TI 2024-030/III
Tinbergen Institute Discussion Paper

Taylor Rules with Endogenous Regimes

Knut Are Aastveit[1]
Jamie Cross[2]
Francesco Furlanetto[3]
Herman K. Van Dijk[4]

1 Norges Bank and BI Norwegian Business School

2 The University of Melbourne, & BI Norwegian Business School

3 Norges Bank

4 Erasmus University Rotterdam, Tinbergen Institute and Norges Bank

Taylor Rules with Endogenous Regimes*

Knut Are Aastveit[†], Jamie Cross[‡], Francesco Furlanetto[§], Herman K. Van Dijk[¶]

May 6, 2024

Abstract

The Fed's policy rule switches during the different phases of the business cycle. This finding is established using a dynamic mixture model to estimate regime-dependent Taylor-type rules on US quarterly data from 1960 to 2021. Instead of exogenously partitioning the data based on tenures of the Fed chairs, a Bayesian framework is introduced in order to endogenously select timing and number of regimes in a data-driven way. This agnostic approach favors a partitioning of the data based on two regimes related to business cycle phases. Estimated policy rule coefficients differ in two important ways over the two regimes: the degree of gradualism is substantially higher during normal times than in recessionary periods while the output gap coefficient is higher in the recessionary regime than in the normal one. The estimate of the inflation coefficient largely satisfies the Taylor principle in both regimes. The results are substantially reinforced when using real-time data.

Keywords: Monetary policy, Taylor rules, mixed distributions, regime-switching
JEL classification codes: C32, C51, E42, E52, E58

*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. We would like to thank Fabio Canova, Refet Gürkaynak, Christian Matthes, Gary Koop, Lorenzo Mori, seminar participants at Norges Bank, University of Glasgow, ESOBE conference in Salzburg, SNDE conference in Orlando, IAAE conference in Oslo, CFE conference in London, Norges Bank Mini Workshop on Monetary Policy, and Workshop in Empirical Macroeconomics at King's college for useful discussions and comments on a preliminary draft of this paper.

[†]Norges Bank & BI Norwegian Business School

[‡]Melbourne Business School, The University of Melbourne, & BI Norwegian Business School.

[§]Norges Bank

[¶]Erasmus University, Tinbergen Institute and Norges Bank

1 Introduction

From the late 1960s to the early 1980s, the US economy experienced a period of deep and frequent recessions. Since the early 1980s, however, macroeconomic dynamics changed: the inflation rate stabilized at low levels in the context of rare recessions. The leading explanation for such a better macroeconomic performance has been related to a change in the conduct of monetary policy, as shown in the seminal paper by Clarida et al. (2000). According to this view, the appointment of Paul Volcker as chair in 1979 led to a shift in the response of the Federal Reserve to inflationary pressures from highly accomodative to highly responsive. More specifically, the Federal Reserve started raising the nominal interest rate above one in response to an increase in inflation, thus increasing the real rate and satisfying the Taylor principle that guarantees equilibrium determinacy in most New Keynesian models. While challenged by a series of papers highlighting the role of lower shocks volatility to explain the more stable macroeconomic environment (cf. Sims and Zha (2006) and the references therein), the original result of Clarida et al. (2000) has been confirmed by a large subsequent literature (see Lubik and Schorfheide (2004) among many others).

One assumption that characterizes the analysis of Clarida et al. (2000) and most of the subsequent literature (Coibion and Gorodnichenko, 2011) has received relatively little scrutiny: the exogenous partitioning of the data into a pre-Volcker and a post-Volcker era or, alternatively, into different periods corresponding to the tenure of various Fed chairs (Carvalho et al. (2021) and Barnichon and Mesters (2023)). This practice leaves open the possibility that changes in the conduct of monetary policy are coincidental (rather than causal) and/or that these are more related to the occurrence of specific economic conditions rather than to the leadership of a specific Fed chair, as discussed in Bianchi (2013). It is in fact conceivable that the pre-Volcker and post-Volcker periods feature a different composition of shocks, a different balance of booms and recessions or of any other economic conditions across the two periods. If monetary policy is state-dependent, the prevalence of a specific state may also influence the estimated coefficients of simple reaction functions.

One possible solution is to adopt a methodology that endogenously partitions the data into different regimes (not necessarily adjacent as in the pre-Volcker/post-Volcker example) and compares the behavior of monetary policy across regimes. The goal of this paper is to propose such a framework based on a novel dynamic mixture model of the Taylor rule (Taylor, 1993) with an endogenously selected number of components. In the model each observation is generated from one of the component distributions with a certain probabilistic weight. Consider the special case where we know from which component each observation comes from, then we can partition the data set into sub-periods and estimate a standard regression model on each of the these periods. The dynamic mixture model does this endogenously and in an agnostic data-driven manner. It is important to emphasize that the mixture model has been used extensively in economics but it has been mainly applied to cross-sectional data (see Compiani and Kitamura (2016) for an extensive review of the literature and Frühwirth-Schnatter (2006) for a textbook treatment). Mixture models have also been used for clustering the cross-sectional dimension of a panel of time series (see Paap and Van Dijk (1998), Canova (2004) and Frühwirth-Schnatter and Kaufmann (2008) among others), for forecasting purposes (see Casarin et al. (2023) among others) and for model combination or selection (see Waggoner and Zha (2012) and Loria et al. (2022)). However, to the best of our knowledge, we are the first to apply the mixture model to estimate simple macroeconomic relationships like monetary policy rules.

In terms of empirical results we obtain credible evidence about the following.

First, given our model, the data favor a partitioning based on two regimes. The timing of the regimes and the implications for the evolution of monetary policy, however, are substantially different from the conventional approach based on exogenous partitioning. Our two regimes correspond roughly to one regime, labeled normal, corresponding to economic expansions and a second more recessionary regime, which materializes occasionally and mainly in the first part of the data period where most recessions are concentrated.

Second, the estimated policy rule coefficients are different in the two regimes. The main difference concerns the degree of interest rate smoothing. In fact, we find that the estimated degree of gradualism is credibly lower in the recessionary regime with estimates in the range of 0.5 to 0.7, depending on the specification, against a value of around 0.9 in the normal regime. Such an asymmetric behavior of policy reflects the fact that contractions are shorter but swallower than expansions, in keeping with the properties of many macroeconomic time series. Notably, the lower degree of gradualism in the recessionary regime is reinforced in a battery of sensitivity exercises. Another result emerging from our estimation is that the estimate of the inflation coefficient is rather similar across regimes and largely satisfies the Taylor principle in both regimes. Interestingly, the output gap coefficient is higher in the recessionary regime. While the difference in these estimates across regimes is slightly less strong than in the case of the gradualism coefficient, it appears robust in our exercises. Somewhat intuitively, it signals that the central bank cares relatively more about output dynamics in downturns.

In a nutshell, the main result of our paper is that the data seem to prefer a partitioning based on the state of the business cycle rather than on the tenure of a specific Fed chair. Yet, when imposing an exogenous partitioning in 1979:Q3 and estimating the mixture model on both the pre-Volcker and post-Volker periods, we recover the familiar result that the Fed has responded more aggressively to inflation over time, as in Clarida et al. (2000). In addition, when we impose dogmatically the presence of a third regime, such a regime features a low response to inflation and is prevalent at the beginning of the data period. Therefore, our model can identify time-variation in the response to inflation but endogenously choose to put more weight on the asymmetry in gradualism based on the state of the business cycle. Importantly, such asymmetry in gradualism is substantially larger when using the Greenbook/Tealbook data. Therefore, our main result is robust to the use of real-time data, unlike the asymmetry in the response to inflation which is substantially weaker when using real-time data in models with exogenous partitioning, as shown by Coibion and Gorodnichenko (2011).

The results of this paper contribute to the literature on the estimation of Taylor-type rules. Following the seminal contribution of Clarida et al. (2000), several papers have extended the baseline framework by increasing the set of regressors (Bernanke and Gertler (2000) and Castelnuovo (2003) among many others), using real-time data (Orphanides (2001) and Coibion and Gorodnichenko (2011)), taking a cross-country perspective (Clarida et al. (1998)) or using multivariate systems (Arias et al., 2019). Important earlier contributions on non linearities in the policy rule include Dolado et al. (2004) and Surico (2007).[1] We contribute to this literature by using the dynamic mixture model that so far has not been applied to estimate policy rules.

Within the literature on monetary policy rules, our results are related to a series of papers discussing the role of gradualism in monetary policy. In a wellknown speech, Bernanke (2004) listed the benefits of conducting gradual adjustments in monetary policy. From a normative point of view, Woodford (2003) and Caballero and Simsek (2022) propose models in which smoothing emerges as a feature of optimal monetary policy. In contrast, Rudebusch (2006) argues that the estimated degree of monetary policy inertia reflects the role of omitted variables in the reaction function which manifest itself in the form of persistent monetary policy shocks. If gradualism was a proper feature of monetary policy, according to Rudebusch (2006), the interest rate should be highly predictable. While this debate is not settled, both Coibion and Gorodnichenko (2012) and Caballero and Simsek (2022) have provided counterarguments and substantial additional evidence that gradualism is in fact a feature of modern monetary policy. Interestingly, while the overwhelming majority of empirical papers on policy rules estimate a substantial degree of gradualism, little evidence on the state-dependence of gradualism has been provided. One exception is Florio (2006) who estimates a smooth transition regression that allows for asymmetric interest rate smoothing.

We note that several alternatives to our endogenous partitioning exist in the literature. These refer to the estimation of policy rules with time varying parameters either in a single equation context (Boivin (2006) and Coibion and Gorodnichenko (2011)) or in a

[1] Most of the more recent literature focuses on the state-dependent effects of monetary policy *shocks* but ignores the *systematic* component of policy which obviously affects in and of itself the propagation of shocks. A non-exaustive list of recent contributions include Aastveit and Anundsen (2022), Alpanda et al. (2021), Aruoba et al. (2022), Ascari and Haber (2022), Barnichon and Matthes (2018), Debortoli et al. (2023), Eichenbaum et al. (2022), Gargiulo et al. (2024) and Tenreyro and Thwaites (2016).

Structural Vector Autoregressions with time-varying parameters and stochastic volatility with a focus on the interest rate equation (Cogley and Sargent (2005), Primiceri (2005), Aastveit et al. (2023)). Of particular interest are models which allow explicitly for policy regimes like Owyang and Ramey (2004) and Sims and Zha (2006). We differ in using a simple methodology which imposes little structure and is thereby less prone to specification errors. This becomes advantageous for estimating and classifying regimes in cases where the data period includes brief and abrupt episodes, such as recessions.

The contents of this paper is organized as follows: Section 2 describes our econometric approach. Section 3 contains our empirical results. Section 4 considers alternative specifications for mixture weights. Section 5 puts our results in perspective with respect to the existing literature. Section 6 concludes. Some additional results are provided in the Supplementary Material that serves as an online Appendix.

2 Model Structure and Inference Procedure

In this Section, we describe the dynamic mixture model starting from a Bayesian interpretation of the Taylor rule regression and then describing the details of the estimation of our model.

2.1 The baseline monetary policy rule in a Bayesian context

We rely on the specification used by Carvalho et al. (2021) who show that the monetary policy rule can be safely estimated with OLS, without the need for instrumental variables, because the monetary policy shocks are so small that they induce only a minimal bias in the estimated coefficients.[2] They estimate the following reduced form regression

$$r_t = \alpha_{\text{aux}} + \rho_{1,\text{aux}} r_{t-1} + \rho_{2,\text{aux}} r_{t-2} + \beta_{\text{aux}} \pi_t + \gamma_{\text{aux}} y_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2), \tag{1}$$

and solve for the structural parameters using the relationships: $\rho = \rho_{1,\text{aux}} + \rho_{2,\text{aux}}$, $\beta = \beta_{\text{aux}}/(1-\rho)$, $\gamma = \gamma_{\text{aux}}/(1-\rho)$. These three parameters are our object of interest and summarize the degree of gradualism, the response to inflation and the response to the output gap, respectively. The constant α_{aux} can be interpreted as the natural rate of interest. The policy rate (r_t) is the effective federal funds rate, inflation (π_t) is the year-on-year rate of change in core PCE, and the output gap (y_t) is constructed using the Congressional Budget Office (CBO) estimate of potential GDP. The use of two lags of the interest rate is consistent with the preferred specification in Coibion and Gorodnichenko (2012) who find strong evidence for interest rate smoothing of order two.

[2] Carvalho et al. (2021) show that the asymptotic OLS estimation bias is proportional to the fraction of the variance of regressors due to monetary policy shocks. Since monetary policy shocks explain only a small fraction of the variance of typical Taylor rule regressors, the bias tends to be small.

The reduced form regression in (1) can be rewritten for notational convenience as:

$$r_t = \mathbf{x}_t'\boldsymbol{\beta} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2), \tag{2}$$

which implies that each observation r_t is modeled as coming from the same distribution $(r_t|\boldsymbol{\theta}) = N(\mathbf{x}_t'\boldsymbol{\beta}, \sigma^2)$, where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma^2)'$, $\boldsymbol{\beta} = (\alpha_{\text{aux}}, \rho_{1,\text{aux}}, \rho_{2,\text{aux}}, \beta_{\text{aux}}, \gamma_{\text{aux}})$ and $\mathbf{x}_t = (1, r_{t-1}, r_{t-2}, \pi_t, y_t)'$.

