Monetary Policy across Inflation Regimes*

Valeria Gargiulo
Universitat Pompeu Fabra

Christian Matthes
Indiana University

Katerina Petrova †
Federal Reserve Bank of New York, Universitat Pompeu Fabra and BSE

December 25, 2024

Abstract

Does the effect of monetary policy depend on the prevailing level of inflation? In order to answer this question, we construct a parsimonious nonlinear time series model that allows for inflation regimes. We find that the effects of monetary policy are markedly different when year-over-year inflation exceeds 5.5 percent. Below this threshold, changes in monetary policy have a short-lived effect on prices, but no effect on the unemployment rate, giving a potential explanation for the recent "soft-landing" in the United States. Above this threshold, the effects of monetary policy surprises on both inflation and unemployment can be larger and longer-lasting.

Keywords: Monetary policy shocks, Inflation, Regime-dependence, Outliers, Nonlinear time series models

*The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve System. We would like to thank Davide Debortoli, Jordi Galí, Argia Sbordone, as well as seminar participants at the University of Würzburg, the ECB, the Central Bank of Italy, the RCEA ICEEF 2024, and the BSP International Research Fair for very useful comments. We thank Silvia Miranda-Agrippino for providing the updated monetary instrument used in Appendix D.

†Petrova acknowledges support from the General Directorate for Research of the Government of Catalonia through the Beatriu de Pinós grant no. 2019/BP/00239 and the Spanish Ministry of Science and Innovation, through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S) for the period prior to September 2023.

1 Introduction

What are the effects of monetary policy on the economy? Although this question has long been a cornerstone of macroeconomic research (see Christiano et al. (1999) and the references therein), it has recently become extremely pertinent again, as, after the COVID-19 pandemic, most major economies experienced inflation rates not seen since the early 1980s. In this paper, we investigate whether the effects of monetary policy and the tradeoffs that policymakers face depend on the level of inflation that prevails in the economy. Policymakers rely on economic models to decide their course of action, but much of the research on the effects of monetary policy is based on linear models with constant parameters, disregarding any potential state-dependence, for example on the underlying level of inflation. There are reasons to expect that economic agents' behavior is different when inflation deviates considerably from its target. For example, Weber et al. (2023), in a series of randomized controlled trials (RCTs), show that both consumers and firms react to information and form expectations differently when inflation is high. Households might start to pay more attention to inflation, in line with rational-inattention-based theories (Sims (2003) and Weber et al. (2023)); firms change their price-setting behavior in high-inflation environments (Golosov and Lucas, 2007); and the central bank might adjust its behavior to more aggressively combat inflation. Ascari and Haber (2021) also document nonlinearities in the response of inflation and real activity to monetary policy shocks. [1] Historical evidence from the oil price crises of the 1970s also indicates that high inflation is associated with more persistent macroeconomic variables and greater volatility. Ignoring this state-dependence can lead to flawed empirical conclusions.

In this paper, we address this problem by building a parsimonious and computationally tractable nonlinear vector autoregressive (VAR) model, a self-exciting threshold (SET-) Bayesian VAR model that delivers easily interpretable nonlinearities and allows the data to identify the inflation regimes as well as the regime-dependent parameters. We identify the effects of monetary policy using an instrument for monetary policy shocks, an approach commonly used in the literature. As such, our identification strategy within each regime is the same as that in many papers that focus on linear models.

We find that our approach estimates regime changes at high levels of inflation, thus endogenously separating periods where inflation rates are not representative of most of the sample. Seventy-five percent of our sample falls into one inflation regime where year-over-year CPI-

[1] They find that inflation reacts stronger in high inflation setups, which is consistent with our empirical findings.

based inflation is less than 5.5 percent.[2] In this regime, the persistence of macroeconomic variables is low, and hence the effects of shocks are short-lived. Monetary policy can reduce inflation, but it has no meaningful effect on the unemployment rate, thus providing a rationale for the recent "soft landing" of the U.S. economy, at least once year-over-year inflation became lower than 5.5 percent. This recent episode is not included in our sample[3], so it is not the case that our model simply fits these recent data. Instead, this result is driven by the different reduced-form relationships that we uncover between inflation and unemployment when inflation levels are low and stable. Once inflation becomes larger, monetary policy has larger and more persistent effects on prices and a significant effect on the unemployment rate. Finally, we find that the persistence and the effects of policy shocks do not monotonically change with the level of inflation, and once the underlying inflation rate becomes double-digit (larger than 11 percent), the policy effects disappear and we find a price puzzle. A standard linear VAR model, applied to our sample, would incorrectly suggest that monetary policy has no effect on prices or unemployment, as it averages effects across the distinct regimes.

