	GPB(N) Update
$\alpha_{t\|t-1}^{(\mathcal{H}_t)} = c_{\alpha,s_t} + T_{s_t}\alpha_{t-1\|t-1}^{(\mathcal{C}_{t-1})}$,	$\alpha_{t\|t}^{(\mathcal{H}_t)} = \alpha_{t\|t-1}^{(\mathcal{H}_t)} + P_{t\|t-1}^{(\mathcal{H}_t)} Z'_{s_t} \left[F_{t\|t-1}^{(\mathcal{H}_t)}\right]^{-1} v_{t\|t-1}^{(\mathcal{H}_t)}$,
$P_{t\|t-1}^{(\mathcal{H}_t)} = T_{s_t} P_{t-1\|t-1}^{(\mathcal{C}_{t-1})} T'_{s_t} + R_{s_t} R'_{s_t}$,	$P_{t\|t}^{(\mathcal{H}_t)} = \left(I - P_{t\|t-1}^{(\mathcal{H}_t)} Z'_{s_t} \left[F_{t\|t-1}^{(\mathcal{H}_t)}\right]^{-1} Z_{s_t}\right) P_{t\|t-1}^{(\mathcal{H}_t)}$,

where

$$v_{t|t-1}^{(\mathcal{H}_t)} = y_t - Z_{s_t}\alpha_{t|t-1}^{(\mathcal{H}_t)} - c_{y,s_t}, \quad F_{t|t-1}^{(\mathcal{H}_t)} = Z_{s_t} P_{t|t-1}^{(\mathcal{H}_t)} Z'_{s_t} + H_{s_t}, \quad H_{s_t} = g_{s_t} g'_{s_t}.$$

Compute the associated likelihood

$$\Lambda_t^{(\mathcal{H}_t)} = f\left(y_t \mid \mathcal{H}_t, Y_{t-1}\right) = (2\pi)^{-p/2} \left|F_{t|t-1}^{(\mathcal{H}_t)}\right|^{-1/2} \exp\left(-\frac{1}{2} v_{t|t-1}^{(\mathcal{H}_t)\prime} \left[F_{t|t-1}^{(\mathcal{H}_t)}\right]^{-1} v_{t|t-1}^{(\mathcal{H}_t)}\right).$$

Step 2. Compute probabilities $\mu_t^{(\mathcal{H}_t)}$ according to[2]

$$\begin{aligned}
\mu_{t|t}^{(\mathcal{H}_t)} &= \Pr\left[\mathcal{H}_t \mid Y_t\right] = \Pr\left[\mathcal{H}_t \mid Y_{t-1}, y_t\right] \\
&= \frac{f\left(y_t, \mathcal{H}_t \mid Y_{t-1}\right)}{f\left(y_t \mid Y_{t-1}\right)} = \frac{f\left(y_t \mid \mathcal{H}_t, Y_{t-1}\right) \Pr\left[\mathcal{H}_t \mid Y_{t-1}\right]}{f\left(y_t \mid Y_{t-1}\right)} \\
&= \frac{f\left(y_t \mid \mathcal{H}_t, Y_{t-1}\right) \Pr\left[s_t \mid \mathcal{C}_{t-1}, Y_{t-1}\right]}{f\left(y_t \mid Y_{t-1}\right)} \Pr\left[\mathcal{C}_{t-1} \mid Y_{t-1}\right] \\
&\simeq \frac{\Lambda_t^{(\mathcal{H}_t)} Q\left(s_{t-1}, s_t\right)}{\sum_{\mathcal{H}_t} \Lambda_t^{(\mathcal{H}_t)} Q\left(s_{t-1}, s_t\right) \mu_{t-1|t-1}^{(\mathcal{C}_{t-1})}} \mu_{t-1|t-1}^{(\mathcal{C}_{t-1})},
\end{aligned}$$

where in the last line we used approximation

$$\Pr\left[s_t \mid \mathcal{C}_{t-1}, Y_{t-1}\right] \simeq \Pr\left[s_t \mid s_{t-1}\right] = Q\left(s_{t-1}, s_t\right), \tag{3}$$

and

$$\begin{aligned}
f\left(y_t \mid Y_{t-1}\right) &= \sum_{\mathcal{H}_t} f\left(y_t \mid Y_{t-1}, \mathcal{H}_t\right) \Pr\left[\mathcal{H}_t \mid Y_{t-1}\right] \\
&= \sum_{\mathcal{H}_t} f\left(y_t \mid Y_{t-1}, \mathcal{H}_t\right) \Pr\left[s_t \mid \mathcal{C}_{t-1}, Y_{t-1}\right] \Pr\left[\mathcal{C}_{t-1} \mid Y_{t-1}\right] \\
&\simeq \sum_{\mathcal{H}_t} \Lambda_t^{(\mathcal{H}_t)} Q\left(s_{t-1}, s_t\right) \mu_{t-1|t-1}^{(\mathcal{C}_{t-1})}.
\end{aligned} \tag{4}$$

[2] This procedure is a generalised version of the Hamilton (1989) filter.

where the sum is taken over all possible realisations of $\mathcal{H}_t$.

Step 3. Next, h^N KF outputs are merged into h^{N-1} using conditional probabilities $\Pr[s_{t-N+1}|Y_t, \mathcal{C}_t]$ computed as:

$$\Pr[s_{t-N+1}|Y_t, \mathcal{C}_t] = \frac{\Pr(s_{t-N+1}, \mathcal{C}_t|Y_t)}{\Pr(\mathcal{C}_t|Y_t)} = \frac{\Pr(\mathcal{H}_t|Y_t)}{\Pr(\mathcal{C}_t|Y_t)} = \frac{\mu_{t|t}^{(\mathcal{H}_t)}}{\sum_{s_{t+1-N}=1}^{h} \mu_{t|t}^{(\mathcal{H}_t)}}.$$

Thus,

$$\alpha_{t|t}^{(\mathcal{C}_t)} = \sum_{s_{t-N+1}=1}^{h} \Pr[s_{t-N+1}|Y_t, \mathcal{C}_t] \alpha_{t|t}^{(\mathcal{H}_t)}, \tag{5}$$

$$P_{t|t}^{(\mathcal{C}_t)} = \sum_{s_{t-N+1}=1}^{h} \Pr[s_{t-N+1}|Y_t, \mathcal{C}_t] \left\{ P_{t|t}^{(\mathcal{H}_t)} + \left(\alpha_{t|t}^{(\mathcal{C}_t)} - \alpha_{t|t}^{(\mathcal{H}_t)}\right)\left(\alpha_{t|t}^{(\mathcal{C}_t)} - \alpha_{t|t}^{(\mathcal{H}_t)}\right)' \right\} \tag{6}$$

$$\mu_{t|t}^{(\mathcal{C}_t)} = \Pr[\mathcal{C}_t \mid Y_t] = \sum_{s_{t-N+1}=1}^{h} \Pr[\mathcal{H}_t \mid Y_t] = \sum_{s_{t-N+1}=1}^{h} \mu_{t|t}^{(\mathcal{H}_t)}. \tag{7}$$

Updated $\alpha_{t|t}^{(\mathcal{C}_t)}$, $P_{t|t}^{(\mathcal{C}_t)}$ and $\mu_{t|t}^{(\mathcal{C}_t)}$ serve as initialisations for the next time period ($t = 2, 3, ..., n$) in a recursion at Step 1.

The t-increment likelihood

$$L_t = \log f(y_t \mid Y_{t-1}),$$

is computed using (4) as part of the filter algorithm.

To continue the recursion to the end of the sample we only need to compute $\left\{\alpha_{t|t}^{(\mathcal{C}_t)}, P_{t|t}^{(\mathcal{C}_t)}, \mu_{t|t}^{(\mathcal{C}_t)}\right\}_{(\mathcal{C}_t)\in\mathbb{H}_{N-1,t}}$ at each step of the algorithm. These quantities are also used to compute the state vectors and the MSE matrices $\left\{\alpha_{t|t}, P_{t|t}\right\}_{t=1:n}$ for each time t, and the probability $\mu_t^{(s_t)}$ of the system being in regime s_t conditional on information at time t:

$$\alpha_{t|t} = \sum_{\mathcal{C}_t} \Pr[\mathcal{C}_t \mid Y_t] \, \alpha_{t|t}^{(\mathcal{C}_t)} = \sum_{\mathcal{C}_t} \mu_{t|t}^{(\mathcal{C}_t)} \alpha_{t|t}^{(\mathcal{C}_t)}, \tag{8}$$

$$\begin{aligned} P_{t|t} &= \sum_{\mathcal{C}_t} \Pr[\mathcal{C}_t \mid Y_t] \left\{ P_{t|t}^{(\mathcal{C}_t)} + \left(\alpha_{t|t} - \alpha_{t|t}^{(\mathcal{C}_t)}\right)\left(\alpha_{t|t} - \alpha_{t|t}^{(\mathcal{C}_t)}\right)' \right\} \\ &= \sum_{\mathcal{C}_t} \mu_{t|t}^{(\mathcal{C}_t)} \left\{ P_{t|t}^{(\mathcal{C}_t)} + \left(\alpha_{t|t} - \alpha_{t|t}^{(\mathcal{C}_t)}\right)\left(\alpha_{t|t} - \alpha_{t|t}^{(\mathcal{C}_t)}\right)' \right\}, \end{aligned} \tag{9}$$

$$\mu_{t|t}^{(s_t)} = \Pr[s_t \mid Y_t] = \sum_{\mathcal{C}_{t-1}} \Pr[\mathcal{H}_t \mid Y_t] = \sum_{\mathcal{C}_{t-1}} \mu_{t|t}^{(\mathcal{H}_t)}. \tag{10}$$

These objects can be computed outside of the recursion.

This algorithm works for any $N \geq 1$. Note that for $N = 1$ we have $\mathcal{C}_t = \varnothing, \mathcal{H}_t = s_t$. This implies that the GPB(1) algorithm considers possible regime histories only at their latest instant and merges all preceding regime sequences into one, using *common* initial conditions $\alpha_{t-1|t-1}$ at each time step.

2.2.3 Family of IMM Filters

IMM(N) maintains the dimensionality of histories at h^N at every time period. This is achieved by *mixing* histories in every recursion of the algorithm immediately after making a step forward in time, which would otherwise result in an increase of dimension to h^{N+1}. Effectively, at every t, mixing replaces h^{N+1} extended 'exact' histories by h^N reduced approximate histories weighted by probabilities of transition from the earliest state, s_{t-N}, into the sequence of most recent states, $\{s_{t-N+1}, ..., s_t\}$. These h^N histories are then filtered and updated in the usual way.

A comparison of the IMM and the GPB shows that the dimension reduction is performed at the different stages of the algorithms.[3] In the IMM, mixing is done after the state update and before the measurement update, and in the GPB, collapsing is done after the state and measurement updates.[4]

Let $\mathcal{Q}_{t|t-1} := Q_{t|t-1} \otimes Q_{t-1|t-2} \otimes ... \otimes Q_{t-N+1|t-N}$ denote the grand transition matrix of format $h^N \times h^N$.

Algorithm 2 *IMM(N) Algorithm.*

Step 0. Initialise h^N versions of the state vector $\alpha_{t-1|t-1}^{(\mathcal{H}_{t-1})}$, the MSE matrix $P_{t-1|t-1}^{(\mathcal{H}_{t-1})}$, and regime probabilities $\mu_{t-1|t-1}^{(\mathcal{H}_{t-1})}$, $\mathcal{H}_{t-1} \in \mathbb{H}_{N,t-1}$. Compute $\mathcal{Q}_{t|t-1}$.

Step 1. Compute the mixing probabilities defined as

$$\mu_{t-1|t-1}^{(\mathcal{H}_{t-1}|\mathcal{H}_t)} := \Pr\left[\mathcal{H}_{t-1} \mid Y_{t-1}, \mathcal{H}_t\right].$$

Note that

$$\begin{aligned}
\Pr[\mathcal{H}_{t-1} \cup \mathcal{H}_t \mid Y_{t-1}] &= \Pr\left[\mathcal{H}_{t-1} \mid Y_{t-1}, \mathcal{H}_t\right] \Pr[\mathcal{H}_t \mid Y_{t-1}] \\
&= \Pr\left[\mathcal{H}_t \mid \mathcal{H}_{t-1}, Y_{t-1}\right] \Pr[\mathcal{H}_{t-1} \mid Y_{t-1}],
\end{aligned}$$

and

$$\Pr[\mathcal{H}_t \mid Y_{t-1}] = \sum_{\mathcal{H}_{t-1}} \Pr\left[\mathcal{H}_t \mid \mathcal{H}_{t-1}, Y_{t-1}\right] \Pr\left[\mathcal{H}_{t-1} \mid Y_{t-1}\right].$$

[3]See Tables A1 and A2 in Appendix A.

[4]We use *mixing* in the description of the IMM and *collapsing* in the description of the GPB following the convention in the literature.