It is well known that parameters $\boldsymbol{\beta}$ and σ^2 from this model can be estimated in a simple simulation based Bayesian procedure using a Gibbs Sampler.[3] In online Appendix A.1, we show estimation results for the four sub-periods used by Carvalho et al. (2021) and for our full estimation data period using this Bayesian framework.

2.2 From a Taylor rule regression model to a mixture model

We move now to the description of the mixture model. We are interested in considering different monetary policy regimes (or components) in which the Taylor rule coefficients are potentially different.

If we know which component each observation comes from, we can partition the data set according to the information in the k-components and estimate a standard linear regression model for each of the sub-periods. This is then mostly done exogenously by splitting the data into sub-periods. It is problematic, however, since governors are not always hawkish/dovish; sub-periods may contain different economic conditions like recessions/expansions; and policy rule coefficients may change. As a consequence basic inferential procedures like OLS and Bayesian linear regression average these results out.

However, in many applications, it is unclear which distribution r_t comes from. This includes scenarios where the researcher is uncertain about the policy stance in a given period, or when the time series of interest behaves differently over a period; perhaps due to a change in policy regime. Then it is natural to consider a model specification that allows each observation r_t to come from a possibly different distribution. This is the basis of the finite mixture model. The underlying idea of this model is that each observation is generated from one of the component distributions within a mixture model. In our case, each component represents a Taylor-type monetary policy rule (Clarida et al., 2000) and is assigned a specific probabilistic weight. If we specify that all mixture components follow a Normal distribution akin to the above regression model, then we can model each observation r_t as coming from a mixture of Normals:

[3] The algorithm works as follows: Start by choosing initial conditions $\boldsymbol{\beta}^{(0)} = \mathbf{a}$, and $\sigma^{2(0)} = b > 0$. Then, repeat the following steps from $d = 1, \ldots, D$: Draw $\boldsymbol{\beta}^{(d)} \sim f(\boldsymbol{\beta}|\mathbf{r}, \sigma^{2(d-1)})$ (multivariate Normal) and $\sigma^{2(d)} \sim f(\sigma^2|\mathbf{r}, \boldsymbol{\beta}^{(d)})$ (inverse-Gamma). In practice, the first D_0 draws are discarded as a burn-in period to eliminate any effects of the initial conditions. For more details and background, see Koop (2003).

$$f(r_t|\mathbf{p}, \mathbf{x}_t, \boldsymbol{\theta}) = \sum_{j=1}^{k} p_j f_j(r_t|\mathbf{x}_t, \boldsymbol{\theta}_j), \tag{3}$$

where $\mathbf{p} = (p_1, \ldots, p_k)'$ are the *component/mixing probabilities*, that satisfy $p_j \geq 0$ and $\sum_{j=1}^{k} p_j = 1$; $\mathbf{x}_t = (1, r_{t-1}, r_{t-2}, \pi_t, y_t)'$ are data on the explanatory variables given in equation (1); $\boldsymbol{\theta}_j = (\boldsymbol{\beta}_j', \sigma_j^2)'$ are the parameters of j-th component density with $\boldsymbol{\beta}_j = (\alpha_{j,\text{aux}}, \rho_{1j,\text{aux}}, \rho_{2j,\text{aux}}, \beta_{j,\text{aux}}, \gamma_{j,\text{aux}})'$ and σ_j^2 the variance of the disturbance term ε_{jt}. The *k-component/mixing densities* $f_j(r_t|\mathbf{x}_t, \boldsymbol{\theta}_j)$, $j = 1, \ldots, k$, are all Normal densities, i.e., $(r_t|\mathbf{x}_t, \boldsymbol{\theta}_j) \sim N\left(\mathbf{x}_t'\boldsymbol{\beta}_j, \sigma_j^2\right)$.[4]

The Taylor rule representation in (3) endogenously classifies each interest rate observation, r_t, as belonging to one of k possible mixture components (also known as regimes) conditional on the economic environment, $\mathbf{x}_t$, and associated policy response, $\boldsymbol{\theta}_j$. To see this more clearly, we introduce a latent time-varying component indicator $S_t \in \{1, \cdots, k\}$, such that the probability that S_t equals to j is equal to p_j, i.e.:

$$Pr(S_t = j|\mathbf{p}) = p_j. \tag{4}$$

Given the component labels, $\mathbf{S} = (S_1, \ldots, S_T)'$, the model can be viewed as partitioning the interest rate data, $\mathbf{r} = (r_1, \ldots, r_T)'$, into k regimes such that:

$$(r_t|S_t, \mathbf{x}_t, \boldsymbol{\theta}) \sim N(\mathbf{x}_t'\boldsymbol{\beta}_{S_t}, \sigma_{S_t}^2), \tag{5}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_k')'$ and $\boldsymbol{\theta}_{S_t} = \left(\boldsymbol{\beta}_{S_t}', \sigma_{S_t}^2\right)'$ with $S_t \in \{1, \ldots, k\}$. To see that this latent variable representation gives the same model, note that integrating out the component label gives

$$\begin{aligned} f(r_t|\mathbf{p}, \mathbf{x}_t, \boldsymbol{\theta}) &= \sum_{j=1}^{k} f(r_t|S_t = j, \mathbf{x}_t, \boldsymbol{\theta}) Pr(S_t = j|\mathbf{p}), \\ &= \sum_{j=1}^{k} p_j f(r_t|\mathbf{x}_t, \boldsymbol{\theta}_j), \end{aligned} \tag{6}$$

where the final expression is the same as (3).

We make here a remark about the interpretation of the unobserved states in a mixture process. In business cycle analysis states in the economy are interpreted as expansionary and recessionary regimes. The data-driven states from a mixture process do not have this direct interpretation. However, in our empirical analysis we find a close correspondence between the data-driven states and the economic regimes. Henceforth, we make regularly use of the term regimes to indicate the states, as shown empirically in Section 3.

[4]For a classic treatment of finite mixtures we refer to the textbook of Frühwirth-Schnatter (2006).

2.3 Estimation of the mixture model with known number of components.

We now discuss how to estimate the mixture model. For expository purposes, we consider first the case in which the number of components is known. Note that this is the case considered in Section 5.1 where we impose dogmatically the presence of three regimes.

The mixture model is more difficult to estimate than the regression model. The main reason for this is that the usual prior for the parameters of the regression model is no longer conjugate, and a standard Gibbs sampling method can not directly be applied.

This problem can be overcome by using data augmentation. As shown in (5), given the component labels in $\mathbf{S}$, the mixture model can be seen as a series of k regime-specific linear regression models, each with a Normal likelihood function. The parameter vectors $\boldsymbol{\theta}_j$, $j = 1 \ldots, k$, can therefore be estimated with standard linear regression methods using the observations from each regime.

In line with the linear regression framework, we specify standard independent Normal and inverse-Gamma priors for the regression parameters of each component, i.e., $f(\boldsymbol{\theta}_j) = f(\boldsymbol{\beta}_j, \sigma_j^2) = f(\boldsymbol{\beta}_j)f(\sigma_j^2)$, in which:

$$\boldsymbol{\beta}_j \sim N(\boldsymbol{\beta}_{j0}, \mathbf{V}_{j0}), \quad \sigma_j^2 \sim IG(\nu_{j0}, \mathbf{S}_{j0}), \quad j = 1, \ldots, k. \tag{7}$$

In principle, the econometrician has the flexibility to assign distinct values for the component-specific prior hyperparameters, however we here set them globally, i.e., $\boldsymbol{\beta}_{j0} = \boldsymbol{\beta}_0$, $\mathbf{V}_{j0} = \mathbf{V}_0$, $\nu_{j0} = \nu_0$, and $\mathbf{S}_{j0} = \mathbf{S}_0$, for all $j = 1, \ldots, k$.

So far we have discussed the estimation of the mixture model conditional on the component indicator for each period. A crucial final step is to specify an appropriate prior for the component probabilities in order to obtain a well-specified posterior distribution. We therefore use an independent Dirichlet prior (also know as the generalized Beta distribution) specified over the component probabilities:

$$\mathbf{p} \sim D(\alpha_1, \ldots, \alpha_k) \tag{8}$$

where $\alpha_j > 0$, $j = 1, \ldots, k$, are positive, real-valued, concentration parameters. The concentration parameters shape the distribution by influencing how the probability mass is spread across the components. For instance, if one α_j is much larger than the others, then the probability mass will be highly concentrated on the corresponding component. To ensure that no prior information favors one particular component over another, we therefore remain agnostic and specify a symmetric concentration parameter, $\alpha_j = \alpha$, for all $j = 1, \ldots, k$. When $\alpha = 1$, the Dirichlet distribution becomes uniform over the $(k-1)$ simplex, meaning that all possible outcomes are equally likely. When $\alpha > 1$, the distribution tends to be more concentrated around the center of the distribution simplex, indicating that the probabilities are more evenly spread across the different categories. When $\alpha < 1$, the distribution is more concentrated towards the corners and edges of the simplex, suggesting a preference for distributions where fewer categories are likely to have higher probabilities, and the rest are closer to zero.

Using Bayes rule, it is straightforward to show that the conditional posterior distributions of the mixture weights, the component indicators, and the equation parameters are, respectively:

$$(\mathbf{p}|\mathbf{r}, \mathbf{S}, \boldsymbol{\theta}) \sim D(e_1(\mathbf{S}), \ldots, e_k(\mathbf{S})), \quad (\mathbf{S}|\mathbf{r}, \mathbf{p}, \boldsymbol{\theta}) \sim M(1, \mathbf{p}, T),$$
$$(\boldsymbol{\beta}_j|\mathbf{r}, \mathbf{S}, \mathbf{p}, \boldsymbol{\theta}_{-\boldsymbol{\beta}_j}) \sim N(\boldsymbol{\mu}_j, \mathbf{V}_j), \quad (\sigma_j^2|\mathbf{r}, \mathbf{S}, \mathbf{p}, \boldsymbol{\theta}_{-\sigma_j^2}) \sim IG(\nu_j, S_j), \quad j = 1, \ldots, k, \tag{9}$$

where $\boldsymbol{\theta}_{-\gamma}$ denotes $\boldsymbol{\theta}$ excluding $\gamma \in \{\boldsymbol{\beta}_j, \sigma_j^2\}$, $e_j(\mathbf{S}) = \alpha + T_j(\mathbf{S})$, with $T_j(\mathbf{S})$ denoting the number of observations assigned to component j, and $M(1, \mathbf{p}, T)$ is used to denote T independent draws from the Multinomial posterior distribution, i.e. $S_t \sim M(1, \mathbf{p})$, $t = 1, \ldots, T$. For notation convenience, the data $\mathbf{x}_t$, $t = 1, \ldots, T$, is omitted from (9).

We emphasize that while the prior Dirichlet distribution for the mixture weights is symmetric, the posterior Dirichlet distribution is not symmetric. Intuitively, this is because the mass placed on each of the component weights is determined by the dynamic classification of each observation through the latent component indicators, that learns about the state of the macroeconomic environment through data given the model's likelihood function. To see this, note that since the component indicators are conditionally independent given the data and model parameters, the joint conditional mass function of $\mathbf{S}$ is the product of T conditionally independent mass functions:

$$f(\mathbf{S}|\mathbf{r}, \boldsymbol{\theta}, \mathbf{p}) \propto \prod_{t=1}^{T} f(r_t|S_t, \boldsymbol{\theta}) f(S_t|\mathbf{p}), \tag{10}$$

and we can sample each of them separately. At each date t, the classification probabilities are given by:

$$Pr(S_t = j|r_t, \boldsymbol{\theta}, \mathbf{p}) = c_t f(r_t|S_t = j, \boldsymbol{\theta}_j) Pr(S_t = j|p_j), \tag{11}$$

where $f(r_t|S_t = j, \boldsymbol{\theta}_j) = \phi\left(r_t; \mathbf{x}_t'\boldsymbol{\beta}_j, \sigma_j^2\right)$ is a Normal density, $j = 1, \ldots, k$, and c_t is a normalizing constant that is given by:

$$c_t = \frac{1}{\sum_{j=1}^{k} p_j \phi\left(r_t; \mathbf{x}_t'\boldsymbol{\beta}_j, \sigma_j^2\right)}. \tag{12}$$

Combining (11) and (12) shows that elements in $\mathbf{S}$ are drawn independently from the multinomial distribution in (9). The time pattern of this classification sequence is shown in Figure 2 and in later figures.

We emphasize that the conditionally independent distribution in (10) arises because the mixture model implicitly assumes that the classification sequence is independent. An alternative specification of the mixture model would be to impose more structure on this sequence. A common example of such a structure is a Markov process, i.e., $P(S_t = j | S_{t-1} = i)$, for all $i, j \in \{1, \ldots, k\}$. Another example would be to impose different regimes a priori as in Carvalho et al. (2021), i.e., $P(S_t = j) = 1$, for a given sub-period of time. We discuss these possibilities in Section 4, where we compare the results of three cases in increasing order of structural information: (1) time-varying classification sequence that are independent, (2) time-varying classification sequence that is dependent and follow a Markov structure, and (3) time-invariant classification sequence that follow exogenous regimes for the Pre- and Post-Volcker periods as in Carvalho et al. (2021).