Part of the contribution of the paper is methodological. We built a self-exciting threshold Bayesian VAR model, allowing for regime-dependence in the variance of the series for full likelihood-based identification of the threshold parameter, which is particularly relevant for macroeconomic series that have been documented to undergo volatility regimes over time. This is novel relative to existing approaches based on sum-of-square-residuals (SSR) minimization which either assume a constant variance or only allow for variance regimes in the second stage estimation, but do not utilize volatility regimes to aid the threshold parameter identification. We demonstrate in a small Monte Carlo exercise that when regime-dependence in the volatility is present, not taking this into account in the first stage threshold estimation can lead to inconsistent threshold estimates.

Our second methodological contribution is computational: we combine two-step frequentist estimation procedures with Bayesian regularization via the use of priors on the VAR parameters in order to deliver a parsimonious nonlinear time series model. We can ignore posterior uncertainty on the threshold parameters without distorting posterior inference on the VAR

[2]Using a different modeling approach and only allowing for two regimes, Canova and Pérez Forero (2024)) discover a strikingly similar threshold of 5.3 percent.

[3]We use the updated Romer and Romer (2004) monetary policy shock series as our instrument, which is based on staff forecasts from the Board of Governors before each FOMC meeting. These forecasts are made public with a five-year lag, preventing us from extending our analysis to the most recent period. Similar issues arise with the instrument proposed by Aruoba and Drechsel (2022). Other instruments based on high-frequency variation in asset prices are available for more recent periods, but these samples start in the late 1980s, missing the high inflation periods of the 1970s and early 1980s.

parameters since the contraction rate of the posterior of the threshold parameters is faster and hence conditioning on the entire posterior of the threshold or just on a consistent point estimate is equivalent when the sample is large. The main advantages of our econometric approach are: (i) it can handle a large number of variables allowing for standard Bayesian treatment on the large dimensional regime-dependent VAR parameters, (ii) it is considerably easier and faster to estimate than the model with a fully Bayesian treatment of the threshold parameters, and (iii) it provides a simple and easily interpretable source of nonlinearity, relative to Markov-switching models widely used in the literature, where the regimes are driven by an unobserved latent process.

The remainder of the paper is organized as follows. Section 2 presents a stylized example to highlight the pitfalls of disregarding nonlinearities when estimating impulse responses/causal effects of monetary policy and motivate the usefulness of the threshold model. Section 3 establishes in detail the econometric methodology and explains the novelties of our model relative to existing threshold models in the literature. Section 3.2 presents a small Monte Carlo exercise that demonstrates the merits of our novel approach relative to existing approaches. Section 4 contains our empirical application to monetary policy in the US across inflation regimes. Section 5 concludes and the supplementary Appendix contains some additional results.

2 The Pitfalls of Disregarding Inflation Regimes - An Example

To highlight that the causal effects of monetary policy can change dramatically once we allow for nonlinearities, which can endogenously remove outliers/unusual periods of high inflation, we present a stylized Monte Carlo exercise to illustrate this issue in a controlled environment. For simplicity, we focus on one endogenous variable z_t, one shock of interest m_t and a variable u_t that summarizes the persistent effects of all other variables in the economy. Our data-generating process is:

$$u_t = \rho u_{t-1} + v_t$$

$$z_t = \begin{cases} \beta m_t + \rho z_{t-1} + u_t \text{ if } z_{t-1} < \overline{z} & (1) \\ \gamma m_t + \rho z_{t-1} + u_t \text{ otherwise} & (2) \end{cases}$$

where v_t and m_t are zero-mean i.i.d. Gaussian random variables, and $\overline{z}$ is the threshold value, which we calibrate so that z_t infrequently exceeds the threshold and so the model is more often in the first regime. We assume opposing effects of the shock in the two regimes: $\beta < 0$ and $\gamma > 0$. The exact calibration is $\rho = 0.8$, $\overline{z} = 13$, $\beta = -0.5$, $\gamma = 7$, $m_t \sim \mathcal{N}(0,1)$, and $v_t \sim \mathcal{N}(0,1.5)$. We simulate 5,000 samples of size 500 each. For each sample, we run two ordinary least squares regressions to estimate the policy effect of m_t on z_t: one regression for the entire sample and another where we run the same regression, but only for the observations where $z_{t-1} < \overline{z}$ so that we do not consider the outliers. This is equivalent to estimating the threshold model that we present later with knowledge of the true threshold value.