Therefore,

$$\begin{aligned}
\mu_{t-1|t-1}^{(\mathcal{H}_{t-1}|\mathcal{H}_t)} &= \Pr\left[\mathcal{H}_{t-1} \mid Y_{t-1}, \mathcal{H}_t\right] = \frac{\Pr\left[\mathcal{H}_t \mid \mathcal{H}_{t-1}, Y_{t-1}\right]\Pr[\mathcal{H}_{t-1} \mid Y_{t-1}]}{\Pr[\mathcal{H}_t \mid Y_{t-1}]} \\
&= \frac{\Pr\left[\mathcal{H}_t \mid \mathcal{H}_{t-1}, Y_{t-1}\right]\Pr[\mathcal{H}_{t-1} \mid Y_{t-1}]}{\sum_{\mathcal{H}_{t-1}}\Pr\left[\mathcal{H}_t \mid \mathcal{H}_{t-1}, Y_{t-1}\right]\Pr\left[\mathcal{H}_{t-1} \mid Y_{t-1}\right]} \\
&\simeq \frac{\mathcal{Q}_{t|t-1}(\mathcal{H}_{t-1},\mathcal{H}_t)\mu_{t-1|t-1}^{(\mathcal{H}_{t-1})}}{\sum_{\mathcal{H}_{t-1}}\mathcal{Q}_{t|t-1}(\mathcal{H}_{t-1},\mathcal{H}_t)\mu_{t-1|t-1}^{(\mathcal{H}_{t-1})}}
\end{aligned}$$

Step 2. Compute the mixed state vectors and MSE matrices for each history:

$$\hat{\alpha}_{t-1|t-1}^{(*,\mathcal{H}_t)} = \sum_{\mathcal{H}_{t-1}}\Pr\left[\mathcal{H}_{t-1} \mid Y_{t-1},\mathcal{H}_t\right]\alpha_{t-1|t-1}^{(\mathcal{H}_{t-1})} = \sum_{\mathcal{H}_{t-1}}\mu_{t-1|t-1}^{(\mathcal{H}_{t-1}|\mathcal{H}_t)}\alpha_{t-1|t-1}^{(\mathcal{H}_{t-1})}, \tag{11}$$

$$\begin{aligned}
\hat{P}_{t-1|t-1}^{(*,\mathcal{H}_t)} &= \sum_{\mathcal{H}_{t-1}}\Pr\left[\mathcal{H}_{t-1} \mid Y_{t-1},\mathcal{H}_t\right] \qquad (12)\\
&\quad\times\left\{P_{t-1|t-1}^{(s_{t-1})} + \left(\alpha_{t-1|t-1}^{(\mathcal{H}_{t-1})} - \hat{\alpha}_{t-1|t-1}^{(*,\mathcal{H}_{t-1})}\right)\left(\alpha_{t-1|t-1}^{(\mathcal{H}_{t-1})} - \hat{\alpha}_{t-1|t-1}^{(*,\mathcal{H}_{t-1})}\right)'\right\} \\
&= \sum_{\mathcal{H}_{t-1}}\mu_{t-1|t-1}^{(\mathcal{H}_{t-1}|\mathcal{H}_t)}\left\{P_{t-1|t-1}^{(\mathcal{H}_{t-1})} + \left(\alpha_{t-1|t-1}^{(\mathcal{H}_{t-1})} - \hat{\alpha}_{t-1|t-1}^{(*,\mathcal{H}_{t-1})}\right)\left(\alpha_{t-1|t-1}^{(\mathcal{H}_{t-1})} - \hat{\alpha}_{t-1|t-1}^{(*,\mathcal{H}_{t-1})}\right)'\right\}.
\end{aligned}$$

Here, $\hat{\alpha}_{t-1|t-1}^{(*,\mathcal{H}_t)}$ is conditional on a particular sequence of regimes, and it is computed for all possible sequences, or histories, in $\mathbb{H}_{N,t}$. When computing $\hat{P}_{t-1|t-1}^{(*,\mathcal{H}_t)}$ in (12) we take the sum over all possible sequences $\mathcal{H}_{t-1}$ of length N ending at $t-1$ such that they overlap with $\mathcal{H}_t$ between times $t-N+1$ and $t-1$. Note that once states and MSEs are mixed, the memory of s_{t-N} is 'cleared', so we put an asterisk in place of the now non-existent index $s_{t-N} = \mathcal{H}_{t-1}\backslash(\mathcal{H}_t \cap \mathcal{H}_{t-1})$. This reduces the dimensionality from h^{N+1} to h^N.

Step 3. For each history compute the standard KF to obtain

IMM(N) Forecast	IMM(N) Update
$\alpha_{t\|t-1}^{(\mathcal{H}_t)} = c_{\alpha,s_t} + T_{s_t}\hat{\alpha}_{t-1\|t-1}^{(*,\mathcal{H}_t)}$,	$\alpha_{t\|t}^{(\mathcal{H}_t)} = \alpha_{t\|t-1}^{(\mathcal{H}_t)} + P_{t\|t-1}^{(\mathcal{H}_t)}Z'_{s_t}\left[F_{t\|t-1}^{(\mathcal{H}_t)}\right]^{-1}v_{t\|-1}^{(\mathcal{H}_t)}$,
$P_{t\|t-1}^{(\mathcal{H}_t)} = T_{s_t}\hat{P}_{t-1\|t-1}^{(*,\mathcal{H}_t)}T'_{s_t} + R_{s_t}R'_{s_t}$,	$P_{t\|t}^{(\mathcal{H}_t)} = \left(I - P_{t\|t-1}^{(\mathcal{H}_t)}Z'_{s_t}\left[F_{t\|t-1}^{(\mathcal{H}_t)}\right]^{-1}Z_{s_t}\right)P_{t\|t-1}^{(\mathcal{H}_t)}$,

where

$$v_{t|t-1}^{(\mathcal{H}_t)} = y_t - Z_{s_t}\alpha_{t|t-1}^{(\mathcal{H}_t)} - c_{y,s_t}, \quad F_{t|t-1}^{(\mathcal{H}_t)} = Z_{s_t}P_{t|t-1}^{(\mathcal{H}_t)}Z'_{s_t} + H_{s_t}, \quad H_{s_t} = g_{s_t}g'_{s_t}.$$

Compute the associated likelihood:

$$\begin{aligned}
\Lambda_t^{(\mathcal{H}_t)} &= f\left(y_t \mid \mathcal{H}_t, Y_{t-1}\right) \\
&= (2\pi)^{-p/2}\left|F_{t|t-1}^{(\mathcal{H}_t)}\right|^{-1/2}\exp\left(-\frac{1}{2}v_{t|t-1}^{(\mathcal{H}_t)\prime}\left[F_{t|t-1}^{(\mathcal{H}_t)}\right]^{-1}v_{t|t-1}^{(\mathcal{H}_t)}\right).
\end{aligned}$$

Step 4. Update the probabilities $\mu_{t|t}^{(\mathcal{H}_t)}$,

$$\begin{aligned}
\mu_{t|t}^{(\mathcal{H}_t)} &= \Pr\left[\mathcal{H}_t \mid Y_t\right] = \Pr\left[\mathcal{H}_t \mid Y_{t-1}, y_t\right] \\
&= \frac{f\left(y_t, \mathcal{H}_t \mid Y_{t-1}\right)}{f\left(y_t \mid Y_{t-1}\right)} = \frac{f\left(y_t \mid \mathcal{H}_t, Y_{t-1}\right) \Pr\left[\mathcal{H}_t \mid Y_{t-1}\right]}{f\left(y_t \mid Y_{t-1}\right)} \\
&= \frac{f\left(y_t \mid \mathcal{H}_t, Y_{t-1}\right) \sum_{\mathcal{H}_{t-1}} \Pr\left[\mathcal{H}_t \mid \mathcal{H}_{t-1}, Y_{t-1}\right] \Pr\left[\mathcal{H}_{t-1} \mid Y_{t-1}\right]}{f\left(y_t \mid Y_{t-1}\right)} \\
&\simeq \frac{\Lambda_t^{(\mathcal{H}_t)} \sum_{\mathcal{H}_{t-1}} \mathcal{Q}_{t|t-1}(\mathcal{H}_{t-1}, \mathcal{H}_t) \mu_{t-1|t-1}^{(\mathcal{H}_{t-1})}}{\sum_{\mathcal{H}_t} \Lambda_t^{(\mathcal{H}_t)} \sum_{\mathcal{H}_{t-1}} \mathcal{Q}_{t|t-1}(\mathcal{H}_{t-1}, \mathcal{H}_t) \mu_{t-1|t-1}^{(\mathcal{H}_{t-1})}},
\end{aligned}$$

where in the last line we used

$$\begin{aligned}
f\left(y_t \mid Y_{t-1}\right) &= \sum_{\mathcal{H}_t} f\left(y_t \mid Y_{t-1}, \mathcal{H}_t\right) \Pr\left[\mathcal{H}_t \mid Y_{t-1}\right] \\
&= \sum_{\mathcal{H}_t} f\left(y_t \mid Y_{t-1}, \mathcal{H}_t\right) \\
&\quad \times \sum_{\mathcal{H}_{t-1}} \left(\Pr\left[\mathcal{H}_t \mid \mathcal{H}_{t-1}, Y_{t-1}\right] \Pr\left[\mathcal{H}_{t-1} \mid Y_{t-1}\right]\right) \\
&\simeq \sum_{s\mathcal{H}_t} \Lambda_t^{(\mathcal{H}_t)} \sum_{\mathcal{H}_{t-1}} \mathcal{Q}_{t|t-1}(\mathcal{H}_{t-1}, \mathcal{H}_t) \mu_{t-1|t-1}^{(\mathcal{H}_{t-1})} \qquad (13)
\end{aligned}$$

The KF outputs and the updated probabilities, $\left\{\alpha_{t|t}^{(\mathcal{H}_t)}, P_{t|t}^{(\mathcal{H}_t)}, \mu_{t|t}^{(\mathcal{H}_t)}\right\}$, serve as initialisations for the next time step in a recursion. The t-increment likelihood is

$$L_t = \log f\left(y_t \mid Y_{t-1}\right)$$

and it is computed using formula (13) as part of the filtering algorithm. The state vectors, the MSE matrices, and updated regime probabilities $\left\{\alpha_{t|t}, P_{t|t}, \mu_{t|t}^{(s_t)}\right\}_{t=1:n}$ for each time t are computed using the same formulae (8)-(10) as in the GPB algorithm.

2.3 Smoothing

After estimating the states and regime probabilities through forward recursion, we can improve the inference on s_t and α_t using the information from the entire sample. This process, known as smoothing, is usually conducted by backward recursion. However, since smoothing algorithms employ approximations, the smoothed estimates may not necessarily be more precise than the filtered and updated estimates for every time period. Thus, the overall performance of smoothing algorithms is evaluated using simulations.

A practical algorithm for computing smoothed *probabilities* for a KN-GPB(2) filter is detailed in Kim (1994). In this paper we generalise it for arbitrary length of history to make it applicable to filters of higher order.

Smoothing *state vectors*, even in single-regime models, can be challenging, and is even more so in models with multiple regimes. Kim (1994) introduces an algorithm specifically designed to work with the KN-GPB(2) filter. However, this algorithm is computationally unstable as it requires inverting large auxiliary matrices. Existing smoothers developed for engineering applications typically exploit measurement errors and require the invertibility of matrix $H_{s_t} = g_{s_t} g'_{s_t}$, a condition often not met in economic applications. In this section, we adapt a single-regime smoothing algorithm proposed by Durbin and Koopman (2012), based on De Jong (1988), for use in a Markov-switching multiple-regime model with arbitrary history lengths. This adapted algorithm requires only matrix inversions that are part of the corresponding filter and would have been computed at the filtering stage.

2.3.1 Smoothed Probabilities

Smoothed probabilities are computed using the total probability theorem. The probability of history $\mathcal{H}_t$ conditional on the information contained in the full sample, Y_n, can be written as

$$\mu_{t|n}^{(\mathcal{H}_t)} := \Pr[\mathcal{H}_t \mid Y_n] = \sum_{s_{t+1}} \Pr[\mathcal{H}_t, s_{t+1} \mid Y_n].$$

To compute $\Pr[\mathcal{H}_t, s_{t+1} \mid Y_n]$, we use the following approximation:

$$\begin{aligned}
\Pr[\mathcal{H}_t, s_{t+1} \mid Y_n] &= \Pr\left[s_{t+1} \mid Y_n\right] \Pr\left[\mathcal{H}_t \mid s_{t+1}, Y_n\right] \\
&\simeq \Pr\left[s_{t+1} \mid Y_n\right] \Pr\left[\mathcal{H}_t \mid s_{t+1}, Y_t\right] \\
&= \Pr\left[s_{t+1} \mid Y_n\right] \frac{\Pr\left[\mathcal{H}_t, s_{t+1} \mid Y_t\right]}{\Pr\left[s_{t+1} \mid Y_t\right]} \\
&= \Pr\left[s_{t+1} \mid Y_n\right] \frac{\Pr\left[\mathcal{H}_t \mid Y_t\right] \Pr\left[s_{t+1} \mid \mathcal{H}_t, Y_t\right]}{\sum_{\mathcal{H}_t} \left(\Pr\left[s_{t+1} \mid \mathcal{H}_t, Y_t\right] \Pr\left[\mathcal{H}_t \mid Y_t\right]\right)} \\
&= \mu_{t+1|n}^{(s_{t+1})} \frac{\mu_{t|t}^{(\mathcal{H}_t)} Q\left(s_t, s_{t+1}\right)}{\sum_{\mathcal{H}_t} Q\left(s_t, s_{t+1}\right) \mu_{t|t}^{(\mathcal{H}_t)}}
\end{aligned}$$

Upon substitution, we get

$$\mu_{t|n}^{(\mathcal{H}_t)} \simeq \sum_{s_{t+1}} \mu_{t+1|n}^{(s_{t+1})} \frac{\mu_t^{(\mathcal{H}_t)} Q\left(s_t, s_{t+1}\right)}{\sum_{\mathcal{H}_t} Q\left(s_t, s_{t+1}\right) \mu_{t|t}^{(\mathcal{H}'_t)}} \tag{14}$$

The smoothing algorithm is implemented by backward recursion as follows.

Algorithm 3 *Smoothed Probabilities*

Step 0. Initialise $\mu_{n|n}^{(s_n)} = \Pr[s_n \mid Y_n]$, $s_n = 1, ..., h$.

Step 1. For $t = n-1$ use (14) to compute the smoothed probability of $\mu_{t|n}^{(\mathcal{H}_t)}$ for history $\mathcal{H}_t$.

Step 2. Compute smoothed probabilities of each regime:

$$\mu_{t|n}^{(s_t)} = \Pr[s_t \mid Y_n] = \sum_{\mathcal{C}_{t-1}} \Pr[\mathcal{H}_t \mid Y_n] = \sum_{\mathcal{C}_{t-1}} \mu_{t|n}^{(\mathcal{H}_t)}$$

Use $\mu_{t|n}^{(s_t)}$ to initialise the algorithm for $t = n-2$.

2.3.2 Smoothed Variables

The smoother is based on the properties of the joint Gaussian distribution of the forecast errors of the vector of latent state variables and the estimation errors of the vector of observations produced by the filter.