Now we can estimate the mixture regression model using the Gibbs sampler, as specified in Algorithm 1 on the model parameters and here extended with the inclusion of the mixture weights.

Algorithm 1 Gibbs Sampler for the Mixture Regression Model with known k

1: Select the number of components, k, and initialize component weights, $p_j^{(0)} = \frac{1}{k}$, parameter values, $\boldsymbol{\theta}_j^{(0)} = (\boldsymbol{\beta}_j^{(0)\prime}, \sigma_j^{2(0)})' = (\mathbf{a}', b)'$, $b > 0$, $j = 1, \ldots, k$, and component indicators, $S_t^{(0)} \sim M(1, \mathbf{p}^{(0)})$, $t = 1, \ldots, T$.

2: **for** $d = 1$ to D **do**

- a) Given component indicators, $\mathbf{S}^{(d-1)}$, classify the observations $\mathbf{r}_j(\mathbf{S})$, $\mathbf{x}_{tj}(\mathbf{S})$, $T_j(\mathbf{S})$, and sample parameter values $\boldsymbol{\theta}_j^{(d)} = (\boldsymbol{\beta}_j^{(d')}, \sigma_j^{2(d)})'$, $j = 1, \ldots, k$ (independent multivariate Normal and inverse-Gamma)
- b) Given the parameter values, $\boldsymbol{\theta}_j^{(d)}$, and component weights $\mathbf{p}^{(d-1)}$, sample the component indicators $\mathbf{S}^{(d)} = (S_1^{(d)}, \ldots, S_T^{(d)})'$ (independent Multinomial)
- c) Given the parameter values, $\boldsymbol{\theta}_j^{(d)}$, and component indicators $\mathbf{S}^{(d)}$, sample component weights $\mathbf{p}^{(d)} = (p_1^{(d)}, \ldots, p_k^{(d)})'$ (independent Dirichlet)

3: **end for**

2.4 Estimation of the mixture model with unknown number of components

We next consider the mixture regression model with an unknown number of components k which constitutes our baseline model. Here, the main challenge is to estimate this number reliably and efficiently. This has been a nontrivial issue in the literature on this topic.[5] We are using the Sparse Finite Mixture Markov-Chain Monte Carlo (SFM-

[5] Several alternative methods exist like the Reversible Jump MCMC (Green, 1995; Richardson and Green, 1997), which is computationally complex and suffers from convergence problems in several cases, and the Dirichlet Process mixture (Miller and Harrison 2013), which suffers from unreliability. Koop and

MCMC) algorithm due to Malsiner-Walli et al. (2016). An important advantage of this MCMC method is its simplicity and ease of implementation. The algorithm can be viewed as a two-step process. In step 1, the researcher starts by deliberately overfitting the mixture model by specifying the number of components larger than the number of components expected to describe the data. In step 2, a regularization prior is used which, in combination with likelihood and data, shrinks the number of mixture components towards a credible number with substantial posterior probability. This method of selecting the number of components is also appealing because it overcomes well-known problems associated with using marginal likelihoods for component selection (Frühwirth-Schnatter, 2006), and it avoids the computational burden of having to estimate infinite mixture models using non-parametric approaches (Frühwirth-Schnatter and Malsiner-Walli, 2019).

To learn from the information set (data, likelihood and prior) how much sparsity is needed in the regularization prior, we treat the concentration parameter in the Dirichlet distribution α as an unknown parameter to be estimated. To that end, we specify a Gamma hyperprior of the form

$$\alpha \sim G(\alpha_0, \alpha_0 k), \tag{13}$$

where α_0 is to be selected by the researcher and k is the number of mixture components. This prior is desirable for two reasons. First, it is informative on near-empty components, i.e. $Var(\alpha) = \frac{1}{\alpha_0 K^2}$, while being uninformative across components, i.e, $\mathbb{E}(\alpha) = \frac{1}{K}$. Second, it ensures that the Dirichlet prior approximates a Dirichlet process prior with concentration parameter α as k becomes large (Ishwaran et al., 2001).[6]

The trade-off in adopting this approach is that the conditional posterior distribution implied by (8) and (13) is given by

$$f(\alpha^{(d)}|\mathbf{p}^{(d)}) \propto f(\alpha^{(d-1)}) \frac{\Gamma(\alpha^{(d-1)}k)}{\Gamma(\alpha^{(d-1)})^k} \left(\prod_{j=1}^{J} p_j^{(d)} \right)^{\alpha^{(d-1)}-1}, \tag{14}$$

where $\Gamma(\cdot)$ is the Gamma function.[7] Sampling from (14) therefore requires an additional Metropolis-step within the Gibbs sampler outlined earlier, as specified below in Algorithm 2. This Algorithm makes it clear that the main difference in estimating the mixture regression model with unknown number of components is the estimation of the concentration parameter.

Potter (2007) propose an alternative based on change-point modelling, allowing the number of change points to be unknown.

[6]The Dirichlet process prior is a commonly used prior for infinite mixture models.

[7]The Gamma function is defined as $\Gamma(x) = \int_0^\infty e^{-z} z^{x-1} dz$ where $x > 0$.

Algorithm 2 Metropolis-within-Gibbs Sampler for the Mixture Regression Model with unknown k

1: Initialize the concentration parameter, $\alpha^0 = c > 0$, and the number of components, component weights, parameter values, and component indicators as in Algorithm 1.

2: **for** $d = 1$ to D **do**

1. Sample:
   a) Parameter values, $\boldsymbol{\theta}_j^{(d)} = (\boldsymbol{\beta}_j^{(d)}, \sigma_j^{2(d)})'$, as in Algorithm 1.
   b) Component indicators, $\mathbf{S}^{(d)}$, as in Algorithm 1.
   c) Component weights, $\mathbf{p}^{(d)}$, as in Algorithm 1.
2. Given the parameter values, $\boldsymbol{\theta}_j^{(d)}$, component indicators $\mathbf{S}^{(d)}$, and weights $\mathbf{p}^{(d)}$, sample the concentration parameter $\alpha^{(d)}$ (Random Walk Metropolis step).

3: **end for**

Algorithm 2 provides estimates of the model parameters $\boldsymbol{\theta}$, $\mathbf{p}$, α, and $\mathbf{S}$, with k_0 components such that $k_0 > k^{\text{exp}}$, where the latter is the expected value of k to describe the data. The posterior distribution of the number of non-empty components, $Pr(k_0 = h|\mathbf{r})$, $h = 1, \ldots, k_0$, can be directly obtained from the MCMC output. To that end, note that the number of non-empty components in a given draw is given by

$$k^{(d)} = k_0 - \sum_{j=1}^{k_0} 1(T_j(\mathbf{S})^{(d)} = 0) \tag{15}$$

where $1(\cdot)$ is the indicator function. The relative frequency of each value across all draws, $\frac{1}{D}\sum_{d=1}^{D} k^{(d)}$, therefore returns the estimated posterior probability $Pr(k_0 = h|\mathbf{r})$, $h = 1, \ldots, k_0$. To estimate the credible number, k, of mixture components with substantial probability, we use the posterior mode, $\hat{k}$, as a point estimate from this distribution.[8] After computing the posterior mode, we consider only the sub-sequent MCMC draws such that the number of non-empty components is exactly equal to $\hat{k}$. This step is important because the main objective of this paper is to conduct component-specific parameter inference on the resulting Taylor rule coefficients. If we were instead interested in using the mixture model to fit an unknown distribution, then the number and exact nature of the underlying distributions is less relevant, and all draws from the above algorithm can be retained.[9]

The next step is to ensure that each of the draws within this sub-sequence is in the correct order. To illustrate the problem, suppose that $\hat{k} = 2$ and the first retained draw,

[8]The posterior mode is the natural choice for two reasons. First, it is optimal under the 0/1 loss function. Second, provided that $k_0 > k^{\text{exp}}$ it is invariant to the initial choice of k_0. We emphasize that this is not the case for the posterior mean or median which would be effected by the initial choice of k_0 due to the possibly large right tail when selecting large values of k_0.

[9]As a recent example Cross et al. (2024) introduce a mixture of shifted-Poisson distributions for modal inference in discrete data distributions. They find that inflation expectations in the Michigan survey exhibit substantial heterogeneity, which is indicative of de-anchored expectations.

$d = 1$, has a classification sequence $(1, 2)$, while the second retained draw, $d = 2$, has a classification sequence $(2, 1)$. This is known as *label switching*, and is overcome by clustering the retained draws into $\hat{k}$ clusters using K-centroids cluster analysis (Leisch, 2006), where the distance between a point and a cluster centroid is determined by a distance metric. Here we use the Mahalanobis distance which is widely used in cluster analysis and classification techniques.

The final step is to ensure that the classification sequence from the previous step results in a permutation of $(1, \ldots, \hat{k})$. Since classifications involving overlapping clusters are indicative of overfitting, this step is akin to having a parsimonious set of well-identified components (Frühwirth-Schnatter, 2011). In practice, this is done by simply checking whether or not the clustered draws is a permutation of $(1, \ldots, \hat{k})$. If it is, then we retain it. If it is not, then we discard it. For instance, in the case that $\hat{k} = 2$ we require that the classification sequence is either $(1, 2)$, or $(2, 1)$. Draws that are clustered into a single component, i.e., $(1, 1)$ or $(2, 2)$ are discarded. The remaining identified draws are then resorted and used for inference.

The different steps in the algorithms are summarized in Figure 1. The Gibbs sampler of Algorithm 1 is shown in steps 1a, 1b and 2 (excluding the role of the parameter α). The Metropolis-Hastings step is shown as the addition for Algorithm 2 in step 1c. Given this output one generates a complete vector of component indicators, $\mathbf{S} = (S_1, ..., S_T)'$, in one step from a Multinomial distribution. As discussed above, the reduction procedure going from a selection of components from $k_0 > k^{\exp}$ to a credible value of k makes direct use of the MCMC output of Algorithm 2.

It is important to emphasize that this estimation procedure is particularly suitable for our case of a parsimonious model structure due to the efficiency and flexibility of the algorithm. To the best of our knowledge, this is the first application of the sparse finite mixture approach of Malsiner-Walli et al. (2016) to Taylor rules, and the broader class of dynamic regressions more generally.

We end this section with a remark. Given that with substantial posterior probability we end up in our empirical analysis with two components or regimes, the Dirichlet prior on the concentration parameter α reduces to a Beta prior, and the Multinomial conditional posterior distribution of the component indicator S_t reduces to a Binomial distribution. However, in practice, one starts the estimation procedure with a much larger number of components or regimes than two.

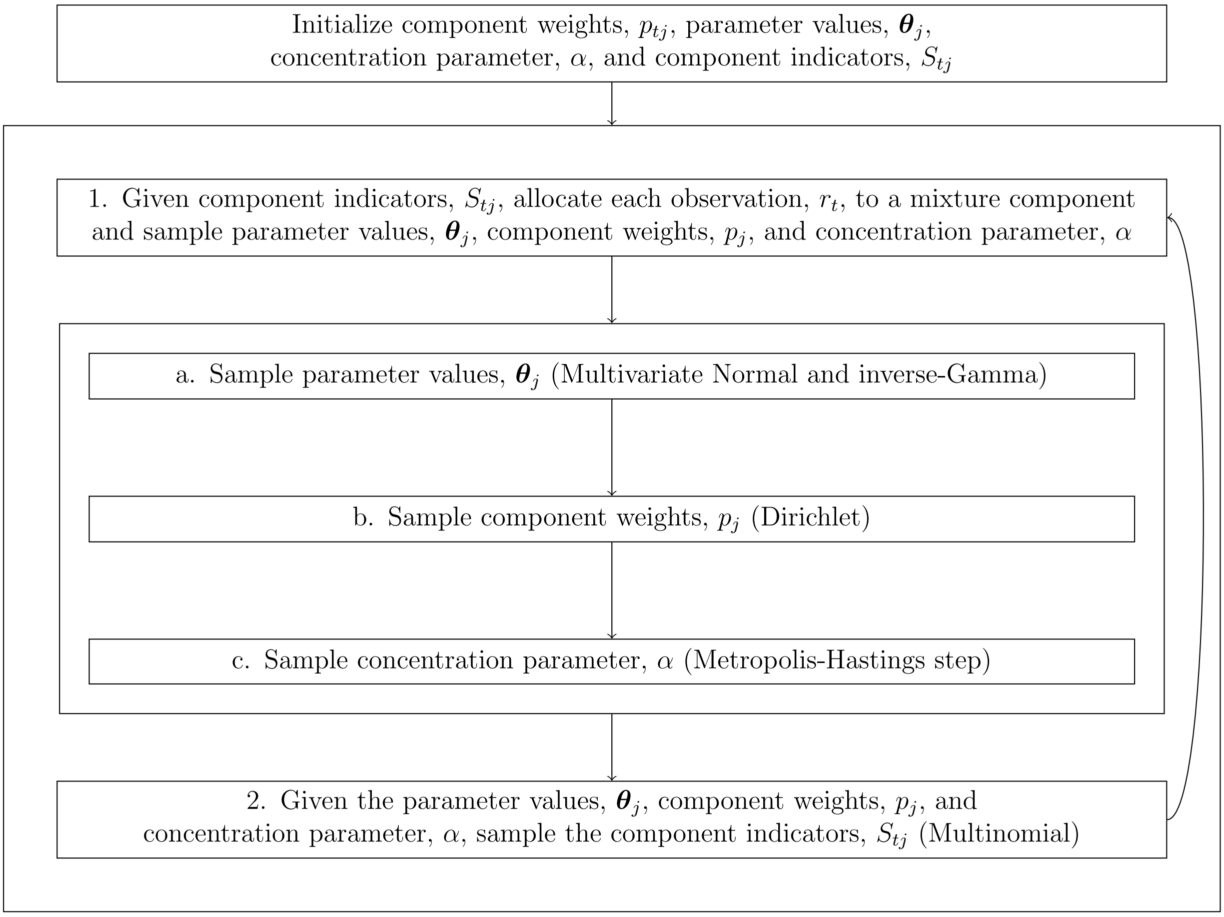


Figure 1: Summary of the algorithm for the Sparse Finite Mixture Markov Chain Monte Carlo (SFM-MCMC) sampling procedure

3 Results

We present estimation results on regime classification and on finite sample posterior estimates of the policy rule parameters. In addition, we discuss the fit of the model in comparison with alternative specifications.

3.1 Regime classification and inference on policy parameters

We estimate our baseline mixture model with unknown number of regimes for the period 1960:Q1-2021:Q1. Since we lose one year of observations due to data transformations, the period used for analysis is therefore 1961:Q1-2021:Q1.

Our first result is that the mixture model endogenously selects a two component density. The time series of the estimated component weights or probabilities are plotted in Figure 2. We emphasize that these probabilistic weights can be associated with a regime classification as done in business cycle analysis. Note the prevalence of the probabilistic weights denoted by S1 which constitutes a "normal time" regime. On the contrary, the probability of being in regime S2 episodically increases substantially, in particular in the

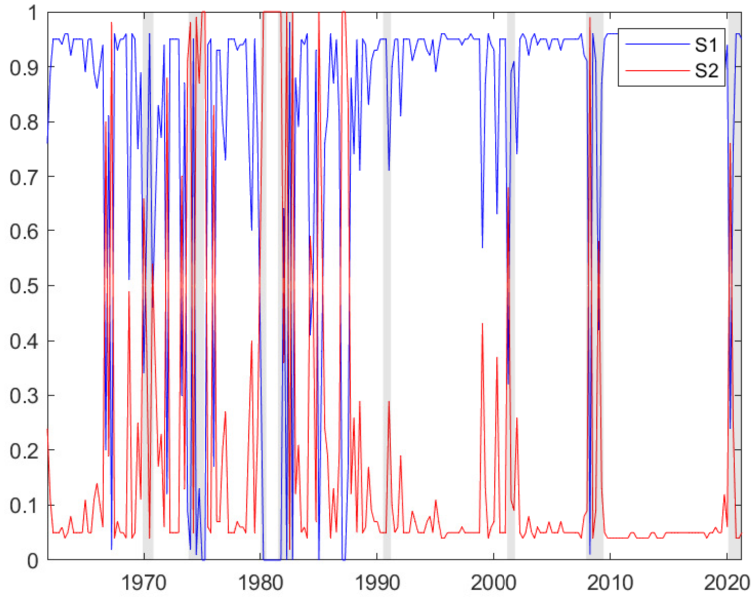


Figure 2: Time series of the estimated component (regime) indicators. The shaded areas show NBER recession periods.

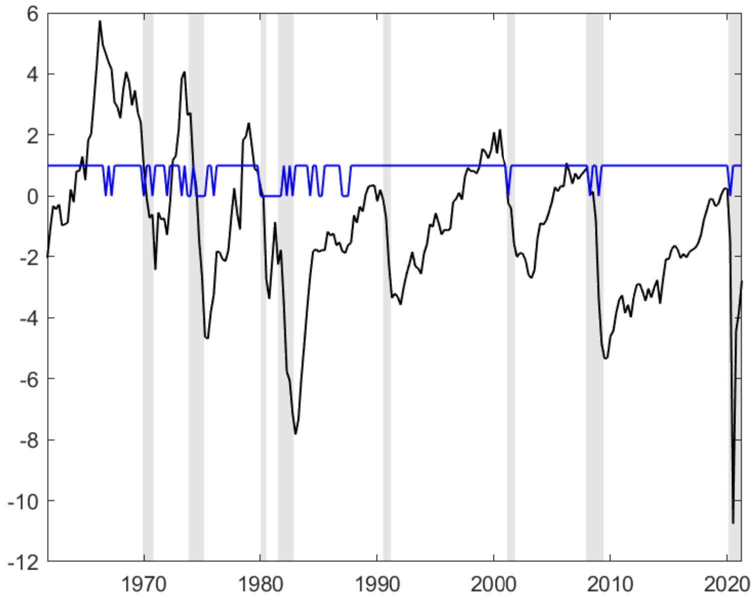


Figure 3: Regime indicator, defined as the average probability of being in regime S1 is larger than 50 percent (blue line) and the CBO output gap (black line). The shaded areas show NBER recession periods.

first part of the data sample. In fact, given our time series, regime S2 can be associated with recessions. This is more visible in Figure 3 where we plot the CBO output gap together with a regime indicator associated with periods when the probability of being in regime S1 is larger than 50 percent. The regime indicator captures all recessions present in the data (marked with grey bars) and a few additional episodes. Since most recessions materialize in the first part of the data sample, regime S2 is substantially more prevalent during that period. Clearly, the mixture model does not select two adjacent regimes, as often assumed in specifications with exogenous partitioning.

The posterior moments and densities of the structural parameters from the mixture model are shown in Table 1 and Figure 4. One major result is that the interest rate smoothing coefficient, ρ, is clearly and credibly different across the two regimes. According to our estimates, the reaction function exhibits a high degree of gradualism (with posterior mean equal to 0.92) in normal times and a much lower degree of gradualism (with posterior mean equal to 0.72) in the recessionary regime. Put simply, the Fed seems to respond much more gradually to the state of the economy in normal times.[10] This asymmetry is important from an economic point of view but also credible from a statistical point. The posterior mean of ρ in S1 is outside the 68% and 90% credible region of the posterior of ρ in S2. We further show in the upper panel of Figure 4 how different the two posterior densities are for the smoothing parameter. Not surprisingly, estimates are more precise in regime S1 which contains a much larger number of observations.

Table 1: Posterior moments for structural parameters in the mixture model

	Posterior Mean	68% CI	90% CI
	Mixture $S = 1$		
β	1.87	(1.48,2.25)	(1.25,2.62)
γ	1.04	(0.76,1.32)	(0.62,1.60)
ρ	0.92	(0.90,0.94)	(0.88,0.95)
	Mixture $S = 2$		
β	1.94	(1.32,2.59)	(0.97,3.32)
γ	1.44	(0.56,2.37)	(0.27,3.59)
ρ	0.72	(0.60,0.85)	(0.50,0.90)

Notes: The table shows the posterior moments and credible intervals (68% and 90%) for the structural parameters of each components in the mixture model. S1 denotes the "normal time" regime and S2 the "recession" regime. β is the inflation coefficient, γ the output gap coefficient and ρ the interest smoothing coefficient.

Our second result concerns the response coefficients to inflation β and the output gap γ. The response to inflation is essentially the same across regimes while the response to the output gap is substantially larger in the recessionary regime (a posterior median of 1.44 against a value of 1.04 in normal times). Thus, the Fed seems to respond more

[10] One may wonder whether the asymmetric degree of smoothing is capturing a symmetric response to a very asymmetric variable like the unemployment rate which is not included in our baseline regression. In order to investigate this concern, we replace the output gap with the unemployment rate in the estimated mixture Taylor rule in online Appendix A.2. Our results are not only confirmed but even reinforced with an estimated degree of smoothing of around 0.5 in the recessionary regime in this experiment.

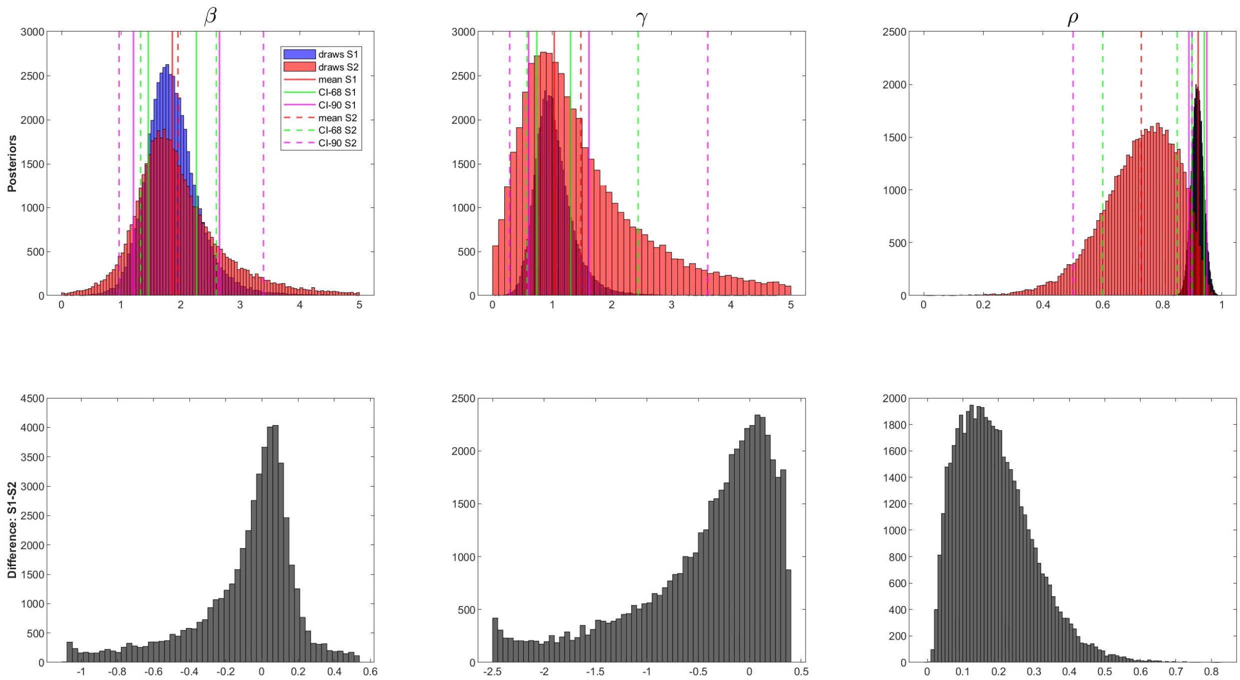


Figure 4: Posterior densities of structural parameters (top row) and difference between posterior densities of structural parameters (bottom row) in the mixture model. For visualization purposes the areas under the densities in the top panel are not scaled to be the same.

to real economic activity when economic conditions deteriorate sharply as in recessions. Intriguingly, our methodology suggest that the most stable parameter is the coefficient on inflation which is precisely the coefficient whose magnitude changes the most in models with exogenous partitioning.

Are the mixture regimes different? To investigate this question, we look at the difference between the estimated posterior densities of the structural parameters in the mixture model for regimes S1 and S2. The bottom panel in Figure 4 shows that generated draws in the two regimes are quite different. For all three parameters (β, γ, ρ) the evidence indicates that generated draws in regime S2 occur over a wider range than for S1, which confirms the results about the credibility intervals reported in Table 1. For the coefficient ρ it is seen that the posterior density is very skew in regime S2 and the generated draws occur much more in a wider stationary region than for regime S1. For the output coefficient γ the posterior density is also strongly asymmetric. For the inflation coefficient β the posterior densities in both regimes are more symmetric but in regime S2 more draws are generated in both the left and the right tail. These results stem from a simulation-based Bayesian inference procedure which holds for the exact finite sample data.

We believe that our evidence of asymmetric gradualism is sensible. In an influential speech, Panetta (2022), at that time Member of the Executive Board of the ECB, states that gradualism was clearly appropriate in the euro area in 2022 because of the strength of recent shocks that was generating extreme uncertainty about the outlook for economic activity and because of the unprecedented nature of the shocks. However, in the same speech, Panetta (2022) clearly states that *"...a gradual approach is not appropriate in all circumstances. For example, when faced with deflationary shocks that risk rooting interest rates at the lower bound, it pays to act more decisively"*. Therefore, it seems that gradualism is desirable mainly during tightening cycles. Note that financial stability considerations are often invoked (see Bernanke (2004)) as a second justification for gradualism in addition to the robustness argument used by Panetta (2022). However, the need to proceed cautiously in order to avoid financial market disruptions is also asymmetric in nature: it is invoked when central banks increase interest rates but not when central banks quickly lower interest rates in response to a deteriorating outlook. Finally, the logic of asymmetric gradualism was stressed also during the FOMC video conference meeting at the Fed on October 15, 2010 when it was stated *"In their discussion of the relative merits of smaller and more frequent adjustments versus larger and less frequent adjustments . . . , [FOMC] participants generally agreed that large adjustments had been appropriate when economic activity was declining sharply in response to the financial crisis. In current circumstances, however, most saw advantages to a more incremental approach that would involve smaller changes . . . calibrated to incoming data."*.