By definition, smoothed state vectors and MSE matrices are:

$$\alpha_{t|n}^{(\mathcal{H}_t)} = \mathbb{E}\left[\alpha_t^{(\mathcal{H}_t)} \mid Y_n, \mathcal{H}_t\right], \tag{15}$$

$$P_{t|n}^{(\mathcal{H}_t)} = \mathbb{E}\left[\left(\alpha_t^{(\mathcal{H}_t)} - \alpha_{t|t-1}^{(\mathcal{H}_t)}\right)\left(\alpha_t^{(\mathcal{H}_t)} - \alpha_{t|t-1}^{(\mathcal{H}_t)}\right)' \mid Y_n, \mathcal{H}_t\right]. \tag{16}$$

Define the forecast error of the state vector at time t with history $\mathcal{H}_t$ as

$$\xi_{t|t-1}^{(\mathcal{H}_t)} := \alpha_t - \alpha_{t|t-1}^{(\mathcal{H}_t)} \tag{17}$$

Then,

$$P_{t|t-1}^{(\mathcal{H}_t)} = \mathbb{E}\left[\xi_{t|t-1}^{(\mathcal{H}_t)} \xi_{t|t-1}^{(\mathcal{H}_t)\prime} \mid Y_{t-1}, \mathcal{H}_t\right]. \tag{18}$$

Define the Kalman gain matrix:

$$K_{t|t-1}^{(\mathcal{H}_t)} = P_{t|t-1}^{(\mathcal{H}_t)} Z_{s_t}' \left[F_{t|t-1}^{(\mathcal{H}_t)}\right]^{-1}. \tag{19}$$

To calculate $\alpha_{t|n}^{(\mathcal{H}_t)}$ defined in (15), we split the history into two components at $t-1$ and use the formula for the conditional mean of multivariate Gaussian distribution:

$$\begin{aligned}
\alpha_{t|n}^{(\mathcal{H}_t)} &= \mathbb{E}[\alpha_t \mid \mathcal{H}_t, Y_n] = \mathbb{E}\left[\alpha_t \mid \mathcal{H}_t, Y_{t-1}, \{v_{k|k-1}\}_{k=t:n}\right] \\
&= \alpha_{t|t-1}^{(\mathcal{H}_t)} + \sum_{k=t}^{n} \mathbb{E}\left[\alpha_t v_{k|k-1}' \mid Y_{t-1}, \mathcal{H}_t\right] \left[F_{k|k-1}\right]^{-1} v_{k|k-1} \\
&= \alpha_{t|t-1}^{(\mathcal{H}_t)} + \sum_{k=t}^{n} \mathbb{E}\left[\left(\xi_{t|t-1}^{(\mathcal{H}_t)} + \alpha_{t|t-1}^{(\mathcal{H}_t)}\right)\left(\xi_{k|k-1}' Z_{s_k}' + [g_{s_k}\varepsilon_k]'\right) \mid Y_{t-1}, \mathcal{H}_t\right] \left[F_{k|k-1}\right]^{-1} v_{k|k-1}
\end{aligned}$$

where we used

$$v_{t|t-1}^{(\mathcal{H}_t)} = y_t - Z_{s_t}\alpha_{t|t-1}^{(\mathcal{H}_t)} - c_{y,s_t} = y_t - Z_{s_t}\left(\alpha_t - \xi_{t|t-1}^{(\mathcal{H}_t)}\right) - c_{y,s_t} = g_{s_t}\varepsilon_t + Z_{s_t}\xi_{t|t-1}^{(\mathcal{H}_t)}.$$

So, finally,

$$\alpha_{t|n}^{(\mathcal{H}_t)} = \alpha_{t|t-1}^{(\mathcal{H}_t)} + \sum_{k=t}^{n} \mathbb{E}\left[\xi_{t|t-1}^{(\mathcal{H}_t)}\xi_{k|k-1}' \mid Y_{t-1}, \mathcal{H}_t\right] Z_{s_k}' \left[F_{k|k-1}\right]^{-1} v_{k|k-1}. \tag{20}$$

In this derivation we do not specify the *future* regime sequences starting from s_t under the summation. We introduce them in the calculations of expectations $\mathbb{E}[\cdot \mid \cdot]$ for every step going backward, as shown later.

For now, we will need the following recursion for $\xi_{t|t-1}^{(\mathcal{H}_t)}$. The recursion is slightly different for the two families of filters.

Lemma 1 *1. For GPB(N) filter*

$$\xi_{t|t-1}^{(\mathcal{H}_t)} = T_{s_t} \sum_{s_{t-N}=1}^{h} \Pr[s_{t-N}|Y_{t-1}, \mathcal{C}_{t-1}] \left(I - K_{t-1|t-2}^{(\mathcal{H}_{t-1})} Z_{s_{t-1}}\right) \xi_{t-1|t-2}^{(\mathcal{H}_{t-1})} + \omega_{t-1}, \tag{21}$$

where

$$\omega_{t-1} = R_{s_t}\eta_t - T_{s_t} \sum_{s_{t-N}=1}^{h} \Pr[s_{t-N}|Y_{t-1}, \mathcal{C}_{t-1}] K_{t-1|t-2}^{(\mathcal{H}_{t-1})} g_{s_{t-1}} \varepsilon_{t-1}.$$

2. For IMM filter

$$\xi_{t|t-1}^{(\mathcal{H}_t)} = T_{s_t} \sum_{\mathcal{H}_{t-1}} \Pr\left[\mathcal{H}_{t-1} \mid Y_{t-1}, \mathcal{H}_t\right] \left(I - K_{t-1|t-2}^{(\mathcal{H}_{t-1})} Z_{s_{t-1}}\right) \xi_{t-1|t-2}^{(\mathcal{H}_{t-1})} + \omega_{t-1}, \tag{22}$$

where

$$\omega_{t-1} = R_{s_t}\eta_t - T_{s_t} \sum_{\mathcal{H}_{t-1}} \Pr\left[\mathcal{H}_{t-1} \mid Y_{t-1}, \mathcal{H}_t\right] K_{t-1|t-2}^{(s_{t-1})} g_{s_{t-1}} \varepsilon_{t-1}.$$

Proof. For GPB(N) we have

$$\xi_{t|t-1}^{(\mathcal{H}_t)} = \alpha_t - \alpha_{t|t-1}^{(\mathcal{H}_t)} = T_{s_t}\left(\alpha_{t-1} - \alpha_{t-1|t-1}^{(\mathcal{C}_{t-1})}\right) + R_{s_t}\eta_t,$$

and, using $\alpha_{t-1|t-1}^{(\mathcal{C}_{t-1})}$ from equation (5),

$$\xi_{t|t-1}^{(\mathcal{H}_t)} = T_{s_t}\left(\alpha_{t-1} - \sum_{s_{t-N}=1}^{h} \Pr[s_{t-N} \mid Y_{t-1}, \mathcal{C}_{t-1}] \alpha_{t-1|t-1}^{(\mathcal{H}_{t-1})}\right) + R_{s_t}\eta_t. \tag{23}$$

Similarly, for IMM(N) we have

$$\xi_{t|t-1}^{(\mathcal{H}_t)} = T_{s_t}\left(\alpha_{t-1} - \hat{\alpha}_{t-1|t-1}^{(\mathcal{H}_{t-1}|\mathcal{H}_t)}\right) + R_{s_t}\eta_t$$

and using $\hat{\alpha}_{t-1|t-1}^{(\mathcal{H}_{t-1}|\mathcal{H}_t)}$ from (11),

$$\xi_{t|t-1}^{(\mathcal{H}_t)} = T_{s_t}\left(\alpha_{t-1} - \sum_{\mathcal{H}_{t-1}} \Pr\left[\mathcal{H}_{t-1} \mid Y_{t-1}, \mathcal{H}_t\right] \alpha_{t-1|t-1}^{(\mathcal{H}_{t-1})}\right) + R_{s_t}\eta_t. \tag{24}$$

Denote

$$M(\psi) = \begin{cases} \Pr[\psi \mid Y_{t-1}, \mathcal{C}_{t-1}], & \psi = s_{t-N}, \quad \text{for GPB(N)}, \\ \Pr[\psi \mid Y_{t-1}, \mathcal{H}_t], & \psi = \mathcal{H}_{t-1}, \quad \text{for IMM(N)}. \end{cases}$$

Then expressions (23) and (24) can be written in the same form,

$$\xi_{t|t-1}^{(\mathcal{H}_t)} = T_{s_t}\left(\alpha_{t-1} - \sum_{\psi} M(\psi)\, \alpha_{t-1|t-1}^{(\mathcal{H}_{t-1})}\right) + R_{s_t}\eta_t,$$

and the rest of the proof is identical for both families of filters.

Use the KF update for $\alpha_{t-1|t-1}^{(\mathcal{H}_{t-1})}$ and the definition of Kalman gain (19) for $K_{t-1|t-2}^{(\mathcal{H}_{t-1})}$ to rewrite the last expression as:

$$\begin{aligned}
\xi_{t|t-1}^{(\mathcal{H}_t)} &= T_{s_t}\Bigg(\alpha_{t-1} - \sum_{\psi} M(\psi) \\
&\quad \times \left(\alpha_{t-1|t-2}^{(\mathcal{H}_{t-1})} + P_{t-1|t-2}^{(\mathcal{H}_{t-1})} Z'_{s_{t-1}} \left[F_{t-1|t-2}^{(\mathcal{H}_{t-1})}\right]^{-1} v_{t-1|t-2}^{(\mathcal{H}_{t-1})}\right)\Bigg) + R_{s_t}\eta_t \\
&= T_{s_t}\sum_{\psi} M(\psi)\left(\alpha_{t-1} - \alpha_{t-1|t-2}^{(\mathcal{H}_{t-1})}\right) \\
&\quad - T_{s_t}\sum_{\psi} M(\psi)\, K_{t-1|t-2}^{(\mathcal{H}_{t-1})} v_{t-1|t-2}^{(\mathcal{H}_{t-1})} + R_{s_t}\eta_t
\end{aligned}$$

Next, use the KF output for $v_{t-1|t-2}^{(\mathcal{H}_{t-1})}$ along with the definition (17) for $\xi_{t-1|t-2}^{(\mathcal{H}_{t-1})}$ and equation (1) for y_{t-1} to obtain the recursions in Lemma 1:

$$\begin{aligned}
\xi_{t|t-1}^{(\mathcal{H}_{N,t})} &= T_{s_t}\sum_{\psi} M(\psi)\left(\alpha_{t-1} - \alpha_{t-1|t-2}^{(\mathcal{H}_{t-1})}\right) \\
&\quad - T_{s_t}\sum_{\psi} M(\psi)\, K_{t-1|t-2}^{(\mathcal{H}_{t-1})}\left(y_{t-1} - Z_{s_{t-1}}\alpha_{t-1|t-2}^{(\mathcal{H}_{t-1})} - c_{y,s_{t-1}}\right) + R_{s_t}\eta_t
\end{aligned}$$

$$
\begin{aligned}
&= T_{s_t} \sum_{\psi} M(\psi)\, \xi_{t-1|t-2}^{(\mathcal{H}_{t-1})} \\
&\quad - T_{s_t} \sum_{\psi} M(\psi)\, K_{t-1|t-2}^{(\mathcal{H}_{t-1})} Z_{s_{t-1}} \xi_{t-1|t-2}^{(\mathcal{H}_{t-1})} \\
&\quad + R_{s_t}\eta_t - T_{s_t} \sum_{\psi} M(\psi)\, K_{t-1|t-2}^{(\mathcal{H}_{t-1})} g_{s_{t-1}} \varepsilon_{t-1} \\
&= T_{s_t} \sum_{\psi} M(\psi) \left(I - K_{t-1|t-2}^{(\mathcal{H}_{t-1})} Z_{s_{t-1}} \right) \xi_{t-1|t-2}^{(\mathcal{H}_{t-1})} + \omega_{t-1}^{(\mathcal{H}_{t-1})}
\end{aligned}
$$

where we used the notation

$$\omega_{t-1} = R_{s_t}\eta_t - T_{s_t} \sum_{\psi} M(\psi)\, K_{t-1|t-2}^{(\mathcal{H}_{t-1})} g_{s_{t-1}} \varepsilon_{t-1}.$$

■

Note that in this derivation for GPB(N) the sum is taken over all possible regimes at time $t-N$ when s_{t-N} is unknown. If the regime s_{t-N} is known, this recursion is given by

$$\xi_{t|t-1}^{(s_{t-N},\mathcal{H}_t)} = T_{s_t} \left(I - K_{t-1|t-2}^{(s_{t-N},\mathcal{H}_{t-1})} Z_{s_{t-1}} \right) \xi_{t-1|t-2}^{(s_{t-N},\mathcal{H}_{t-1})} + \omega_{t-1}. \tag{25}$$

Similarly, for IMM, when $\mathcal{H}_{t-1} = \tilde{\mathcal{H}}_{t-1}$ is known, then

$$\xi_{t|t-1}^{\left(\tilde{\mathcal{H}}_{t-1},\mathcal{H}_t\right)} = T_{s_t} \left(I - K_{t-1|t-2}^{\left(\tilde{\mathcal{H}}_{t-1},\mathcal{H}_{t-1}\right)} Z_{s_{t-1}} \right) \xi_{t-1|t-2}^{\left(\tilde{\mathcal{H}}_{t-1},\mathcal{H}_{t-1}\right)} + \omega_{t-1}. \tag{26}$$

The remaining derivations are identical for GPB and IMM.