3.2 Shocks and model fit

We evaluate the implied monetary policy shocks from the structural estimates and the in-sample fit of our mixture model. The residuals represent non-systematic deviations from Taylor Rule fundamentals and can consequently be interpreted as monetary policy shocks.

As shown in Figure 5, monetary shocks are small with the exception of the end of the 1970s when we identify a series of large and volatile shocks. This is evidence that monetary policy was less systematic in that specific period and more generally in the first part of the sample (see also Bernanke et al. (1997)). This is a common result in the literature showing that fluctuations in the interest rate reflect mostly the endogenous response of the central bank to the state of the economy rather than unsystematic interventions.

Next, we compare models using the Bayesian R^2 (Gelman et al., 2019), which is the Bayesian analogue of the widely used coefficient of determination (also known as R^2). Gelman et al. (2019) define this R^2 as:

$$R^2_{\text{Bayes}} = \frac{\text{var}_{\text{fit}}}{\text{var}_{\text{fit}} + \text{var}_{\text{res}}}. \tag{16}$$

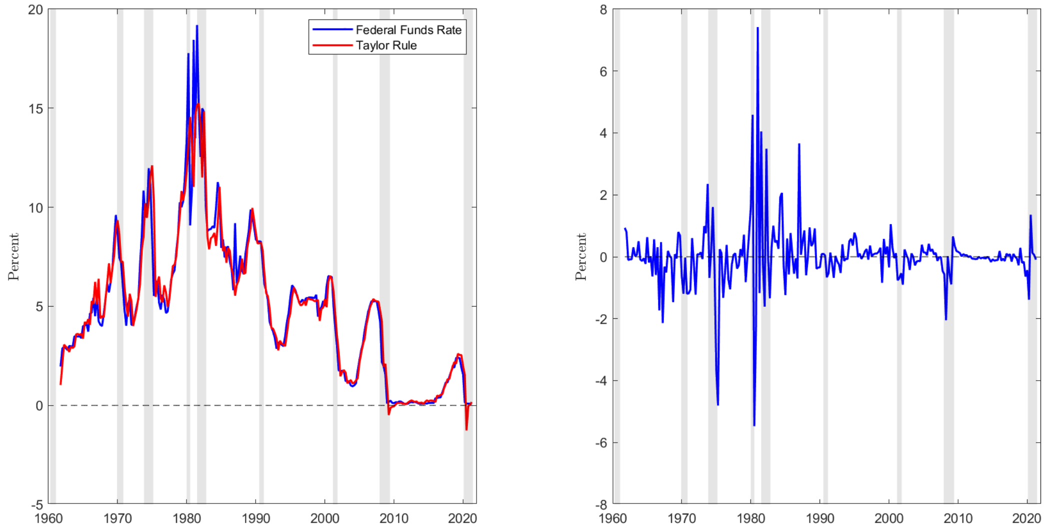


Figure 5: Fitted Mixture Taylor Rule and actual Federal funds rate (left) and monetary policy shocks from the mixture regression (right)

The procedure to evaluate R^2_{Bayes} is as follows: one generates a random draw from the posterior of the parameters and evaluates one value of R^2_{Bayes} using equation (16) (similar to the Frequentist approach where one obtains one R^2 using, for instance, OLS estimators of the parameters). The generation of parameter draws and the evaluation of R^2_{Bayes} is repeated many times. Formally, R^2_{Bayes} is integrated/averaged over the parameter space, for details see Gelman et al. (2019). Note that R^2_{Bayes} is a regular function of the parameters, as a consequence the simulated finite sample posterior density of R^2_{Bayes} follows directly from the simulated values of the posterior density of the model parameters. This is in contrast to the Frequentist approach in this respect.

Table 2: Model Comparison Metrics

	Mean	90% credible interval
Two Regimes Mixture	90.73	(89.46,91.60)
Linear Regression	89.81	(89.06,90.39)

Notes: The table shows the posterior mean and 90% credible interval for Bayesian R^2 for estimated Taylor rules based on the mixture regression with two regimes, and the linear regression over the full sample.

Table 2 shows that the mixture model provides a good fit with a tight finite sample posterior density of R^2_{Bayes} and a mean which is outside the 90% credible interval of the density for the full-sample linear regression model. The mean fit of the linear regression model is also slightly lower. This suggests that, on average, the mixture model provides a superior in-sample fit.

To further investigate the significance of this improvement, we plot the posterior distributions of each models R^2_{Bayes} together in Figure 6. The figure highlights that the empirical support of the posterior distribution for the mixture model tends to be larger than the linear regression. To determine the credibility of this improvement, we make use of the two-sample Kolmogorov–Smirnov function for pair-wise equality of the Bayesian R-squared distribution from the mixture model with the linear regression. In this test the null hypothesis is that the MCMC draws of the Bayesian R-squared from the two models are drawn from the same distribution, and the alternative hypothesis is that they are not drawn from the same distribution. The test provides strong evidence that the distributions are not equal. We therefore conclude that the mixture model provides a superior in-sample fit relative to the linear regression model.

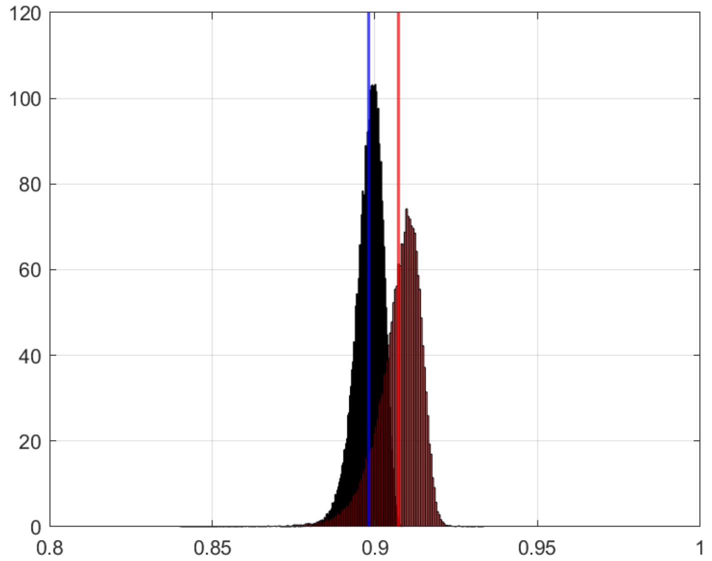


Figure 6: Bayesian R-squared for the Mixture model (red) and linear regression model (Blue). Solid lines are the respective posterior means.

4 Increasing structural information on mixture weights

So far, we considered the case of independent time-varying classification of the data. However, various scholars have examined changes in monetary policy regimes by specifying the Taylor Rule as a Markov-Switching regression model, or by exogenous partitioning based on the Fed chairperson. In this section we show that both of these models are in the general class of mixture regression models. However, the unrestricted mixture with independent time-varying classification provides a superior fit of the data.[11]

[11]See Davig and Leeper (2007) for a seminal contribution in the use of Markov-Switching Taylor rules. Sims and Zha (2006) also provide a related multivariate analysis with a Markov-Switching vector autoregression (VAR) model.

In a Markov-Switching regression model the vector of discrete latent indicators, $\mathbf{S} = (S_1, \ldots, S_T)'$, is defined such that each S_t is a discrete latent process, with finite state space $\{1, \ldots, k\}$, whose dynamics are described by an irreducible and aperiodic Markov chain. Specifically, the dynamics of S_t are governed by the $k \times k$ transition matrix P, in which each element p_{ij} represents a transition probability from state i in period $t-1$ to state j in period t, defined as:

$$Pr(S_t = j | S_{t-1} = i) = p_{ij}, \tag{17}$$

where $i, j \in \{1, \ldots, k\}$.

The core difference between our mixture and Markov switching mixture models is seen by comparing the component indicator specifications in (4) and (17). We here highlight two important differences between these different specifications. First, the Markov process in (17) introduces autocorrelation into the regime indicator. In contrast, (4) implies that the component indicator in the mixture model has no built-in a priori memory. Second, the specification of a latent Markov process for the latent component indicator means that the states S_t must be estimated using forward and backward filtering recursions. These recursions are well known and are often referred to as the 'Hamilton filter' after the seminal paper by Hamilton (1989).[12] This additional step greatly increases the computational complexity of the model. In contrast, the dynamic independence of the component indicator in (10) implies that S_t in the mixture model is sampled directly from a multinomial distribution.

The equation parameters in the Markov-Switching model can be estimated using a similar Bayesian method as in our mixture regression model. In fact, the priors for elements of $\boldsymbol{\theta}_j$ are set in exactly the same way as in our mixture model, and result in exactly the same (conditional) posterior distributions across regimes. We also assume that the number of regimes k is known, and specify a Dirichlet distribution for the transition probabilities such that $(p_{i,1}, \ldots, p_{i,k})$, $i = 1, \ldots, k$, with symmetric concentration parameter α. From the viewpoint of prior specification, the key difference between the mixture model and Markov-Switching model is that specifying the latent indicator S_t as a Markov process introduces the transition probability matrix P. Since the Federal Funds Rate is known to evolve with a high degree of gradualism, we specify an informative prior on the regime transition probability matrix such that $P = 1_k + \zeta I_k$, in which 1_k is a $k \times k$ matrix with all entries equal to one, I_k is the identity matrix of size k, and $\zeta > 0$ is a parameter that controls the degree of regime persistence. To see this note that the expected probability of staying in regime j is given by $\mathbb{E}(p_{jj}) = \frac{1+\zeta}{\zeta+k}$. Thus, a higher value of ζ indicates a higher persistence and vice-versa. Here we set $k = 2$ and $\zeta = 18$ which

[12] A comprehensive textbook treatment of these forward filtering and backward smoothing procedures can be found in Frühwirth-Schnatter (2006).

translates to a 95% prior probability of staying in the same regime.[13]

Results from the Markov-Switching model are presented in Table 3. While the magnitude of the estimated coefficients is similar to our baseline model, the characterization of the regimes is quite different. As shown in Figure 7, the partitioning of the regimes is less interpretable in the Markov-Switching model with the less frequent regime being never selected after the late 1980s. In contrast, the model assigns a high probability to the less frequent regime for longer periods in the first part of the data sample.

Table 3: Posterior moments of estimated structural parameters a two-regime Markov-Switching model

	Posterior Mean	68% CI	90% CI
	Mixture $S = 1$		
β	1.36	(1.03,1.69)	(0.85,2.04)
γ	0.95	(0.74,1.17)	(0.60,1.36)
ρ	0.90	(0.87,0.92)	(0.86,0.94)
	Mixture $S = 2$		
β	1.68	(1.13,2.21)	(0.78,2.89)
γ	1.18	(0.35,2.01)	(0.14,3.23)
ρ	0.72	(0.59,0.85)	(0.48,0.90)

Notes: The table shows the posterior moments and credible intervals (68% and 90%) for the structural parameters in a two-regime markov-switching model. S1 denotes the "normal time" regime and S2 the "recession" regime. β is the inflation coefficient, γ the output gap coefficient and ρ the interest smoothing coefficient.

This highlights an important difference between our baseline mixture model and the Markov-Switching model. In the mixture model the state variable is independent with respect to time while it follows a Markovian process in the Markov-Switching model. This means that the Markov-Switching model imposes a stronger structure and the estimation of the states requires filtering and smoothing steps, which also makes the states more persistent. Since recessions are less likely to occur than expansions (particularly in the latter part of our data sample) and also less persistent, there is a risk that the Markov-Switching model may underestimate the frequency of recessions. On the contrary, the mixture model imposes very little structure and sampling is done directly from a Multinomial distribution, making the model more parsimonious and less prone to specification errors.[14]

[13] To see this note that $\mathbb{E}(p_{jj}) = \frac{1+\zeta}{\zeta+k} = (1+18)/(18+2) = 0.95$.

[14] The coding of a mixture model is also much simpler, and the computational speed faster. This is primarily because the component indicator in the mixture model can be sampled from a multinomial distribution in one line of code. In contrast, for the case of Markov-Switching, the latent states are sampled using computationally costly forward and backward smoothing recursions that require more computer time.

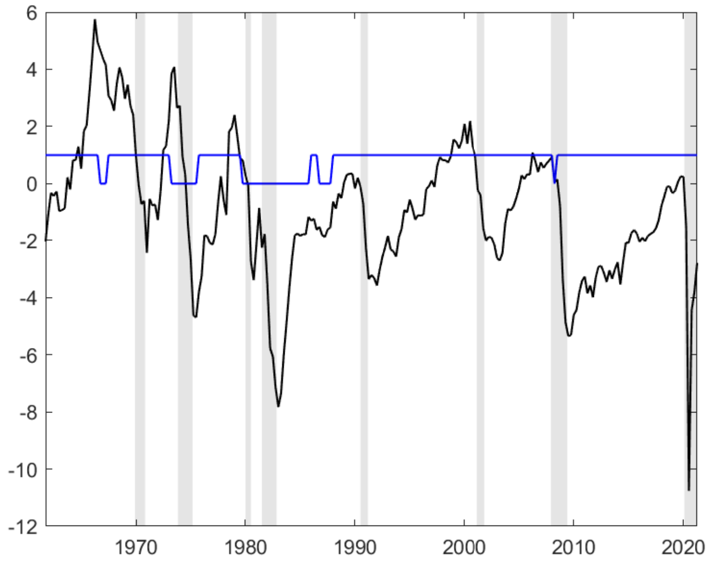


Figure 7: Regime indicator from a two-regime markov-switching model, defined as the probability of being in regime S1 is larger than 50 percent (blue line) and the CBO output gap (black line). The shaded areas show NBER recession periods.