We apply formula (20) recursively, starting from the observation at the final period, n, in regime s_n:

$$
\begin{aligned}
\alpha_{n|n}^{(\mathcal{H}_n)} &= \alpha_{n|n-1}^{(\mathcal{H}_n)} + \mathbb{E}\left[\xi_{n|n-1}^{(\mathcal{H}_n)} \xi_{n|n-1}^{(\mathcal{H}_n)\prime} \mid Y_{n-1}, \mathcal{H}_n \right] Z_{s_n}' \left[F_{n|n-1}^{(\mathcal{H}_n)} \right]^{-1} v_{n|n-1}^{(\mathcal{H}_n)} \\
&= \alpha_{n|n-1}^{(\mathcal{H}_n)} + P_{n|n-1}^{(\mathcal{H}_n)} Z_{s_n}' \left[F_{n|n-1}^{(\mathcal{H}_n)} \right]^{-1} v_{n|n-1}^{(\mathcal{H}_n)} = \alpha_{n|n-1}^{(\mathcal{H}_n)} + P_{n|n-1}^{(\mathcal{H}_n)} r_{n|n-1}^{(\mathcal{H}_n)}
\end{aligned}
$$

where

$$r_{n|n-1}^{(\mathcal{H}_n)} = Z_{s_n}' \left[F_{n|n-1}^{(\mathcal{H}_n)} \right]^{-1} v_{n|n-1}^{(\mathcal{H}_n)}.$$

Next, we move one step back to $t = n-1$.

$$\alpha_{n-1|n}^{(\mathcal{H}_{n-1})} = \alpha_{n-1|n-2}^{(\mathcal{H}_{n-1})} + \sum_{k=n-1}^{n} \mathbb{E}\left[\xi_{n-1|n-2}^{(\mathcal{H}_{n-1})} \xi_{k|k-1}' \mid Y_{n-2}, \mathcal{H}_{n-1} \right] Z_{s_k}' \left[F_{k|k-1} \right]^{-1} v_{k|k-1}$$

$$
\begin{aligned}
&= \alpha_{n-1|n-2}^{(\mathcal{H}_{n-1})} + P_{n-1|n-2}^{(\mathcal{H}_{n-1})} Z'_{s_{n-1}} \left[F_{n-1|n-2}^{(\mathcal{H}_{n-1})} \right]^{-1} v_{n-1|n-2}^{(\mathcal{H}_{n-1})} \\
&\quad + \sum_{s_n=1}^{h} \Pr\left[s_n \mid Y_{n-2}, \mathcal{H}_{n-1} \right] \\
&\quad \times \mathbb{E}\left[\xi_{n-1|n-2}^{(\mathcal{H}_{n-1})} \xi_{n|n-1}^{(s_{n-N}, \mathcal{H}_{n-1})\prime} \mid Y_{n-2}, \mathcal{H}_{n-1}, s_n \right] Z'_{s_n} \left[F_{n|n-1}^{(\mathcal{H}_n)} \right]^{-1} v_{n|n-1}^{(\mathcal{H}_n)} \\
&= \alpha_{n-1|n-2}^{(\mathcal{H}_{n-1})} + P_{n-1|n-2}^{(\mathcal{H}_{n-1})} Z'_{s_{n-1}} \left[F_{n-1|n-2}^{(\mathcal{H}_{n-1})} \right]^{-1} v_{n-1|n-2}^{(\mathcal{H}_{n-1})} \\
&\quad + \sum_{s_n=1}^{h} \Pr\left[s_n \mid Y_{n-2}, \mathcal{H}_{n-1} \right] \mathbb{E}\left[\xi_{n-1|n-2}^{(\mathcal{H}_{n-1})} \xi_{n-1|n-2}^{(\mathcal{H}_{n-1})\prime} \mid Y_{n-2}, \mathcal{H}_{n-1} \right] \\
&\quad \times \left(I - Z'_{s_{n-1}} K_{n-1|n-2}^{(\mathcal{H}_{n-1})\prime} \right) T'_{s_t} Z'_{s_n} \left[F_{n|n-1}^{(\mathcal{H}_n)} \right]^{-1} v_{n|n-1}^{(\mathcal{H}_n)} \\
&= \alpha_{n-1|n-2}^{(\mathcal{H}_{n-1})} + P_{n-1|n-2}^{(\mathcal{H}_{n-1})} Z'_{s_{n-1}} \left[F_{n-1|n-2}^{(\mathcal{H}_{n-1})} \right]^{-1} v_{n-1|n-2}^{(\mathcal{H}_{n-1})} \\
&\quad + P_{n-1|n-2}^{(\mathcal{H}_{n-1})} \sum_{s_n=1}^{h} \Pr\left[s_n \mid Y_{n-2}, \mathcal{H}_{n-1} \right] \\
&\quad \times \left(I - Z'_{s_{n-1}} K_{n-1|n-2}^{(\mathcal{H}_{n-1})\prime} \right) T'_{s_t} Z'_{s_n} \left[F_{n|n-1}^{(\mathcal{H}_n)} \right]^{-1} v_{n|n-1}^{(\mathcal{H}_n)}
\end{aligned}
$$

In this derivation we used $P_{n-1|n-2}^{(\mathcal{H}_{n-1})} = \mathbb{E}\left[\xi_{n-1|n-2}^{(\mathcal{H}_{n-1})} \xi_{n-1|n-2}^{(\mathcal{H}_{n-1})\prime} \mid Y_{n-2}, \mathcal{H}_{n-1}, s_n \right] = \mathbb{E}\left[\xi_{n-1|n-2}^{(\mathcal{H}_{n-1})} \xi_{n-1|n-2}^{(\mathcal{H}_{n-1})\prime} \mid Y_{n-2}, \mathcal{H}_{n-1} \right]$ as conditioning on s_n becomes irrelevant.

This can be written as:

$$
\alpha_{n-1|n}^{(\mathcal{H}_{n-1})} = \alpha_{n-1|n-2}^{(\mathcal{H}_{n-1})} + P_{n-1|n-2}^{(\mathcal{H}_{n-1})} r_{n-1|n-2}^{(\mathcal{H}_{n-1})}
$$

where, using the approximation $\Pr\left[s_n \mid Y_{n-2}, \mathcal{H}_{n-1} \right] \simeq Q\left(s_{n-1}, s_n \right)$, we express $r_{n-1|n-2}^{(\mathcal{H}_{n-1})}$ recursively:

$$
r_{n-1|n-2}^{(\mathcal{H}_{n-1})} = Z'_{s_{n-1}} \left[F_{n-1|n-2}^{(\mathcal{H}_{n-1})} \right]^{-1} v_{n-1|n-2}^{(\mathcal{H}_{n-1})} + \left(I - Z'_{s_{n-1}} K_{n-1|n-2}^{(\mathcal{H}_{n-1})\prime} \right) \sum_{s_n=1}^{h} Q\left(s_{n-1}, s_n \right) T'_{s_n} r_{n|n-1}^{(\mathcal{H}_n)}.
$$

Thus, for $t = n-2$ in regime s_{n-2}, we obtain

$$
\alpha_{n-2|n}^{(\mathcal{H}_{n-2})} = \alpha_{n-2|n-3}^{(\mathcal{H}_{n-2})} + P_{n-2|n-3}^{(\mathcal{H}_{n-2})} r_{n-2|n-3}^{(\mathcal{H}_{n-2})}
$$

where

$$
\begin{aligned}
r_{n-2|n-3}^{(\mathcal{H}_{n-2})} &= Z'_{s_{n-2}} \left[F_{n-2|n-3}^{(\mathcal{H}_{n-2})} \right]^{-1} v_{n-2|n-3}^{(\mathcal{H}_{n-2})} + \left(I - Z'_{s_{n-2}} K_{n-2|n-3}^{(\mathcal{H}_{n-2})\prime} \right) \\
&\quad \times \sum_{s_{n-1}=1}^{h} Q\left(s_{n-2}, s_{n-1} \right) T'_{s_{n-1}} r_{n-1|n-2}^{(\mathcal{H}_{n-1})},
\end{aligned}
$$

and so on.

The procedure can be summarised in the following algorithm.

Algorithm 4 *State Smoothing*

Step 0. Initialise the smoother by setting $r_{n|n-1}^{(\mathcal{H}_n)} = Z'_{s_n}\left[F_{n|n-1}^{(\mathcal{H}_n)}\right]^{-1} v_{n|n-1}^{(\mathcal{H}_n)}$, $\mathcal{H}_n \in \mathbb{H}_{N,n}$. Here $\alpha_{n|n}^{(\mathcal{H}_n)}$ is the output of the corresponding filter at $t = n$.

Step 1. Compute the smoothed estimates of the state vector for each history $\mathcal{H}_t$, using recursion:

$$
\begin{aligned}
L_{t+1,t}^{(\mathcal{H}_t)} &= T_{s_{t+1}}\left(I - K_{t|t-1}^{(\mathcal{H}_t)} Z_{s_t}\right)\\
r_{t|t-1}^{(\mathcal{H}_t)} &= Z'_{s_t}\left[F_{t|t-1}^{(\mathcal{H}_t)}\right]^{-1} v_{t|t-1}^{(\mathcal{H}_t)} + \sum_{s_{t+1}=1}^{h} Q\left(s_t, s_{t+1}\right) L_{t+1,t}^{(\mathcal{H}_t)\prime} r_{t+1|t}^{(\mathcal{H}_{n-1})},\\
\alpha_{t|n}^{(\mathcal{H}_t)} &= \alpha_{t|t-1}^{(\mathcal{H}_t)} + P_{t|t-1}^{(\mathcal{H}_t)} r_{t|t-1}^{(\mathcal{H}_t)},
\end{aligned}
$$

for $t = n-1, n-2, ..., 1$.

Use the smoothed probabilities, $\left\{\mu_{t|n}^{(\mathcal{H}_t)}\right\}$, to compute the smoothed state vectors:

$$
x_{t|n} = \sum_{\mathcal{H}_t} \mu_{t|n}^{(\mathcal{H}_t)} \alpha_{t|n}^{(\mathcal{H}_t)}.
$$

3 Validating the Filters

3.1 Model and Parameterisation

To compare the performance of filters and smoothers, we use the model developed in Fernandez-Villaverde, Guerron-Quintana, and Rubio-Ramirez (2015), hereafter referred to as FGR2015. It is a relatively standard medium-scale New Keynesian DSGE model, which we modify to investigate the aspects of good luck and good policy.

The model consists of a household sector, firms, and a monetary authority. Households derive utility from consumption relative to their habit stock and from leisure. They supply differentiated labour to monopolistically competitive firms and choose wages subject to Calvo wage-setting friction. Firms produce differentiated output using capital, labour, and a neutral technology process. They set prices, also subject to Calvo pricing frictions. The capital stock evolves in the usual way, except for the inclusion of embodied technology in new investment goods. The model is closed by imposing a Taylor-type rule for the monetary authority. We present the full specification of the model in Appendix B.

We base the structural parameters of the model on the estimates reported in FGR2015; see column (1) in Table C1 in Appendix B. Our treatment of policy and shock volatilities is different from FGR2015, who estimated a single-regime nonlinear policy function and a single-regime stochastic volatility process. We introduce two Markov-switching processes into the model. The first, $S_{P,t}$, governs policy parameters in the following monetary policy rule:

$$\frac{r_t}{r_{ss}} = \left(\frac{r_{t-1}}{r_{ss}}\right)^{\gamma_r(S_{P,t})} \left(\left(\frac{\pi_t}{\pi_{\text{targ}}}\right)^{\gamma_\pi(S_{P,t})} \left(\frac{Y_{d,t}}{\lambda_{yd} Y_{d,t-1}}\right)^{\gamma_y(S_{P,t})} \right)^{1-\gamma_r(S_{P,t})} \exp\left(\sigma_\xi\left(S_{V,t}\right)\varepsilon_{\xi,t}\right). \quad (27)$$

The literature typically categorizes monetary policy approaches into hawkish and dovish modes, characterized by more and less aggressive responses to inflation, respectively. Accordingly, we assume that the $\gamma-$parameters are high in state $S_{P,t} = 1$ (hawkish state) and low in state $S_{P,t} = 2$ (dovish state). We explain below how we chose these values. The second two-state process, $S_{V,t}$, governs the shock volatilities for all shocks, including the policy shock in equation (27).

3.2 Monte-Carlo Simulations Design

In our simulations, we aim to differentiate between periods of infrequent large shocks and periods of more frequent regular shocks. We set the probability of remaining in the low volatility state to 0.95. This parameterisation implies an average of 20 quarters between high shocks, with a standard deviation of 19 quarters.[5] This probability accurately reflects the fact that recessions in the US have occurred approximately every 8-10 years since the end of World War II. We set the probability of staying in the high volatility state to 0.8, resulting in an average duration of high shock periods of 5 quarters (with a standard deviation of 4.5 quarters). Interpreting periods of large shocks as recessions suggests that a typical recession lasts slightly for less than a year, a duration that our parameterisation appropriately captures.

In formulating our policy model, we applied considerations similar to those used in the assumptions in the shock volatility experiments. The existing literature tends to report that hawkish policies have been predominant since the 1980s, spanning approximately 40 years.[6] However, considering the data starting from 1955 and acknowledging the evident dovish tendencies since 2008, we infer that the time split between these regimes is roughly equal. Therefore, we assume symmetric diagonal elements in the transition probability matrix. As the benchmark case, we calibrate the probability to remain in either of these states at 0.95. This implies an average of

[5] If probability to leave one of the two Markov states is q, then the expected length of stay in this state is $1/q$ with the standard deviation of $\sqrt{1-q}/q$.

[6] See e.g. Bianchi and Melosi (2017), Chen, Kirsanova, and Leith (2017).

20 quarters between policy changes, allowing for a wide range of durations between policy shifts. In addition, we consider an alternative calibration, with this probability set to 0.1. All transition matrices are presented in Table 1.

Table 1: Parameterisation of shock and policy regimes
Transition matrices

Shocks	Benchmark Case	Alternative Case
$P_s = \begin{bmatrix} 0.95 & 0.05 \\ 0.2 & 0.8 \end{bmatrix}$	$P_p^I = \begin{bmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{bmatrix}$	$P_p^{II} = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}$

Parameters of Taylor Rule:

	Hawkish Feedback	Dovish Feedback
γ_π^{Base}	1.7	0.9
$\gamma_\pi^{Altern.}$	1.5	0.9

As for the policy coefficients that are time-varying (or depend on the state), we describe the hawkish policy mode with feedback on inflation $\gamma_\pi^{Base} = 1.7$ in the hawkish state and $\gamma_\pi^{Base} = 0.9$ in the dovish state, consistent with findings in other studies[7]. We also consider an alternative parameterisation where these two feedbacks are less distinct, as shown in Table 1. In these simulations, we keep the feedback on output and the interest rate smoothing parameter the same in both hawkish and dovish states.

As reported in column (1) in Table C1, the standard deviations of all shocks in the low-volatility state, $S_{V,t} = 1$, are set to be equal to the mean estimates of corresponding variables in FGR2015, and they are doubled in the high-volatility state, $S_{V,t} = 2$.

In order to generate artificial data, we solve and simulate this non-linear model using a perturbation approach with the functional iteration algorithm developed for RISE© (Maih, 2015).

We chose to generate 500 samples of 1,000 observations each. We consider output growth, price inflation, wage inflation, the Federal Funds rate, and the relative price of investment goods as observable variables. The latent variables are listed in Table 2 and other relevant tables. We then use the simulation results to investigate the performance of the discussed filters, controlling for the sample length. Within each sample, we use the initial 300 observations as a proxy for a typical real-life scenario with post-WW2 quarterly data, where the influence of initial conditions can be substantial. Additionally, we analyze the full sample of 1000 observations, in which we

[7]See, e.g. Bianchi (2012), Chang, Kwak, and Qiu (2021), Chen, Leeper, and Leith (2022).

expect the impact of initial conditions to be significantly diminished.