As a third case of information on the component indicator, we consider a time-invariant specification that uses exogenous partitioning before and after the Volcker period as in Clarida et al. (2000). Following Carvalho et al. (2021), we specify the break date such that the pre-Volcker period ends in 1979:Q2 and the post-Volcker period starts from 1979:Q3. If we define the pre-Volcker period as regime 1, then this corresponds to defining a component indicator such that:

$$\begin{aligned} Pr(S_t = 1) &= 1, \quad \text{if } t \leq 1979Q2 \text{ and } 0 \text{ otherwise}, \\ Pr(S_t = 2) &= 1, \quad \text{if } t > 1979Q2 \text{ and } 0 \text{ otherwise}. \end{aligned} \tag{18}$$

For completeness, this indicator is illustrated in Figure 8, and the estimated Taylor rule parameters are presented in Table 4. In line with the real-time analysis of Carvalho et al. (2021), we find that the Taylor rule in the post-Volcker regime places a much larger emphasis on inflation, along with slightly lower gradualism.[15]

To compare the in-sample fit of the models, we use the Bayesian R-squared. Results from the three cases are plotted together in Figure 9, and summarized in Table 5. The posterior mean and 90% credible interval of the Bayesian R-squared for the Markov switching and exogenous break models are 90.19 $(88.83, 91.19)$, and 89.63 $(88.87, 90.23)$ respectively. Comparing this to results in Table 2 suggests that, on average, the mixture model provides a superior in-sample fit over these two model, however the Markov-

[15] We note that the results here differ slightly from those in Appendix A.1 since these results are obtained with the most recent data vintage as opposed to real-time data.

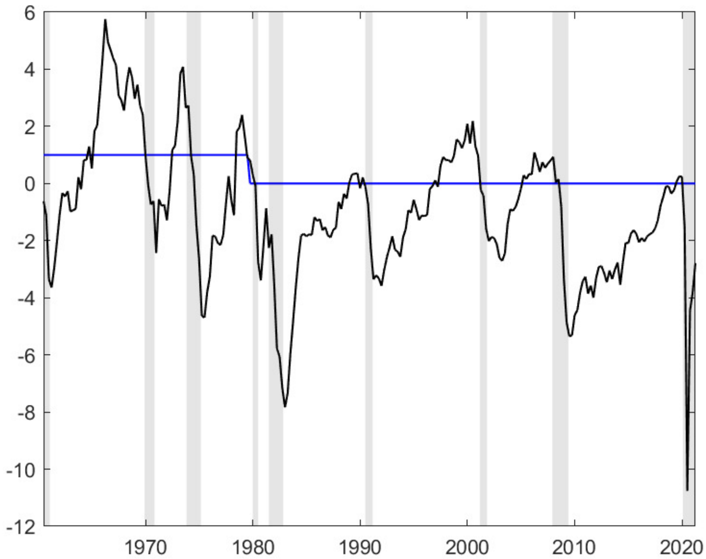


Figure 8: Regime indicator for the regression with an exogenous break in 1979:Q2 (blue line) and the CBO output gap (black line). The shaded areas show NBER recession periods.

Table 4: Posterior mean for Bayesian estimation of the Taylor rule structural parameters in various periods

	Pre-Volcker 1960Q1-1979Q2	Post-Volcker 1979Q3-2021Q1
β	0.84	2.17
γ	0.83	0.84
ρ	0.75	0.69

Notes: The table shows the posterior moments and credible intervals (68% and 90%) for the structural parameters when imposing a time-invariant specification that uses exogenous partitioning before and after the Volcker period as in Clarida et al. (2000). β is the inflation coefficient, γ the output gap coefficient and ρ the interest smoothing coefficient.

Switching model still provides a superior in-sample fit over the two regression models. To determine the credibility of the mixture models improvements, we again use the two-sample Kolmogorov–Smirnov function. In both cases, we find strong evidence that the Bayesian R^2 distributions from the mixture model is not equal to that of the Markov switching or linear regression models. We therefore conclude that the mixture model provides a superior in-sample fit relative to both of these model specifications.

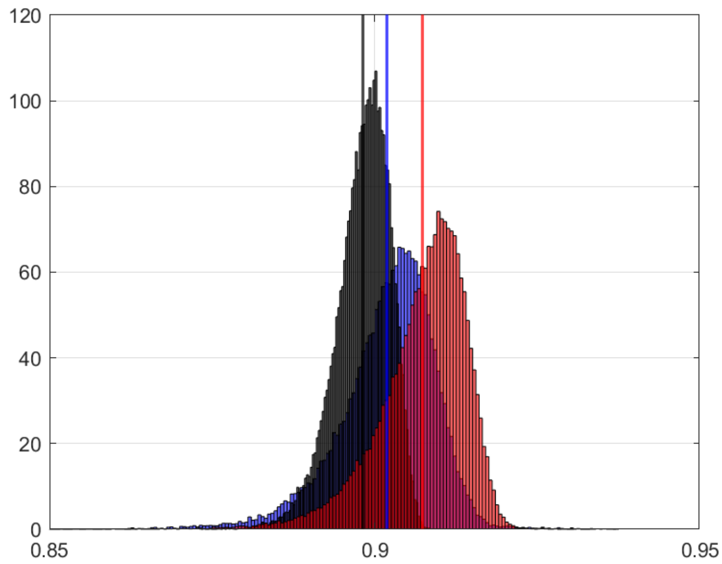


Figure 9: Bayesian R-squared for the Mixture model (red), Markov Switching model (Blue) and exogenous break model (Black). Solid lines are the respective posterior means.

Table 5: Model Comparison Metrics

	Mean	90% credible interval
Two Regimes Markov-Switching	90.19	(88.83,91.19)
Two Dogmatic Regimes	89.63	(88.87,90.23)

Notes: The table shows the posterior mean and 90% credible interval for Bayesian R^2 for estimated Taylor rules based on the mixture regression with two regimes, Markov switching regression with two regimes, and a regression with exogenous break in 1979Q2, respectively.

5 Our results in perspective

In this section, we compare our results with those in the literature emphasizing a shift over time in the policy response against inflation. In addition, we discuss the robustness of our results with respect to using real-time data, a sample excluding the zero lower bound period or a time-varying natural rate of interest.

5.1 Asymmetric smoothing or changes in the response to inflation?

Our main result is that the mixture model identifies two (and only two) regimes that feature a similar response to inflation but varying degrees of gradualism and responsiveness to real economic activity. At first sight, this result seems in contrast with the seminal paper by Clarida et al. (2000) and the following literature based on exogenous partitioning. In fact, these papers identify a large increase in the policy response to inflation during the post-Volcker period, whereas our model favors partitioning based on the phase of the business cycle.

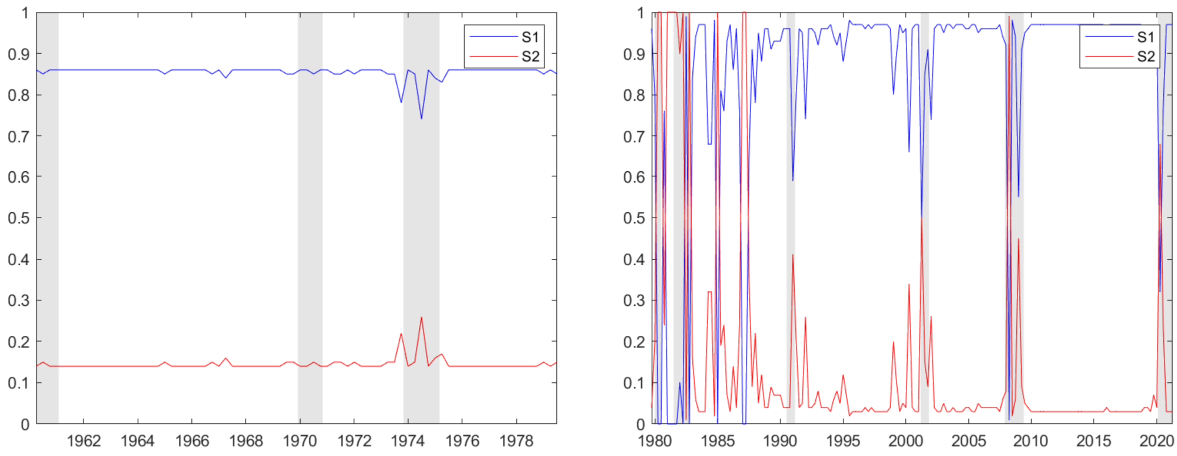
In order to reconcile the two set of results, we impose an exogenous partitioning in 1979:Q3 and estimate the mixture model on both the pre-Volcker and the post-Volcker periods. The results for the post-Volcker period largely confirm the findings of the baseline model based on the full data sample, as shown in Table 6 and Figure 10. The model relies on two regimes with the recessionary regime featuring a lower degree of gradualism. In addition, the response to inflation is estimated largely above one in both regimes. When estimated over the pre-Volcker period, the mixture model still selects two regimes. However, in this case both regimes feature a low response to inflation with a coefficient lower than one (see Table 6) in keeping with Clarida et al. (2000). Since the model is estimated over a short data sample (less than 20 years), the identification of the two regimes is poor, as shown in Figure 10. Nonetheless, this experiment is particularly insightful: it shows that our baseline model selects the asymmetry between expansions and recessions as more salient than the asymmetry between pre-Volcker and post-Volcker.

Table 6: Posterior moments for structural parameters in the mixture model over the "Pre-Volcker" and "Post-Volcker" period

	pre-Volcker			**post-Volcker**		
	Mean	68% CI	90% CI	Mean	68% CI	90% CI
		Mixture $S = 1$			Mixture $S = 1$	
β	0.87	(0.60,1.06)	(0.21,1.38)	2.90	(2.51,3.31)	(2.17,3.67)
γ	0.84	(0.50,1.09)	(0.22,1.64)	0.87	(0.65,1.08)	(0.51,1.28)
ρ	0.69	(0.61,0.81)	(0.28,0.86)	0.89	(0.86,0.91)	(0.84,0.93)
		Mixture $S = 2$			Mixture $S = 2$	
β	0.96	(0.13,1.77)	(0.03,3.02)	2.11	(1.83,2.36)	(1.66,2.63)
γ	0.99	(0.12,1.92)	(0.03,3.50)	0.62	(0.24,0.97)	(0.10,1.39)
ρ	0.50	(0.18,0.78)	(0.05,0.86)	0.24	(0.08,0.40)	(0.02,0.55)

Notes: The table shows the posterior moments and credible intervals (68% and 90%) for the structural parameters of each components in the mixture models estimated over the "Pre-Volcker" and "Post-Volcker" period, respectively. S1 denotes the "normal time" regime and S2 the "recession" regime. β is the inflation coefficient, γ the output gap coefficient and ρ the interest smoothing coefficient.

Intriguingly, if we force the mixture model to select three components (see Table 7), we recover a normal-time regime with high response to inflation and high degree of interest rate smoothing, a regime with high response to inflation and low smoothing which captures mainly economic downturns, and a third regime featuring a very low response to inflation and an intermediate degree of gradualism. This third regime is associated



(a) Pre-Volcker

(b) Post-Volcker

Figure 10: Time series of estimated components (regimes) probabilities (weights) for mixture regressions over the "Pre-Volcker" and "Post-Volcker" period, respectively. The shaded areas show NBER recession periods.

Table 7: Posterior distribution of estimated structural parameters in the Mixture model with three regimes.

	Posterior Mean	68% CI	90% CI
	Mixture $S = 1$		
β	2.90	(2.21,3.63)	(1.90,4.32)
γ	1.19	(0.78,1.60)	(0.58,2.07)
ρ	0.94	(0.92,0.96)	(0.90,0.97)
	Mixture $S = 2$		
β	2.60	(1.97,3.32)	(1.73,4.14)
γ	0.55	(0.14,0.91)	(0.05,1.43)
ρ	0.33	(0.08,0.59)	(0.02,0.71)
	Mixture $S = 3$		
β	0.45	(0.17,0.72)	(0.06,0.90)
γ	1.10	(0.71,1.46)	(0.54,1.90)
ρ	0.76	(0.71,0.81)	(0.67,0.84)

Notes: The table shows the posterior moments and credible intervals (68% and 90%) for the structural parameters of each components in a mixture model with three components (three regimes). S1 denotes the "normal time" regime and S2 and S3 are both associated with more recessionary periods. β is the inflation coefficient, γ the output gap coefficient and ρ the interest smoothing coefficient.

mainly with selected periods in the first part of the data sample (see Figure 11), in keeping with the narrative of Clarida et al. (2000) and the subsequent literature. One can see the recessionary regime in our baseline model as a combination of the second and

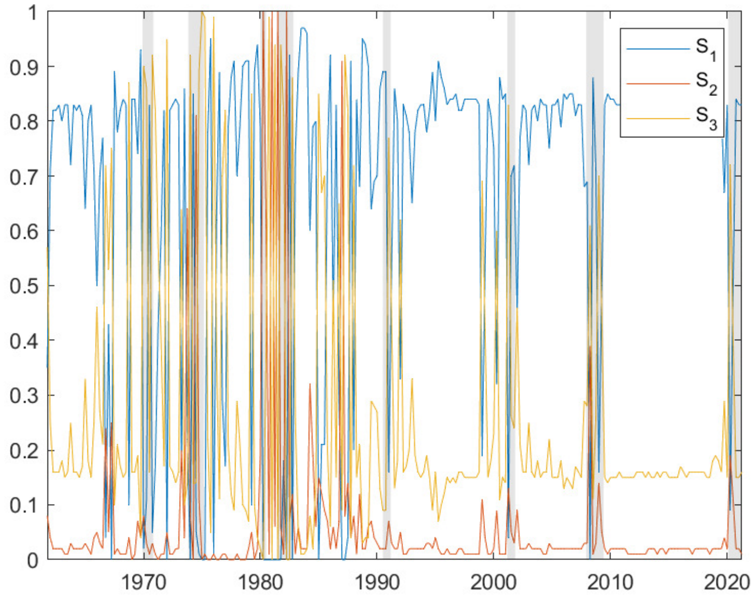


Figure 11: Time series of estimated components (regimes) probabilities (weights) in a three component mixture regression. The shaded areas show NBER recession periods.

third regimes in this extended model. This result confirms that the mixture model can identify different patterns in the response to inflation over time. Nonetheless, the main asymmetry is between expansions and contractions. Given that contractions are rare and shallow, such asymmetry can be captured only by a model with endogenous partitioning.

5.2 The role of real-time data

We re-estimate our model using the Greenbook/Tealbook forecasts prepared by the Federal Reserve staff before each meeting of the Federal Reserve Open Market Committee (FOMC) as real-time measures of expected inflation and the output gap. For both inflation and the output gap, we use the average of the nowcast and the one-quarter ahead forecast.[16] In this case, the data sample is 1975:Q1-2017:Q4 because of data availability. The results presented in Table 8 show a substantially larger difference in the degree of gradualism (measured by ρ) between normal times (0.92) and the recessionary regime (0.54) than in our baseline estimation in Section 3. Interestingly, the difference in the response to the output gap (γ) is also substantially larger in the recessionary regime (a posterior mean of 1.67 compared to 0.92 in normal times). The inflation coefficient β is somewhat larger in normal times (1.53 vs. 1.35) but the Taylor principle is easily satisfied in both regimes.

These results are particularly connected with Coibion and Gorodnichenko (2011), who

[16] In online Appendix A.3, we show that results are very similar when using either only nowcasts or only one-quarter ahead forecasts for inflation and the output gap.

Table 8: Posterior moments and credible intervals for structural parameters in the Mixture model with Greenbook forecasts over the data period 1975:Q1-2017:Q4.

	Posterior Mean	68% CI	90% CI
	Mixture $S = 1$		
β	1.53	(0.63,2.38)	(0.24,3.13)
γ	0.92	(0.30,1.50)	(0.11,2.14)
ρ	0.92	(0.90,0.94)	(0.88,0.95)
	Mixture $S = 2$		
β	1.35	(0.31,2.45)	(0.09,3.66)
γ	1.67	(0.43,3.02)	(0.13,4.11)
ρ	0.54	(0.35,0.73)	(0.19,0.82)

Notes: The table shows the posterior moments and credible intervals (68% and 90%) for the structural parameters of each components in the mixture model when using Greenbook/Tealbook data. We use the average of the nowcast and one-quarter ahead forecasts for inflation and output gap, respectively. For inflation we use data for the GDP deflator. S1 denotes the "normal time" regime and S2 the "recession" regime. β is the inflation coefficient, γ the output gap coefficient and ρ the interest smoothing coefficient.

estimate the Federal Reserve reaction function under exogenous partitioning. Notably, these authors include as regressors the Greenbook forecasts as real-time measures of expected inflation, output growth and the output gap. They find that the long-run response to inflation is higher in the second period but the difference is not statistically significant, thus in contrast with Clarida et al. (2000). Therefore, while the main asymmetry in models with exogenous partitioning is dampened when using real-time data, the asymmetries recovered by our model with endogenous partitioning are amplified. In addition, Coibion and Gorodnichenko (2011) find that the degree of interest rate smoothing has risen in the second part of their sample. This is consistent with our results once one recognizes that recessions are mostly concentrated in the pre-Volcker period.

All in all, we conclude that the asymmetry in gradualism and in the response to the output gap are reinforced when using real-time data while the asymmetry in the response to inflation in models with exogenous partitioning is substantially weaker when using real-time data .

5.3 Alternative data period and the natural rate of interest

Our data sample ends in 2021:Q2 and therefore include the period in which the policy rate was constrained by the zero-lower-bound. It is reasonable to assume that during that specific period, the policy rule did not accurately represent the policy stance. Note, however, that our model could in principle capture the zero-lower-bound period as an independent regime. In practice, this is not the case. In order to check that our results are not driven by this peculiar period, we re-estimate the model by stopping the data sample in 2007:Q4. All results are confirmed in this experiment as shown in Table 9.

In a second exercise, we take into account that the natural rate of interest has declined over our data sample. Rather than imposing a constant value for the natural rate of interest, we assume that the natural rate of interest is time-varying and is computed following the method proposed by Laubach and Williams (2003). Therefore, the term in the left hand side of the regression includes the difference between the federal funds policy rate and the Laubach-Williams estimate of the natural rate of interest. Results in Table 9 are based on estimates over the period 1961:Q1-2021:Q2. While a lower degree of gradualism in the recessionary regime is confirmed, the differences in the output gap coefficient across regimes are now magnified with respect to the baseline model (1.53 vs 0.67) and the response to inflation is estimated to be stronger in the recessionary regime.

Table 9: Posterior moments for structural parameters in the mixture model on a pre Great Recession sample and on a sample allowing for changing natural rate of interest.

	Pre Great Recession			**Natural rate of interest**		
	Mean	68% CI	90% CI	Mean	68% CI	90% CI
		Mixture $S = 1$			Mixture $S = 1$	
β	1.67	(1.17,2.15)	(0.87,2.71)	1.45	(1.08,1.81)	(0.85,2.17)
γ	1.18	(0.70,1.63)	(0.51,2.28)	0.67	(0.40,0.93)	(0.28,1.22)
ρ	0.91	(0.88,0.94)	(0.86,0.95)	0.91	(0.89,0.93)	(0.88,0.95)
		Mixture $S = 2$			Mixture $S = 2$	
β	1.86	(1.17,2.57)	(0.78,3.43)	1.81	(1.18,2.62)	(0.80,3.56)
γ	1.43	(0.42,2.52)	(0.16,3.72)	1.53	(0.51,2.64)	(0.21,3.86)
ρ	0.67	(0.51,0.83)	(0.35,0.89)	0.74	(0.62,0.87)	(0.51,0.91)

Notes: The table shows the posterior moments and credible intervals (68% and 90%) for the structural parameters of each components in the mixture models when the estimation sample ends in 2007:Q4 and when estimating on the full sample but using the Laubach-Williams estimate of the natural rate of interest, respectively. In the latter case, the left hand side of the regression includes the difference between the federal funds rate policy rate and the Laubach-Williams estimate of the natural rate of interest. S1 denotes the "normal time" regime and S2 the "recession" regime. β is the inflation coefficient, γ the output gap coefficient and ρ the interest smoothing coefficient.

6 Conclusion

In this paper, we propose a methodology based on a mixture model to endogenously partition interest rate data into different regimes. We estimated mixture weights and parameters of the component distributions using a relatively simple Markov Chain Monte

Carlo method. The methodology is applied to the estimation of basic Taylor-type monetary policy rules, contrasting with the commonly used approach of exogenous partitioning for estimating such policy rules (cf. Clarida et al. (2000)).

Our approach endogenously selects two regimes that can be associated with normal times and recessions. We obtain a substantially higher degree of monetary gradualism in normal times. In addition, evidence is obtained about a more aggressive response to real economic activity in the recessionary regime. The magnitude of the response to inflation does not exhibit clear variation across regimes. While we find some evidence of changes in the response to inflation in auxiliary experiments, our model endogenously chooses to put more weight on the asymmetry in gradualism based on the state of the business cycle. Put differently, the model sees the asymmetry between expansions and recessions as more salient than the asymmetry between pre-Volcker and post-Volcker. A modelling implication from our results is therefore that models that include the reaction of monetary authorities to various economic variables, such as commonly used DSGE models, should take into account that the degree of gradualism is state-dependent.

In terms of future research, it would be certainly interesting to apply the mixture model to a multivariate set-up. While clearly outside the scope of this paper, such an extension would make it possible to investigate all the questions studied so far in Markov-Switching Vector Autoregressions (cf. Sims and Zha (2006), Bianchi (2013) and Bianchi and Melosi (2017)) in the context of a model that can isolate regimes related to brief and abrupt episodes, as shown in our paper.

References

Aastveit, K. A. and Anundsen, A. K. (2022). Asymmetric effects of monetary policy in regional housing markets. *American Economic Journal: Macroeconomics*, 14(4):499–529.

Aastveit, K. A., Furlanetto, F., and Loria, F. (2023). Has the Fed responded to house and stock prices? a time-varying analysis. *Review of Economics and Statistics*, 105:1314–1324.

Alpanda, S., Granziera, E., and Zubairy, S. (2021). State dependence of monetary policy across business, credit and interest rate cycles. *European Economic Review*, 140:103936.

Arias, J. E., Caldara, D., and Rubio-Ramirez, J. F. (2019). The systematic component of monetary policy in SVARs: An agnostic identification procedure. *Journal of Monetary Economics*, 101:1–13.

Aruoba, S. B., Mlikota, M., Schorfheide, F., and Villalvazo, S. (2022). SVARs with occasionally-binding constraints. *Journal of Econometrics*, 231(2):477–499.

Ascari, G. and Haber, T. (2022). Non-linearities, state-dependent prices and the transmission mechanism of monetary policy. *The Economic Journal*, 132(641):37–57.

Barnichon, R. and Matthes, C. (2018). Functional approximation of impulse responses. *Journal of Monetary Economics*, 99:41–55.

Barnichon, R. and Mesters, G. (2023). Evaluating policy institutions-150 years of US monetary policy. Universitat Pompeu Fabra, Economics Working Paper Series 1873.

Bernanke, B. (2004). Gradualism. Technical report, Board of Governors of the Federal Reserve System (US).

Bernanke, B. S. and Gertler, M. (2000). Monetary policy and asset price volatility. NBER Working Paper 7559.

Bernanke, B. S., Gertler, M., and Watson, M. (1997). Systematic monetary policy and the effects of oil price shocks. *Brookings Papers on Economic Activity*, 1997(1):91–157.

Bianchi, F. (2013). Regime switches, agents' beliefs, and post-World War II US macroeconomic dynamics. *Review of Economic Studies*, 80(2):463–490.

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

Boivin, J. (2006). Has US monetary policy changed? Evidence from drifting coefficients and real-time data. *Journal of Money, Credit, and Banking*, 38(5):1149–1173.

Caballero, R. J. and Simsek, A. (2022). Monetary policy with opinionated markets. *American Economic Review*, 112(7):2353–92.

Canova, F. (2004). Testing for convergence clubs in income per capita: A predictive density approach. *International Economic Review*, 45(1):49–77.

Carvalho, C., Nechio, F., and Tristao, T. (2021). Taylor rule estimation by OLS. *Journal of Monetary Economics*, 124:140–154.

Casarin, R., Grassi, S., Ravazzolo, F., and van Dijk, H. K. (2023). A flexible predictive density combination for large financial data sets in regular and crisis periods. *Journal of Econometrics*, 237:105370.

Castelnuovo, E. (2003). Taylor rules, omitted variables, and interest rate smoothing in the US. *Economics Letters*, 81(1):55–59.

Clarida, R., Gali, J., and Gertler, M. (1998). Monetary policy rules in practice: Some international evidence. *European Economic Review*, 42(6):1033–1067.

Clarida, R., Gali, J., and Gertler, M. (2000). Monetary policy rules and macroeconomic stability: evidence and some theory. *The Quarterly Journal of Economics*, 115(1):147–180.

Cogley, T. and Sargent, T. J. (2005). Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2):262–302.

Coibion, O. and Gorodnichenko, Y. (2011). Monetary policy, trend inflation, and the great moderation: An alternative interpretation. *American Economic Review*, 101(1):341–70.

Coibion, O. and Gorodnichenko, Y. (2012). Why are target interest rate changes so persistent? *American Economic Journal: Macroeconomics*, 4(4):126–62.

Compiani, G. and Kitamura, Y. (2016). Using mixtures in econometric models: a brief review and some new results. *The Econometrics Journal*, 19(3):95–127.

Cross, J. L., Hoogerheide, L., Labonne, P., and Van Dijk, H. K. (2024). Bayesian mode inference for discrete distributions in economics and finance. *Economics Letters*, 235:111579.