3.3 Results

3.3.1 Evaluation Criteria

We need some criteria to rank the filters for practical purpose, based on their accuracy and speed. For accuracy, or goodness-of-fit, in our exercise we cannot use measures linked to the likelihood $L_t = \log f(y_t \mid Y_{t-1})$ returned by the filters. This is because different filters employ different approximations when computing the likelihood, and so comparison based on this measure is not compelling for comparison of the filters. An alternative and, perhaps, more straightforward approach in our case is to use root mean squared errors (RMSE) for each latent variable α_t, given by the formula

$$\mathcal{R}_{\varphi} = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left(\frac{\alpha_t - \alpha_{\varphi}}{\alpha_{ss}} \right)^2}.$$

We present the comparison of the accuracy of the filters based on the updated variables ($\varphi = t \mid t$) and smoothed variables ($\varphi = t \mid n$) in Tables 2-7.[8] Here, n is the length of each data sample, and n_{sim} is the number of simulations.

3.3.2 The Best Performing Filter

Table 2 shows the results for four filters: the IMM(1) and GPB(N) for $N = 1, 2, 3$, which includes the KN filter as it is equivalent to GPB(2).[9] Our simulations reveal that increasing the order of the GPB(N) filter beyond N=3 offers no practical value. We do not present results for IMM(2) as it does not noticeably improve accuracy of the IMM(1).

We focus on the *updated* variables, as these variables contribute to the likelihood used in estimation. The average RMSEs (denoted as $\mathcal{R}_{t|t}$) for all 500 draws in the Monte Carlo experiment are presented in columns (1)-(4) of Table 2. They vary in magnitude, reflecting findings similar to those in Binning and Maih (2015), where it is observed that highly persistent latent variables, such as capital, pose greater challenges for reconstruction.

The relative RMSEs in columns (5)-(8) are computed by dividing the RMSE for each variable by the lowest RMSE for that particular variable across investigated filters. In other words, for the best-performing filter it has the value of 1, while for all other filters its value is greater than one.

[8] In computing RMSEs, we normalise all variables, except state probabilities, by their steady-state levels, as in this model the steady state is identical for all regimes.

[9] Appendix A presents selected filtering and smoothing algorithms in a form convenient for implementation.

Table 2: MRSEs for update variables from four filters

	Absolute RMSEs $R_{t\|t}$				Relative RMSEs			
	IMM(1)	GPB(2) KN	GPB(1)	GPB(3)	IMM(1)	GPB(2) KN	GPB(1)	GPB(3)
variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
consumption	0.032	0.032	0.033	0.032	1.0003	1	1.030	1.0002
capital	0.226	0.226	0.238	0.226	1.0001	1	1.051	1.0006
output	0.030	0.030	0.030	0.030	1.0005	1.00004	1.016	1
real wage	0.002	0.002	0.002	0.002	1.0002	1	1.035	1.0003
Tobin's Q	0.010	0.010	0.010	0.010	1.0001	1	1.025	1.0001
investment	0.302	0.302	0.321	0.302	1	1.00004	1.065	1.0010
lab supply	0.029	0.029	0.030	0.029	1.0005	1.00004	1.016	1
pref shock	0.043	0.043	0.044	0.043	1.0001	1	1.015	1.00006
lab sup shock	0.070	0.070	0.072	0.070	1.0005	1.0001	1.019	1
tech shock	0.001	0.001	0.001	0.001	1.0004	1	1.066	1.0015
shock reg. probs	0.265	0.265	0.266	0.265	1.00002	1	1.002	1.0004
policy reg. probs	0.335	0.335	0.344	0.335	1	1.00001	1.023	1.00003

Table 2 suggests, as expected, that in terms of accuracy GPB(1) is dominated by two other filters. While KN-GPB(2) filter performs the best for more variables than IMM(1), the maximal difference in their performance is only 0.04%. In contrast, GPB(1) is outperformed by about 2-7%. The GPB(3) has shown performance very similar to that of GPB(2) and IMM filters, without clear dominance over GPB(2).

Table 3: Computational times for filtering 1000 observations

A: IMM(1) vs. GPB(2)-KN			B: Relative speed			
	updating only sec	updating and smoothing sec		ratio	updating only	ratio
			GPB(1)	0.28		
IMM(1)	0.27	1.49	GPB(2)	1	IMM(1)	1
GPB(2)	1.38	2.59	GPB(3)	4.21	IMM(2)	5.81
			GPB(4)	17.74	IMM(3)	52.70
			GPB(5)	79.97	IMM(4)	691.14

These RMSEs are visualised by a red dash-dotted line in Figure 1, illustrating the recovery of latent variables and probabilities of being in a particular Markov state in one particular simulation. The true values are shown by the solid black lines. We plot only the initial 300 observations for clarity of visualisation.

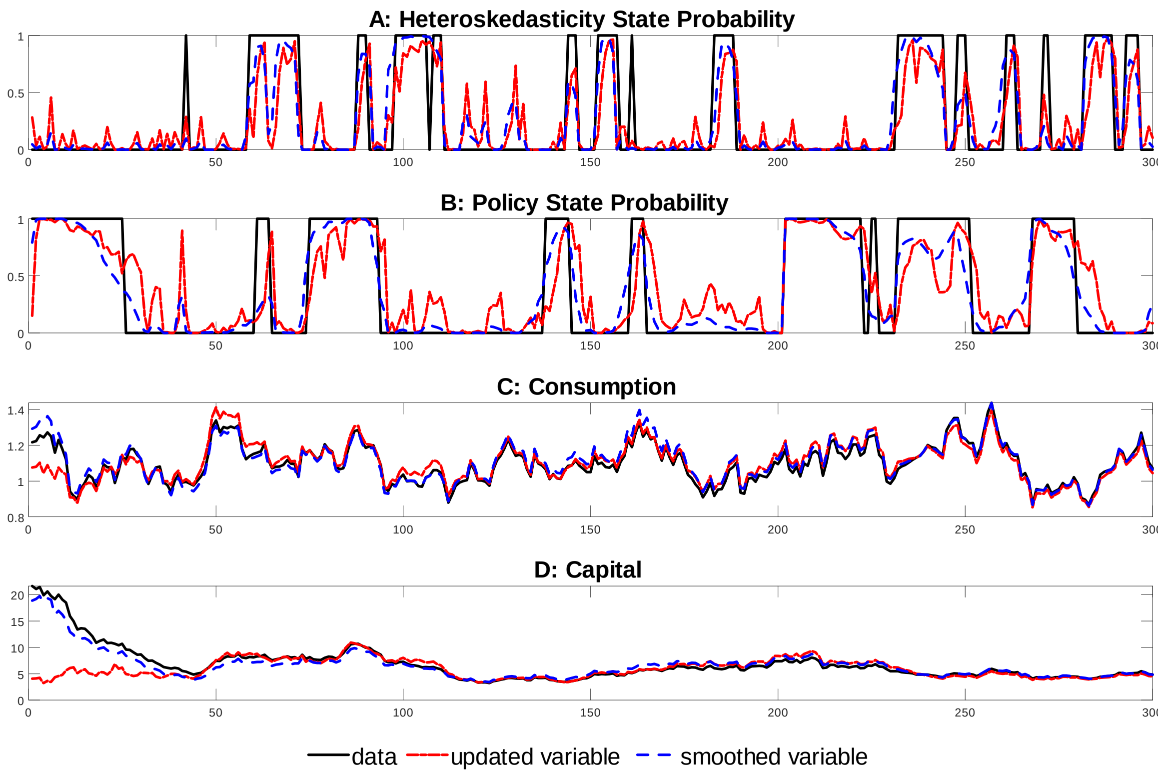

Figure 1: Updating and smoothing produced by IMM filter.

In terms of computational burden, IMM(1) has an advantage over KN-GPB(2) as it works with h histories, rather than with h^2. Panel A of Table 3 shows indicative computational times for these two filters, both independently and in conjunction with the corresponding smoother.[10] These times serve merely as an indication of computational speed, as all filters implemented in RISE© perform additional tasks beyond algorithm computation, which, even if not used, hold the potential to reduce speed.[11] It is important to note that at the estimation stage, where speed is particularly crucial, smoothing is not applied.

One can see in panel B of Table 3 how quickly the computation burden of IMM rises with higher orders. This is because the algorithm keeps track of all possible histories of fixed length

[10] These numbers are achieved on a Ryzen 3950X with 64GB RAM using MATLAB© R2022b.

[11] This includes checking for and accomodating properties such as time-varying states and parameters, missing observations, nonstationarity, and occasionally-binding constraints. See Appendix D for further notes on implementation.

where the histories contain all possible combinations of regimes in every time period of recursion. Therefore, the number of terms containing probabilities of different combinations of regimes is much bigger in IMM than in GPB.

One can reduce the computational burden in IMM(N) by using approximation similar to (3). However, because of the repeated use the disrepancy created by this approximation accumulates and leads to lower accuracy, especially for higher N.

On the balance between accuracy and speed, it is clear that IMM(1) dominates the GPB(N) family of filters. Based on that, we argue that in practical applications, IMM(1) is the best. In what follows, we use IMM to denote IMM(1).

3.3.3 Updating and Smoothing

To visualize the effect of smoothing, in Figure 1 we plot updated and smoothed probabilities and latent variables, alongside their true values. When comparing all lines, we see that smoothing reduces high-frequency noise in the recovered probabilities and often helps to identify the timing of regime changes more accurately. It is also apparent that the initial gap between the updated and actual values of latent variables is substantially corrected, although the improvement is not uniform across the entire sample. There are time periods where smoothing does not improve these particular variables at all. Notably, while the impact of initial conditions on the filter's effectiveness for economic variables is very clear, such an effect is less prominent for probabilities.

Table 4: Accuracy improvement by smoothing, $1-(R_{t|T}/R_{t|t})$

vars:	IMM	GPB(2)	GPB(1)	GPB(3)
consumption	0.29	0.27	0.31	0.27
capital	0.21	0.21	0.23	0.21
output	0.42	0.37	0.47	0.38
real wage	0.18	0.18	0.18	0.18
Tobin's Q	0.25	0.25	0.23	0.25
investment	0.30	0.28	0.32	0.28
labour supply	0.42	0.37	0.47	0.38
preference shock	0.14	0.13	0.14	0.13
labour supply shock	0.37	0.33	0.40	0.34
technology shock	0.10	0.10	0.04	0.10
shock state probs	0.15	0.15	0.15	0.15
policy state probs	0.18	0.18	0.16	0.18

Figure 1 illustrates the work of smoothing for the IMM filter. For all considered filters, Table 4 reports the fractions of RMSEs that is removed by smoothing, $1-\frac{\mathcal{R}_{t|T}}{\mathcal{R}_{t|t}}$.

While the RMSEs for some variables are only improved by 4%, for some others the improvement is as large as 47%, and the average improvement is about 25%. For state probabilities, the average improvement is 16%. Notably, the less computationally intensive smoothers for IMM and GPB(1) improve accuracy better than more complex algorithms for higher order of GPB filters.

3.3.4 Information

The following experiments aim to address the importance of working with longer series of data. In the first experiment, we use the IMM filter; other filters show very similar results. Column (1) of Panel A in Table 5 presents RMSEs for updated variables using the first 300 observations in each simulation. Columns (2) and (3) present RMSEs for smoothed variables for the first 300 observations, where smoothing starts from the last of these 300 observations in column (2) and from the last of all 1000 observations in column (3). The comparison is consistent with the intuition that more information improves accuracy. However, this intuition is not necessarily true in our setting because filtering and smoothing procedures involve numerous approximations. It is remarkable that, despite these approximations, the improvement in accuracy is substantial: close to 10% for some variables.

Table 5: Importance of Information. Panel A: RMSE for updated (1) and smoothed (2,3) variables. Panel B: Accuracy improvement by smoothing.

	Panel A			Panel B			
sample size	300	300	300	300	250	200	100
	$\mathcal{R}_{t\|t}$	$\mathcal{R}_{t\|300}$	$\mathcal{R}_{t\|1000}$	1-$\frac{\mathcal{R}_{t\|300}}{\mathcal{R}_{t\|t}}$	1-$\frac{\mathcal{R}_{t\|250}}{\mathcal{R}_{t\|t}}$	1-$\frac{\mathcal{R}_{t\|200}}{\mathcal{R}_{t\|t}}$	1-$\frac{\mathcal{R}_{t\|100}}{\mathcal{R}_{t\|t}}$
vars:	(1)	(2)	(3)	(1)	(2)	(3)	(4)
consumption	0.051	0.036	0.033	0.30	0.30	0.30	0.26
capital	0.352	0.277	0.266	0.21	0.21	0.20	0.14
output	0.050	0.029	0.026	0.43	0.43	0.42	0.33
real wage	0.003	0.002	0.002	0.17	0.17	0.17	0.11
Tobin's Q	0.010	0.008	0.008	0.26	0.26	0.26	0.22
investment	0.466	0.317	0.300	0.32	0.32	0.32	0.27
labour supply	0.049	0.028	0.026	0.43	0.42	0.42	0.32
pref. shock	0.051	0.042	0.041	0.18	0.19	0.20	0.19
lab. supp. shock	0.110	0.066	0.061	0.40	0.40	0.40	0.30
techn. shock	0.002	0.001	0.001	0.10	0.10	0.11	0.11
shock state probs	0.265	0.225	0.225	0.15	0.15	0.15	0.15
policy state probs	0.335	0.275	0.275	0.18	0.18	0.18	0.19

Panel B of Table 5 shows an improvement in the RMSEs from smoothing obtained in the second experiment. Here we explore different sample sizes, $n \in \{300, 250, 200, 100\}$. We know that RMSEs of updated variables are larger in shorter samples. One might expect that the efficacy of smoothing will also deteriorate in shorter samples. However, Panel B reveals that this is not the case, except for $n = 100$. This suggests that the sample size should be in excess of 100 and, perhaps, at least 200 to ensure that the smoother improves accuracy.

Another important observation is that there is no sample size effect for the smoothing of state probabilities, as can be seen in the last two rows of Panel B. This is consistent with the results presented in Figure 1, where the effect of initial conditions is only observed for latent economic variables: in a shorter sample initial conditions play bigger role, but this not the case for probabilities.

3.3.5 Policy States

For the next experiment, we simulate artificial data for less and more distinct policy states as measured by different probabilities of remaining in a given policy state in the next period, and by larger difference in the feedback coefficient γ_π in two policy states. Table 6 reports the results.

Table 6: RMSE for Pobability of Hawkish Policy State.

policy description	γ_π	$P_{HH} = P_{DD}$	updated	smoothed
more distinct states, less distinct feedback	1.5	0.95	0.366	0.310
less distinct states, less distinct feedback	1.5	0.9	0.412	0.380
more distinct states, more distinct feedback	1.7	0.95	0.335	0.275
less distinct states, more distinct feedback	1.7	0.9	0.386	0.349

We conclude that the greater the difference between the states, the better is their identification. This is true for both updated and smoothed policy state probabilities.

3.3.6 Model Misspecification

We investigate several cases of the model misspecification relevant for the Markov-switching nature of our model. We assume that the true data-generating process contains two Markov-switching processes as described above, but a researcher only considers one of them, either in policy or in volatility.

In the first scenario, we assume that the researcher believes in a single policy stance: the hawkish state.[12] We then execute the filtering and smoothing algorithm with γ_π^{Base}, which is

[12] This is the common assumption in constant-parameter DSGE model estimations, see e.g. Chen, Kirsanova, and Leith (2017)

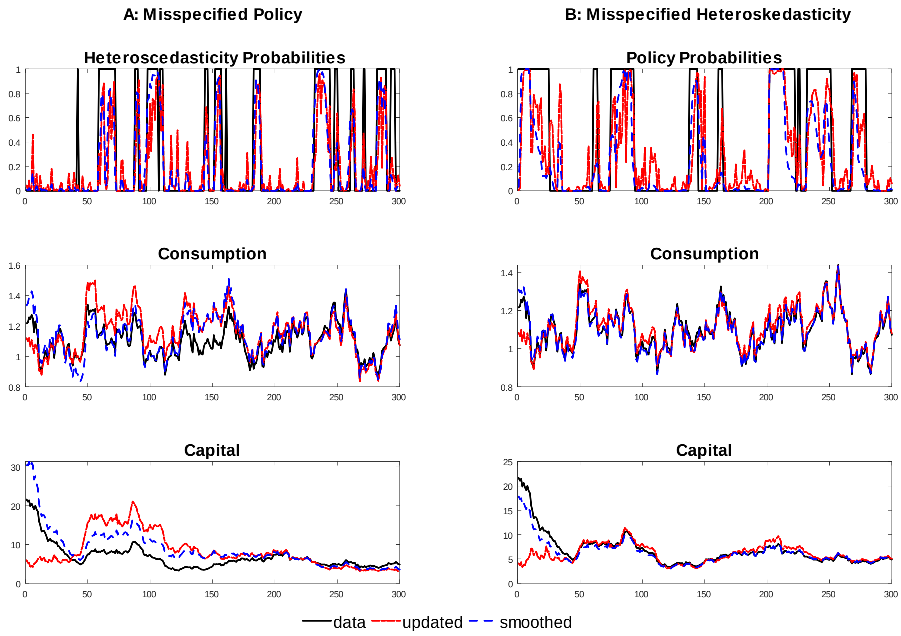

Figure 2: Model Misspecifications

the correct policy feedback but only for one of the two policy Markov states. Panel A in Figure 2 illustrates the outcomes of updating and smoothing for the same simulation, focusing on the initial 300 observations. The updated and smoothed variables are shown to accurately identify the patterns in the data, although with larger errors than in the correctly specified model.

One can get further insights from Table 7 by comparing the RMSEs for updated and smoothed variables obtained from 500 simulations of the entire sample of 1000 observations in columns (1) and (2). Despite effectively handling data patterns, RMSEs for smoothed latent variables show a substantial increase. However, the RMSEs for smoothed heteroskedasticity state probability show only a small increase, as Figure 2 also demonstrates.

In the second scenario, we assume that the researcher believes the volatility is always low. Panel B of Figure 2, which shows the results of this scenario, confirms that the filter correctly identifies the patterns in the data. Column (3) in Table 7 further supports this observation, as it shows a much smaller increase in RMSEs compared to column (1), and even smaller estimation errors for some variables. At the same time, the RMSE for smoothed probability of a policy state is higher than in the correctly specified model. This suggests that incorrectly specifying shock volatilities significantly worsens the identification of policy states.

Table 7: MRSEs of smoothed variables $R_{t|1000}$.

vars:	No missp-b (1)	Missp-d Policy H (2)	Missp-d Shocks L (3)	Missp-d Shocks H (4)
consumption	0.023 [0.032]	0.057 [0.070]	0.022 [0.036]	0.021 [0.035]
capital	0.178 [0.226]	1.472 [1.683]	0.186 [0.259]	0.164 [0.239]
output	0.017 [0.030]	0.020 [0.038]	0.012 [0.032]	0.011 [0.031]
real wage	0.002 [0.002]	0.005 [0.005]	0.002 [0.002]	0.002 [0.02]
Tobin's Q	0.007 [0.010]	0.036 [0.037]	0.011 [0.013]	0.007 [0.010]
investment	0.210 [0.302]	2.151 [2.533]	0.224 [0.349]	0.182 [0.317]
labour supply	0.017 [0.029]	0.019 [0.036]	0.011 [0.031]	0.011 [0.031]
preference shock	0.037 [0.043]	0.112 [0.113]	0.045 [0.051]	0.037 [0.044]
labour supply shock	0.044 [0.070]	0.174 [0.199]	0.045 [0.080]	0.035 [0.073]
technology shock	0.001 [0.001]	0.002 [0.002]	0.001 [0.001]	0.001 [0.001]
shock state probs	0.224 [0.265]	0.247 [0.280]	–	–
policy state probs	0.275 [0.335]	–	0.390 [0.408]	0.282 [0.339]

Note: MRSEs of updated variables $\mathcal{R}_{t|t}$ are in square brackets.

In the final experiment, reported in column (4), we revisit the second scenario, but this time we assume that the researcher believes the volatility is always high. Although the RMSEs for updated latent variables in column (4) are higher than those in column (1), the RMSEs for smoothed variables are sometimes lower than in the correctly specified model. The unexpectedly superior performance of the misspecified model after smoothing can be attributed to the larger variance of shocks. By allowing for a large variance in the shocks distribution, it accommodates both large and small shocks. The smoother then revises the estimated values using the complete sample and adjusts the estimates by factoring in information about the realized shocks.

3.4 Interim Summary

Overall, the findings in this section demonstrate the effectiveness of the canonical IMM filter, particularly when combined with the appropriate smoother, in enhancing the accuracy and efficiency of Bayesian estimation of state-space models.

The canonical IMM outperforms the Kim and Nelson filter in terms of computational speed while delivering comparable accuracy. The implementation of the new smoothing algorithm with the IMM filter substantially enhances precision in estimating latent variables, reducing errors by approximately 25%. We do not find any substantial improvement in accuracy when using higher order filters in our example. It is hard to predict whether the same will be true for other models.

Our simulations confirm that, despite approximations, adding more information improves the performance of the suggested filtering-smoothing procedure. We find that, as long as the sample length remains above 200 observations, there is no reduction in the smoother's efficacy in reducing RMSEs for updated variables. We find that the filter identifies probabilities of more distinct policy regimes with higher accuracy.

Finally, we demonstrate that we can still successfully recover latent variables even when the policy or shock volatility regimes in the model are misspecified.

Having established the superiority of the canonical IMM paired with the matching smoother, we focus on this filter and smoother in the empirical application.

4 Empirical Application

In this section, we further investigate the practicality of the IMM filter with the corresponding smoother. We estimate a modified version of the FGR2015 model but using the same data for 1959Q2-2013Q4 as in that paper (see Table C2 in Appendix C). In our estimation we impose relatively wide priors and use the *Artificial Bee Colony* algorithm by Karaboga and Basturk (2007) for global optimisation.

Table 8 displays the estimated mode of the distribution of transition probabilities and policy parameters. We present the remaining parameters in Table C1 in Appendix C.

Table 8: Estimation of parameters that govern the two Markov processes
Transition matrices

Shocks	Policy
$P_s = \begin{bmatrix} 0.939\,05 & 0.060946 \\ 0.04625 & 0.95375 \end{bmatrix}$	$P_p = \begin{bmatrix} 0.983\,96 & 0.016039 \\ 0.043428 & 0.956\,57 \end{bmatrix}$

Parameters of Taylor Rule:

Hawkish Feedback γ_π	1.6574
Dovish Feedback γ_π	0.93984

We note that both regime-switching processes are highly persistent, and therefore their identification is likely to be correct, as suggested by our simulation results. The policy process, in particular, shows that there is only a 2% probability of leaving the hawkish state.

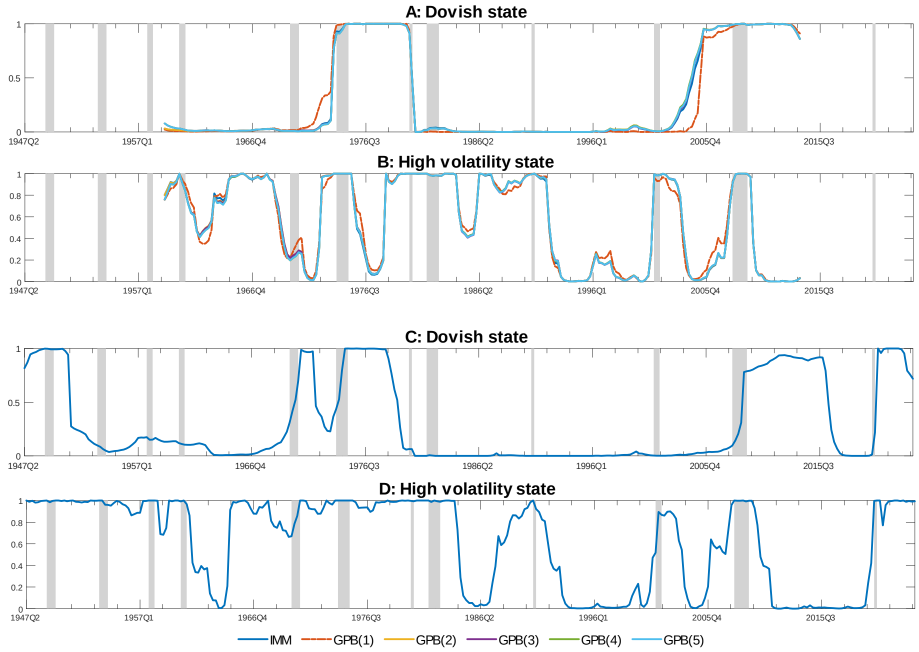

Figure 3: Smoothed State Probabilities

Panels A and B in Figure 3 report the smoothed probabilities of being in the dovish state and the high volatility state. We used the canonical IMM at the estimation stage and six different filters at the filtering stage.

One can notice that the lines plotted for six filters are very close to one another. The GPB filters of order 2 to 5 produce nearly identical results and they are also extremely close to those produced by the canonical IMM. This suggests, first, that using a more computationally intensive higher-order GPB filter does not necessarily improve regime identification compared to the KN-GPB(2) filter, and, second, that the canonical IMM and the KN filter are practically identical in accuracy. While GPB(1) stands out as less accurate, it still identifies all main events similarly to the other filters.

Panel A shows the probability of being in the dovish policy state. Note that it indicates that our approach succeded in identifying all major changes in the US post-war policy stance: the Great Inflation, the Volcker Disinflation, the Great Moderation, the Great Financial Crisis, and the subsequent Zero Lower Bound (ZLB) period. We did not assume a special regime for ZLB monetary policy but identify this period as a dovish state.

Panel B shows the probability of being in the high volatility state. Our approach correctly identifies most of the recessions and suggests that the pre-1990s period experienced larger shocks than the more recent past.

For panels C and D the dataset includes the period 1947Q2-2023Q3 (see Appendix C for details). The extended data covers a longer period, adding observations at the beginning, which should improve the identification of the Great Inflation episode, and at the end, which includes the post-Covid period with rising inflation in 2022-23. In these two panels we only show the results obtained using the IMM filter (with the associated smoother) to this extended dataset.

The message is similar to what is suggested by panels A and B. We identify the dovish state during the ZLB and a shift to hawkish policy a year after the ZLB lift-off. In addition, we see the return to the dovish policy during the Covid-19 pandemic which lasted until 2023Q1, at which time tough measures against inflation were taken. The post-Covid period is also characterised by relatively large shocks.

5 Conclusions

Our focus in this paper has been on improving multiple-regime Bayesian filtering techniques, alongside the development of multiple-regime smoothers.

We introduced the family of IMM filters, along with an extension of the Kim and Nelson filter, to accommodate tracking of longer regime histories. In addition, we developed a robust smoothing algorithm that can be adapted to these extended filters. Our simulation exercises demonstrate that the IMM filter with our proposed smoother deliver the best combination of computational speed and accuracy in a prototypical macroeconomic application of Bayesian filtering.

Our paper provides a comprehensive toolkit for researchers working with complex macroeconomic models. We demonstrate its practical relevance in an empirical application using a NK DSGE model with long U.S. macroeconomic time series.

References

Bar-Shalom, Y., X.-R. Li, and T. Kirubarajan (2001). *Estimation with Applications To Tracking and Navigation*. New York: John Wiley and Sons.

Bianchi, F. (2012). Evolving monetary/fiscal policy mix in the united states. *American Economic Review 102*(3), 167–72.

Bianchi, F. and L. Melosi (2017). Escaping the Great Recession. *American Economic Review 107*(4), 1030–1058.

Binning, A. and J. Maih (2015). Sigma point filters for dynamic nonlinear regime switching models. Working Paper 2015/10, Norges Bank.

Bjørnland, H. C., V. H. Larsen, and J. Maih (2018, October). Oil and macroeconomic (in)stability. *American Economic Journal: Macroeconomics 10*(4), 128–51.

Blom, H. A. and Y. Bar-Shalom (1988). The interacting multiple model algorithm for systems with markovian switching coefficients. *IEEE transactions on Automatic Control 33*(8), 780–783.