Davig, T. and Leeper, E. M. (2007). Generalizing the Taylor principle. *American Economic Review*, 97(3):607–635.

Debortoli, D., Forni, M., Gambetti, L., and Sala, L. (2023). Asymmetric monetary policy tradeoffs. BSE Working Paper 1404.

Dolado, J., Pedrero, R. M.-D., and Ruge-Murcia, F. J. (2004). Nonlinear monetary policy rules: some new evidence for the US. *Studies in Nonlinear Dynamics & Econometrics*, 8(3).

Eichenbaum, M., Rebelo, S., and Wong, A. (2022). State-dependent effects of monetary policy: The refinancing channel. *American Economic Review*, 112(3):721–761.

Florio, A. (2006). Asymmetric interest rate smoothing: The Fed approach. *Economics Letters*, 93(2):190–195.

Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. Springer Science & Business Media.

Frühwirth-Schnatter, S. and Kaufmann, S. (2008). Model-based clustering of multiple time series. *Journal of Business & Economic Statistics*, 26(1):78–89.

Frühwirth-Schnatter, S. and Malsiner-Walli, G. (2019). From here to infinity: sparse finite versus Dirichlet process mixtures in model-based clustering. *Advances in data analysis and classification*, 13:33–64.

Frühwirth-Schnatter, S. (2011). *Dealing with Label Switching under Model Uncertainty*, pages 213 – 239.

Gargiulo, V., Matthes, C., and Petrova, K. (2024). Monetary policy across inflation regimes. Federal Reserve Bank of New York Staff Report 1083.

Gelman, A., Goodrich, B., Gabry, J., and Vehtari, A. (2019). R-squared for Bayesian regression models. *The American Statistician*, 73:307–309.

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

Ishwaran, H., James, L. F., and Sun, J. (2001). Bayesian model selection in finite mixtures by marginal density decompositions. *Journal of the American Statistical Association*, 96(456):1316–1332.

Koop, G. (2003). *Bayesian Econometrics*. Wiley.

Koop, G. and Potter, S. M. (2007). Estimation and forecasting in models with multiple breaks. *The Review of Economic Studies*, 74(3):763–789.

Laubach, T. and Williams, J. C. (2003). Measuring the natural rate of interest. *Review of Economics and Statistics*, 85(4):1063–1070.

Leisch, F. (2006). A toolbox for k-centroids cluster analysis. *Computational Statistics & Data Analysis*, 51(2):526–544.

Loria, F., Matthes, C., and Wang, M.-C. (2022). Economic theories and macroeconomic reality. *Journal of Monetary Economics*, 126:105–117.

Lubik, T. A. and Schorfheide, F. (2004). Testing for indeterminacy: An application to US monetary policy. *American Economic Review*, 94(1):190–217.

Malsiner-Walli, G., Frühwirth-Schnatter, S., and Grün, B. (2016). Model-based clustering based on sparse finite gaussian mixtures. *Statistics and Computing*, 26(1):303–324.

Orphanides, A. (2001). Monetary policy rules based on real-time data. *American Economic Review*, 91(4):964–985.

Owyang, M. T. and Ramey, G. (2004). Regime switching and monetary policy measurement. *Journal of Monetary Economics*, 51(8):1577–1597.

Paap, R. and Van Dijk, H. K. (1998). Distribution and mobility of wealth of nations. *European Economic Review*, 42(7):1269–1293.

Panetta, F. (2022). Normalising monetary policy in non-normal times. *Speech at a policy lecture hosted by the SAFE Policy Center at Goethe University and the Centre for Economic Policy Research (CEPR).*

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852.

Rudebusch, G. D. (2006). Monetary policy inertia: Fact or fiction? *International Journal of Central Banking*, 2:85–135.

Sims, C. A. and Zha, T. (2006). Were there regime switches in US monetary policy? *American Economic Review*, 96(1):54–81.

Surico, P. (2007). The Fed's monetary policy rule and US inflation: The case of asymmetric preferences. *Journal of Economic Dynamics and Control*, 31(1):305–324.

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, 39:195–214.

Tenreyro, S. and Thwaites, G. (2016). Pushing on a String: US Monetary Policy Is Less Powerful in Recessions. *American Economic Journal: Macroeconomics*, 8(4):43–74.

Waggoner, D. F. and Zha, T. (2012). Confronting model misspecification in macroeconomics. *Journal of Econometrics*, 171(2):167–184.

Woodford, M. (2003). Optimal interest-rate smoothing. *The Review of Economic Studies*, 70(4):861–886.

Supplementary Material for the paper "Taylor rules with endogenous regimes"

A Additional Results

A.1 Carvalho et al. (2021) with Bayesian estimation

Results in Table A.1 show that the main OLS point estimates for the Taylor Rule coefficients in from Table 1 of (Carvalho et al., 2021, p.151) are robust to the use of Bayesian estimation under standard non-informative independent Normal and inverse-Gamma priors. We also provide new "Full-Data Sample" estimates for the period that we investigate in Section 3.1.

Table A.1: Posterior median for Bayesian estimation of the Taylor rule structural parameters in various periods

	Pre-Volcker 1960Q1-1979Q2	Volcker-Greenspan 1979Q3-2005Q4	Greenspan-Bernanke 1987Q3-2007Q4	Post-Volcker 1979Q3-2007Q4	Full Data Period 1960Q1-2021Q1
β	0.90	1.99	1.39	2.00	1.50
γ	0.79	0.75	1.01	0.81	1.09
ρ	0.80	0.55	0.82	0.56	0.84

We also provide estimates of the associated fitted Taylor Rules and associated monetary policy shocks from exogenous partitioning of the data using the dates in their paper.

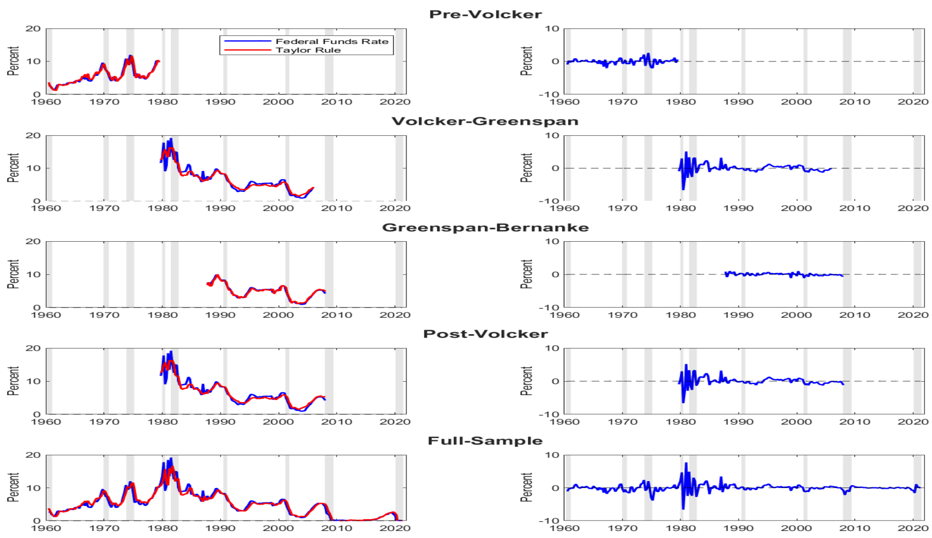


Figure A.1: Fitted Taylor Rule (left) and monetary policy shocks (right) at different periods

A.2 Results with the unemployment rate

The unemployment rate is commonly used to measure the business cycle. For robustness, we therefore run a specification where we replace the output gap with the unemployment rate in our baseline mixture regression.

The data is plotted together with the estimated regime indicator in Figure A.2. In line with the main results we find that the regime indicator captures all recessions except for the 1990s recession.

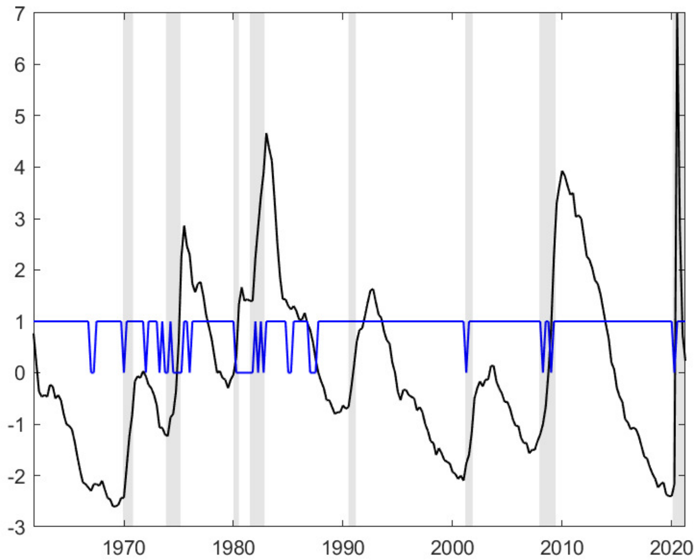


Figure A.2: Regime indicator from the mixture model, defined as the probability of being in regime S1 is larger than 50 percent (blue line) and the demeaned Unemployment rate (black line). The shaded areas show NBER recession periods.

The results in Table A.2 are broadly consistent with the main results. First, we find a credibly different coefficient for interest rate smoothing across the two regimes. This supports our conclusion that the Fed's response to the business cycle is more gradual in normal times, as opposed to during recessions. Second, there is also evidence that the Fed's response to the unemployment rate is larger in the recessionary regime. However, the size of this response is smaller than the one we obtain for our baseline results with the CBO Output gap. Finally, a key difference in these results, compared to those with the output gap, is the relative magnitude of the Fed's responses to inflation and the unemployment rate. In this case, the evidence suggests that the Fed places substantially more weight on targeting inflation in normal times, as opposed to during recessionary periods.

Table A.2: Posterior moments for structural parameters in the mixture model with unemployment instead of the output gap

	Posterior Mean	68% CI	90% CI
	Mixture $S = 1$		
β	2.46	(1.65,3.36)	(1.21,4.23)
γ	0.24	(0.04,0.43)	(0.01,0.77)
ρ	0.95	(0.93,0.97)	(0.92,0.98)
	Mixture $S = 2$		
β	1.27	(0.73,1.74)	(0.38,2.29)
γ	0.53	(0.09,0.93)	(0.03,1.61)
ρ	0.66	(0.49,0.83)	(0.34,0.82)

Notes: The table shows the posterior moments and credible intervals (68% and 90%) for the structural parameters of each components in the mixture model. S1 denotes the "normal time" regime and S2 the "recession" regime. β is the inflation coefficient, γ the unemployment coefficient and ρ the interest smoothing coefficient.

A.3 Results for alternative Greenbook/Tealbook specifications

In section 5.2 we specify mixture Taylor rule estimates when using the average of Greenbook/Tealbook nowcasts and one-quarter ahead forecasts for inflation and the output gap. For robustness, we provide results below when we use only Greenbook/Tealbook nowcasts (Table A.3) and only Greenbook/Tealbook one-quarter ahead forecasts (Table A.4).

Table A.3: Posterior moments and credible intervals for structural parameters in the Mixture model with Greenbook nowcasts over the data period 1975:Q1-2017:Q4.

	Posterior Mean	68% CI	90% CI
	Mixture $S = 1$		
β	1.42	(0.45,2.34)	(0.14,3.12)
γ	1.22	(0.49,2.91)	(0.21,2.64)
ρ	0.91	(0.89,0.94)	(0.74,0.95)
	Mixture $S = 2$		
β	1.34	(0.30,2.45)	(0.09,3.61)
γ	1.81	(0.43,3.35)	(0.13,4.38)
ρ	0.53	(0.33,0.72)	(0.17,0.80)

Notes: The table shows the posterior moments and credible intervals (68% and 90%) for the structural parameters of each components in the mixture model when using Greenbook/Tealbook data. We use nowcasts for inflation and output gap, respectively. For inflation we use data for the GDP deflator. S1 denotes the "normal time" regime and S2 the "recession" regime. β is the inflation coefficient, γ the output gap coefficient and ρ the interest smoothing coefficient.

Table A.4: Posterior moments and credible intervals for structural parameters in the Mixture model with Greenbook one-quarter ahead forecasts over the data period 1975:Q1-2017:Q4.

	Posterior Mean	68% CI	90% CI
	Mixture $S = 1$		
β	1.51	(0.62,2.37)	(0.24,3.09)
γ	0.72	(0.21,1.21)	(0.07,1.73)
ρ	0.92	(0.90,0.94)	(0.88,0.95)
	Mixture $S = 2$		
β	1.39	(0.32,2.52)	(0.10,3.74)
γ	1.52	(0.40,2.72)	(0.13,3.89)
ρ	0.56	(0.36,0.75)	(0.20,0.83)

Notes: The table shows the posterior moments and credible intervals (68% and 90%) for the structural parameters of each components in the mixture model when using Greenbook/Tealbook data. We use one-quareter ahead forecasts for inflation and output gap, respectively. For inflation we use data for the GDP deflator. S1 denotes the "normal time" regime and S2 the "recession" regime. β is the inflation coefficient, γ the output gap coefficient and ρ the interest smoothing coefficient.