Chang, Y., B. Kwak, and S. Qiu (2021). U.S. Monetary and Fiscal Policy Regime Changes and Their Interactions. Mimeo, Indiana University.

Chang, Y., J. Maih, and F. Tan (2021). Origins of Monetary Policy Shifts: A New Approach to Regime Switching in DSGE Models. *Journal of Economic Dynamics and Control 133*, 104235.

Chen, X., T. Kirsanova, and C. Leith (2017). How Optimal is US Monetary Policy? *Journal of Monetary Economics 92*, 96–111.

Chen, X., E. M. Leeper, and C. Leith (2022). Strategic Interactions in U.S. Monetary and Fiscal Policies. *Quantitative Economics 13*(2), 593–628.

Davig, T. and T. Doh (2014). Monetary Policy Regime Shifts and Inflation Persistence. *The Review of Economics and Statistics 96*(5), 862–875.

De Jong, P. (1988). A cross-validation filter for time series models. *Biometrika 75*(3), 594–600.

Durbin, J. and S. J. Koopman (2012). *Time Series Analysis by State Space Models.* Oxford: Oxford University Press.

Fernandez-Villaverde, J., P. A. Guerron-Quintana, and J. Rubio-Ramirez (2015). Estimating Dynamic Equilibrium Models with Stochastic Volatility. *Journal of Econometrics 185*(1), 216–229.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica 57*(2), 357–384.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering 82*(1), 35–45.

Karaboga, D. and B. Basturk (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (abc) algorithm. *Journal of Global Optimization 39*, 459–471.

Kim, C. and C. Nelson (1999). *State-space Models with Regime Switching: Classical and Gibbs-sampling Approaches with Applications*. MIT Press.

Kim, C.-J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics 60*(1-2), 1–22.

Leith, C., T. Kirsanova, C. Machado, and A. P. Ribeiro (2024). (Re)Evaluating recent macroeconomic policy in the US. Working Paper X, University of Glasgow.

Liu, Z., P. Wang, and T. Zha (2013, May). Land Price Dynamics and Macroeconomic Fluctuations. *Econometrica 81*(3), 1147–1184.

Maih, J. (2015). Efficient perturbation methods for solving regime-switching DSGE models. Working Paper 2015/01, Norges Bank.

Online Appendix

to

On Bayesian Filtering for Markov Regime Switching Models

by

Nigar Hashimzade Oleg Kirsanov Tatiana Kirsanova Junior Maih

A Selected Algorithms

Let $M_{j,t}^{y} = \{Z_{j,t}, c_{y,j,t}, T_{j,t}, c_{\alpha,j,t}, g_{j,t}, R_{j,t}; y_t\}$ be state-space system matrices for regime j and information at time t. Let $\mathcal{K}()$ be a KF operator. The filtering algorithms are summarised in Tables A1-A2. Smoothing algorithms are summarised in Table A3.

Table A1: GPB Filtering Algorithms

GPB(1)	GPB(2)
Regime probabilities	
$\mu_{t\|t}^{j} := \Pr\left[s_t = j \mid Y_t\right]$	$\mu_{t-1\|t}^{ij} = \Pr\left[s_{t-1} = i, s_t = j \mid Y_t\right]$ $\mu_{t\|t}^{j} = \sum_{i=1}^{h} \mu_{t-1\|t}^{ij}$
Initialisation	
$\alpha_{t-1\|t-1}, P_{t-1\|t-1}, \mu_{t-1\|t-1}^{j}$	$\alpha_{t-1\|t-1}^{i}, P_{t-1\|t-1}^{i}, \mu_{t-1\|t-1}^{i}$
Filtering and Updating	
$\left[\alpha_{t\|t-1}^{j}, P_{t\|t-1}^{j}, v_{t\|t-1}^{j}, \alpha_{t\|t}^{j}, P_{t\|t}^{j}\right]$ $= \mathcal{K}\left(M_{j,t}^{y}; \alpha_{t-1\|t-1}, P_{t-1\|t-1}\right)$ $\Lambda_t^{j} = (2\pi)^{-t/2} \left\|F_{j,t}\right\|^{-1/2} e^{-\frac{1}{2} v_{t\|t-1}^{j\prime} F_{j.t}^{-1} v_{t\|t-1}^{j}}$	$\left[\alpha_{t\|t-1}^{ij}, P_{t\|t-1}^{ij}, v_{t\|t-1}^{ij}, \alpha_{t\|t}^{ij}, P_{t\|t}^{ij}\right]$ $= \mathcal{K}\left(M_{j,t}^{y}; \alpha_{t-1\|t-1}^{i}, P_{t-1\|t-1}^{i}\right)$ $\Lambda_t^{ij} = (2\pi)^{-t/2} \left\|F_{i,j,t}\right\|^{-1/2} e^{-\frac{1}{2} v_{t\|t-1}^{ij\prime} F_{i,j,t}^{-1} v_{t\|t-1}^{ij}}$
Collapsing (dimension reduction) and Probabilities update	
$\mu_{t\|t}^{j} = \frac{\Lambda_t^{j} \sum_{i=1}^{h} Q_{t-1,t}^{ij} \mu_{t-1\|t-1}^{i}}{\sum_{k,m=1}^{h} \Lambda_t^{m} Q_{t-1,t}^{km} \mu_{t-1\|t-1}^{k}}$ $\alpha_{t\|t} = \sum_{j=1}^{h} \mu_{t\|t}^{j} \alpha_{t\|t}^{j}$ $P_{t\|t} = \sum_{i=1}^{h} \mu_{t\|t}^{j} \Big(P_{t\|t}^{j}$ $+ \left(\alpha_{t\|t} - \alpha_{t\|t}^{j}\right)\left(\alpha_{t\|t} - \alpha_{t\|t}^{j}\right)'\Big)$	$\mu_{t-1\|t}^{ij} = \frac{\Lambda_t^{ij} Q_{t-1,t}^{ij} \mu_{t-1\|t-1}^{i}}{\sum_{k=1}^{h} \Lambda_t^{kj} Q_{t-1,t}^{kj} \mu_{t-1\|t-1}^{k}}$ $\alpha_{t\|t}^{j} = \sum_{i=1}^{h} \mu_{t-1\|t}^{ij} \alpha_{t\|t}^{ij}$ $P_{t\|t}^{j} = \sum_{i=1}^{h} \mu_{t-1\|t}^{ij} \Big(P_{t\|t}^{ij}$ $+ \left(\alpha_{t\|t}^{j} - \alpha_{t\|t}^{ij}\right)\left(\alpha_{t\|t}^{j} - \alpha_{t\|t}^{ij}\right)'\Big)$ $\mu_{t\|t}^{j} = \sum_{i=1}^{h} \mu_{t-1\|t}^{ij}$

Table A2: IMM Filtering Algorithms

IMM(1)	IMM(2)
Regime probabilities	
$\mu_{t\|t}^{j} := \Pr\left[s_t = j \mid Y_t\right]$	$\mu_{t-1\|t}^{ij} = \Pr\left[s_{t-1} = i, s_t = j \mid Y_t\right]$
Initialisation	
$\alpha_{t-1\|t-1}^{i}, P_{t-1\|t-1}^{i}, \mu_{t-1\|t-1}^{i}$	$\alpha_{t-1\|t-1}^{ki}, P_{t-1\|t-1}^{ki}, \mu_{t-1\|t-1}^{ki}$
Mixing (dimension reduction)	
$\mu_{t-1\|t-1}^{i\|j} = \frac{Q_{t-1,t}^{ij}\mu_{t-1\|t-1}^{i}}{\sum_{k=1}^{h} Q_{t-1,t}^{kj}\mu_{t-1\|t-1}^{k}}$ $\alpha_{t-1\|t-1}^{0j} = \sum_{i=1}^{h}\mu_{t-1\|t-1}^{i\|j}\alpha_{t-1\|t-1}^{i}$ $P_{t-1\|t-1}^{0j} = \sum_{i=1}^{h}\mu_{t-1\|t-1}^{i\|j}\Big(P_{t-1\|t-1}^{i}$ $+\left(\alpha_{t-1\|t-1}^{i} - \alpha_{t-1\|t-1}^{0i}\right)$ $\times\left(\alpha_{t-1\|t-1}^{i} - \alpha_{t-1\|t-1}^{0i}\right)^{\prime}\Big)$	$\mu_{t-1\|t-1}^{ki\|ij} = \frac{\mathcal{Q}_{k(h-1)+i,i(h-1)+j}\mu_{t-2\|t-1}^{ki}}{\sum_{m,l=1}^{h}\mathcal{Q}_{m(h-1)+l,l(h-1)+j}\mu_{t-2\|t-1}^{ml}}$ $\alpha_{t-1\|t-1}^{0ij} = \sum_{k,i=1}^{h}\mu_{t-1\|t-1}^{ki\|ij}\alpha_{t-1\|t-1}^{ki}$ $P_{t-1\|t-1}^{0ij} = \sum_{k,i=1}^{h}\mu_{t-1\|t-1}^{ki\|ij}\Big(P_{t-1\|t-1}^{ki}$ $+\left(\alpha_{t-1\|t-1}^{ki} - \alpha_{t-1\|t-1}^{0ki}\right)$ $\times\left(\alpha_{t-1\|t-1}^{ki} - \alpha_{t-1\|t-1}^{0ki}\right)^{\prime}\Big)$
Filtering and Updating	
$\left[\alpha_{t\|t-1}^{j}, P_{t\|t-1}^{j}, v_{t\|t-1}^{j}, \alpha_{t\|t}^{j}, P_{t\|t}^{j}\right]$ $= \mathcal{K}\left(M_{j,t}^{y}; \alpha_{t-1\|t-1}^{0j}, P_{t-1\|t-1}^{0j}\right)$ $\Lambda_t^{j} = (2\pi)^{-t/2}\left\|F_{j,t}\right\|^{-1/2} e^{-\frac{1}{2}v_{t\|t-1}^{j\prime}F_{j,t}^{-1}v_{t\|t-1}^{j}}$	$\left[\alpha_{t\|t-1}^{ij}, P_{t\|t-1}^{ij}, v_{t\|t-1}^{ij}, \alpha_{t\|t}^{ij}, P_{t\|t}^{ij}\right]$ $= \mathcal{K}\left(M_{j,t}^{y}; \alpha_{t-1\|t-1}^{0ij}, P_{t-1\|t-1}^{0ij}\right)$ $\Lambda_t^{ij} = (2\pi)^{-t/2}\left\|F_{ij,t}\right\|^{-1/2} e^{-\frac{1}{2}v_{t\|t-1}^{ij\prime}F_{ij,t}^{-1}v_{t\|t-1}^{ij}}$
Probabilities update	
$\mu_{t\|t}^{j} = \frac{\Lambda_t^{j}\sum_{i=1}^{h} Q_{t-1,t}^{ij}\mu_{t-1\|t-1}^{i}}{\sum_{k,m=1}^{h}\Lambda_t^{m} Q_{t-1,t}^{km}\mu_{t-1\|t-1}^{k}}$	$\mu_{t-1\|t}^{ij} = \frac{\Lambda_t^{ij} Q_{t-1,t}^{ij}\mu_{t-1\|t-1}^{i}}{\sum_{k=1}^{h}\Lambda_t^{kj} Q_{t-1,t}^{kj}\mu_{t-1\|t-1}^{k}}$

Table A3: Smoothing Algorithms

GPB(1) and IMM(1)	GPB(2) and IMM(2)
Smoothed Probabilities	
$\mu_{t\|n}^{j} = \sum_{k=1}^{h}\mu_{t+1\|n}^{k}\frac{\mu_{t\|t}^{j}Q_{t,t+1}^{jk}}{\sum_{m=1}^{h}Q_{t,t+1}^{mk}\mu_{t\|t}^{m}}$	
Smoothed States	
Initialisaion	
$r_{n\|n-1}^{i} = Z_{i,n}^{\prime}F_{i,n}^{-1}v_{n\|n-1}^{i}$	$r_{n\|n-1}^{ij} = Z_{j,n}^{\prime}F_{ij,n}^{-1}v_{n\|n-1}^{ij}$
Recursion	
$L_{t+1,t}^{ij} = T_{j,t+1}\left(I - K_{i,t}Z_{i,t}\right)$ $r_{t\|t-1}^{i} = Z_{i,t}^{\prime}F_{i,t}^{-1}v_{t\|t-1}^{i}$ $+\sum_{j=1}^{h}Q_{t,t+1}^{ij}L_{t+1,t}^{ij\prime}r_{t+1\|t}^{j}$ $\alpha_{t\|n}^{i} = \alpha_{t\|t-1}^{i} + P_{t\|t-1}^{i}r_{t\|t-1}^{i}$	$L_{t+1,t}^{ijk} = T_{k,t+1}\left(I - K_{i,j,t}Z_{j,t}\right)$ $r_{t\|t-1}^{ij} = Z_{t,j}^{\prime}F_{i,j,t}^{-1}v_{t\|t-1}^{ij}$ $+\sum_{k=1}^{h}Q_{t,t+1}^{jk}L_{t+1,t}^{ijk\prime}r_{t+1\|t}^{jk}$ $\alpha_{t\|n}^{ij} = \alpha_{t\|t-1}^{ij} + P_{t\|t-1}^{ij}r_{t\|t-1}^{ij}$
Merge states	
$\alpha_{t\|n} = \sum_{i=1}^{h}\mu_{t\|n}^{i}\alpha_{t\|n}^{i}$	$\alpha_{t\|n}^{j} = \sum_{i=1}^{h}\mu_{t-1\|t}^{ij}\alpha_{t\|n}^{ij}$ $\alpha_{t\|n} = \sum_{i=1}^{h}\mu_{t\|n}^{i}\alpha_{t\|n}^{i}$

B The Model

This section summarises the model in Fernandez-Villaverde et al. (2015). We present the list of variables and all the model equations. We then present parameterisation of the model used in Section 3, and estimated parameters obtained in the empirical investigation discussed in Section 4.

Table B1: List of Variables

d_t	Shifter to intertemp. preference	C_t	Consumption
G_t	Government consumption	Λ_t	Marginal utility of consumption
r_t	gross nominal interest rate	R_{kt}	Rental rate of capital
π_t	Gross inflation	ϕ_t	Cost of use of capital
Q_t	Tobin's Q	ϕ'_t	derivative of the capital adj. cost
X_t	Investment	u_t	capital utilization
s_t	Investment adjustment cost	s'_t	derivative of invest. adj. cost
f_t	Calvo wage parameter	$W_{*,t}$	Optimal real wage
W_t	real wage	$l_{d,t}$	labor demand
φ_t	labor supply shifter	$\pi_{*w,t}$	Relative optimal real wage
$g_{1,t}$	Calvo price process 1	$\pi_{*,t}$	Relative Price
$g_{2,t}$	Calvo price process 2	mc_t	Real marginal cost
$Y_{d,t}$	Output	$v_{p,t}$	Price dispersion
K_t	Capital	A_t	Neutral technology
Z_t	Combined technology	MU_t	Investment-specific tech. level
$v_{w,t}$	Wage dispersion	l_t	hours worked/labor supply
$\varepsilon_{\xi,t}$	Monetary policy shock, scale σ_ξ	$\varepsilon_{\varphi,t}$	labor supply shock, with scale σ_φ
$\varepsilon_{g,t}$	Government spending shock, scale σ_g	$\varepsilon_{\mu,t}$	Invest.-spec. technology shock, scale σ_μ
$\varepsilon_{d,t}$	Preference shock, scale σ_d	$\varepsilon_{A,t}$	Neutral technology shock, scale σ_a

Table B2: Model Equations

Households	
Capital accum-n	$K_t = (1-\delta) K_{t-1} + MU_t \left(1 - s\left[\frac{X_t}{X_{t-1}}\right]\right) X_t$
FOC consum-n	$\frac{d_t}{C_t - hC_{t-1}} - h\beta E_t \frac{d_{t+1}}{C_{t+1} - hC_t} = \Lambda_t$
FOC bonds	$\Lambda_t = \beta E_t \Lambda_{t+1} \frac{r_t}{\pi_{t+1}}$
FOC capital util.	$R_{kt} = \frac{\phi'[u_t]}{MU_t}$
FOC capital	$Q_t = \beta E_t \frac{\Lambda_{t+1}}{\Lambda_t} \left((1-\delta) Q_{t+1} + R_{kt+1} u_{t+1} - \frac{\phi[u_{t+1}]}{MU_{t+1}}\right)$
Capital util-n	$\phi[u] = \phi_1 (u-1) + \frac{\phi_2}{2}(u-1)^2$
its derivative	$\phi'[u] = \phi_1 + \frac{\phi_2}{2}(u-1)$
FOC investment	$1 = Q_t MU_t \left(1 - s\left[\frac{X_t}{X_{t-1}}\right] - s'\left[\frac{X_t}{X_{t-1}}\right]\frac{X_t}{X_{t-1}}\right)$
	$+\beta E_t Q_{t+1} MU_{t+1} \frac{\Lambda_{t+1}}{\Lambda_t} s'\left[\frac{X_{t+1}}{X_t}\right]\left(\frac{X_{t+1}}{X_t}\right)^2$
Invest. adj. cost	$s\left[\frac{X_t}{X_{t-1}}\right] = \frac{\kappa}{2}\left(\frac{X_t}{X_{t-1}} - \lambda_x\right)^2$
its derivative	$s'\left[\frac{X_t}{X_{t-1}}\right] = \kappa\left(\frac{X_t}{X_{t-1}} - \lambda_x\right)$
Firms	
Wage helper 1	$f_t = \frac{\eta-1}{\eta}(W_{*,t})^{1-\eta}\Lambda_t W_t^{\eta} l_{d,t} + \beta\theta_w E_t \left(\frac{\pi_t^{\chi_w}}{\pi_{t+1}}\right)^{1-\eta}\left(\frac{W_{*,t+1}}{W_{*,t}}\right)^{\eta-1} f_{t+1}$
Wage helper 2	$f_t = \psi d_t \varphi_t \pi_{*w,t}^{-\eta(1+\vartheta)} l_{d,t}^{(1+\vartheta)} + \beta\theta_w E_t \left(\frac{\pi_t^{\chi_w}}{\pi_{t+1}}\right)^{-\eta(1+\vartheta)}\left(\frac{W_{*,t+1}}{W_{*,t}}\right)^{\eta(1+\vartheta)} f_{t+1}$
Wage setting	$\pi_{*w,t} = \frac{W_{*,t}}{W_t}$
Wage dynamics	$1 = \theta_w \left(\frac{\pi_{t-1}^{\chi_w}}{\pi_t}\right)^{1-\eta}\left(\frac{W_{t-1}}{W_t}\right)^{1-\eta} + (1-\theta_w)\pi_{*w,t}^{1-\eta}$
Wage dispersion	$v_{w,t} = \theta_w \left(\frac{W_{t-1}}{W_t}\frac{\pi_{t-1}^{\chi_w}}{\pi_t}\right)^{-\eta} v_{w,t-1} + (1-\theta_w)\pi_{*w,t}^{-\eta}$
Price helper 1	$g_{1,t} = \Lambda_t mc_t Y_{d,t} + \beta\theta_p E_t \left(\frac{\pi_t^{\chi}}{\pi_{t+1}}\right)^{-\varepsilon} g_{1,t+1}$
Price helper 2	$g_{2,t} = \Lambda_t \pi_{*,t} Y_{d,t} + \beta\theta_p E_t \left(\frac{\pi_t^{\chi}}{\pi_{t+1}}\right)^{1-\varepsilon}\left(\frac{\pi_{*,t}}{\pi_{*,t+1}}\right) g_{2,t+1}$
Price setting	$\varepsilon g_{1,t} = (\varepsilon - 1) g_{2,t}$
Price dynamics	$1 = \theta_p \left(\frac{\pi_{t-1}^{\chi}}{\pi_t}\right)^{1-\varepsilon} + (1-\theta_p)\pi_{*,t}^{1-\varepsilon}$
Price dispersion	$v_{p,t} = \theta_p \left(\frac{\pi_{t-1}^{\chi}}{\pi_t}\right)^{-\varepsilon} v_{p,t-1} + (1-\theta_p)\pi_{*,t}^{-\varepsilon}$

continued on the next page

Table B2: Model Equations – continued

Market Clearing and Policy	
Production function	$Y_{d,t} = \frac{A_t(u_t K_{t-1})^\alpha \left(l_{d,t}\right)^{1-\alpha} - \phi_y Z_t}{v_{p,t}}$
Capital-labor ratio	$\frac{u_t K_{t-1}}{l_{d,t}} = \frac{\alpha}{1-\alpha} \frac{W_t}{R_{kt}}$
Aggregate labour	$l_t = v_{w,t} l_{d,t}$
Resource constraint	$Y_{d,t} = C_t + G_t + X_t + \frac{\phi[u_t]}{MU_t} K_{t-1}$
Marginal costs	$mc_t = \left(\frac{1}{1-\alpha}\right)^{1-\alpha} \left(\frac{1}{\alpha}\right)^\alpha \frac{W_t^{1-\alpha} R_{kt}^\alpha}{A_t}$
Taylor rule	$\frac{r_t}{r_{ss}} = \left(\frac{r_{t-1}}{r_{ss}}\right)^{\gamma_r} \left(\left(\frac{\pi_t}{\pi_{\text{targ}}}\right)^{\gamma_\pi} \left(\frac{Y_{d,t}}{\lambda_{yd} Y_{d,t-1}}\right)^{\gamma_y}\right)^{1-\gamma_r} \exp\left(\sigma_\xi \varepsilon_{\xi,t}\right)$
Government spending	$\log\left(\frac{G_t}{Z_t}\right) = \left(1-\rho_g\right) \log g + \rho_g \log\left(\frac{G_{t-1}}{Z_{t-1}}\right) + \sigma_g \varepsilon_{g,t}$
Exogenous processes	
Intertemporal preference	$\log\left(d_t\right) = \rho_d \log\left(d_{t-1}\right) + \sigma_d \varepsilon_{d,t}$
Labor supply	$\log\left(\varphi_t\right) = \rho_\varphi \log\left(\varphi_{t-1}\right) + \sigma_\varphi \varepsilon_{\varphi,t}$
Investment-spec. technology	$MU_t = MU_{t-1} \exp\left(\lambda_\mu + \sigma_\mu \varepsilon_{\mu,t}\right)$
Neutral technology	$A_t = A_{t-1} \exp\left(\lambda_a + \sigma_a \varepsilon_{A,t}\right)$
Combined technology	$Z_t = A_t^{\frac{1}{1-\alpha}} MU_t^{\frac{\alpha}{1-\alpha}}$

C Model Parameters and Empirics

Table C1: Model Parameters

Parameters	Description	FGR2015 Values	Estimated Values
		(1)	(2)
β	Time Preference	0.99	0.9992
h	Habit Formation	0.9	0.92747
psi	labor supply coeff in utility	8.0	
vartheta	Disutilty of Labor Scaling	1.17	
δ	Depreciation Rate	0.025	
α	Captial Share in Production	0.21	0.14991
κ	Weight on Investment Adjustment Costs	9.5	3.7946
ε	Elast. of Subst. btw. Differntiated Goods	10	
eta	Elast. of Subst. btw Diff. Types of Labour	10	
phi2	Weight on Adj. Costs for Capital Utilization	0.001	
χ_w	Wage Indexation	0.6340	
χ	Price Indexation	0.6186	0.00011223
θ_w	Probability of not changing wages	0.6869	
θ_p	Probability of not changing prices	0.8139	0.8379

continued on the next page

Table C1: Model Parameters – continued

Parameters	Description	FGR2015 Values	Estimated Values
		(1)	(2)
Policy Parameters			
$\gamma_r\,(S_{P=1})$	Interest rate smoothing	0.7855	0.80302
$\gamma_r\,(S_{P=2})$	Interest rate smoothing	0.7855	0.87472
$\gamma_y\,(S_{P=1})$	Reaction to output growth	exp(-1.4034)	0.41649
$\gamma_y\,(S_{P=2})$	Reaction to output growth	exp(-1.4034)	0.32918
π_{targ}	Inlfation target	1.0005	1.0057
Persistence of Shocks			
ρ_d	Consumption preference	0.1182	0.72554
ρ_φ	Labor supply	0.9331	0.92016
λ_a	Neutral technology	0.0028	
λ_μ	Investment-specific technology	0.0034	
ρ_g	Government spending shock	0.75	0.0024123
Standard Deviation of Shocks			
$\sigma_d\,(S_{V=1})$	Consumption Preference Regime 1	exp(-1.9834)	0.031211
$\sigma_d\,(S_{V=2})^*$	Consumption Preference Regime 2	2*exp(-1.9834)	0.16311
$\sigma_\varphi\,(S_{V=1})$	Labor Supply Regime 1	exp(-2.4983)	0.29752
$\sigma_\varphi\,(S_{V=2})^*$	Labor Supply Regime 2	2*exp(-2.4983)	0.20313
$\sigma_\mu\,(S_{V=1})$	Investment-specific technology Regime 1	exp(-6.0283)	0.003998
$\sigma_\mu\,(S_{V=2})^*$	Investment-specific technology Regime 2	2*exp(-6.0283)	0.0071897
$\sigma_a\,(S_{V=1})$	Neutral technology Regime 1	exp(-3.9013)	0.037752
$\sigma_a\,(S_{V=2})^*$	Neutral technology Regime 2	2*exp(-3.9013)	0.048111
σ_g	Government Spending shock	exp(-3.9013)	0.0074137
σ_ξ	Monetary Policy	exp(-6.000)	0.0020286

Note: * denotes a parameter which is calibrated by the authors.

Table C2: Data Sources

Data series	Description	Units	FRED series
		(1)	(2)
DY_DATA	Output Growth	%pa	A939RX0Q048SBEA
DP_DATA	Inflation rate	%pa	GDPDEF
R_DATA	Federal Funds Rate	%pa	FEDFUNDS
DW_DATA	Wage Inflation	%pa	COMPRNFB
DMU_DATA	Relat. price of invest. goods	%pa	PIRIC

Note: FRED database https://fred.stlouisfed.org/

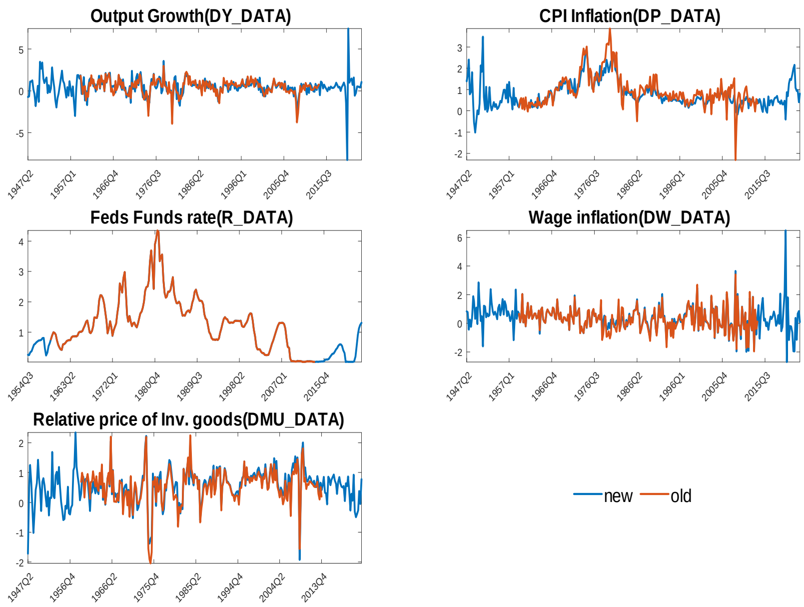

Figure C1: Updated Data

D Notes on Implementation

All computations in this paper were coded in RISE© (Maih, 2015). RISE toolkit accepts the model description as a text file containing list of commands and mathematical expressions, converts it into a state-space form, loads the data, and applies filters and smoothers discussed in this paper.

In panel A of Table 3, the speed results were obtained with implementing IMM(1) and KN-GPB(2) as stand-alone procedure, with optimisation for speed where possible. As discussed in the text, their implementation in RISE allows handling various non-linearities and missing observations.

In panel B of Table 3, for comparison of speed within either GPB or IMM families we employ single GPB(N) and IMM(N) filters that accept an arbitrary order as input, rather than separate codes for different orders of filtration. Consequently, the number of nested loops is not predetermined but is managed throughout the computation.