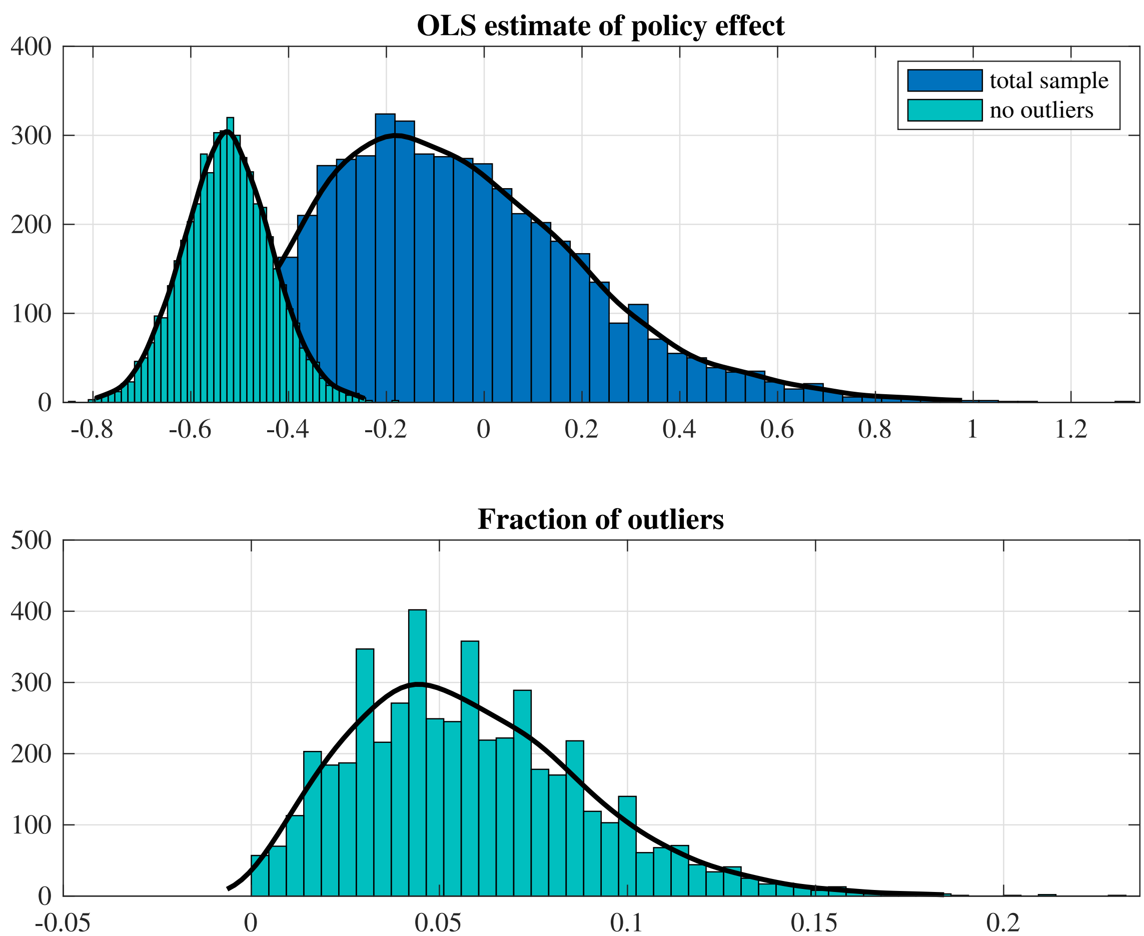

Figure 1: Histogram of OLS coefficients with and without outliers (top panel) and frequency of outliers (bottom panel). Black lines are kernel estimates of the densities associated with each histogram.

Even with infrequent outliers in each sample (the frequency of outliers across the sample is around 5 percent), the outliers substantially contaminate the results if regime-dependence is ignored, as expected. We argue that although the blue histogram in the top panel of Figure 1 would ultimately collapse to the total pooled effect of the shock across the two regimes, this is not a value that most economists would be interested in. Instead, they would be interested in the causal effects of m_t on z_t for most of the sample, which is -0.5 and well approximated by the light blue histogram in the top panel, as well as in the different and possibly relevant causal effects in the extraordinary periods when z_t exceeds the threshold. Pooling the regimes together by fitting a linear model can cause a bias that does not vanish as the sample size increases and can lead to erroneous empirical conclusions.

3 Methodology

In this section, we describe our econometric methodology. This methodology is set up to identify regime-specific causal effects as outlined in the previous section. Modeling nonlinearities is a challenging task, especially in environments with dependent data. The modeling choices we make in this paper are guided by three main goals: (i) parsimony, (ii) transparency and (iii) computational speed. We prefer a simple model in which the factors determining the nonlinearities are clear and easily interpretable. Moreover, our model is set up so that inference is computationally fast and straightforward, while at the same time allowing for regularization via the use of priors, opening up the use of our nonlinear model for applications with many observables. In a nutshell, our model is a Bayesian self-exciting threshold VAR: a piecewise-linear VAR model where breaks in the model's parameters are governed by lagged observables and, within each regime, defined by the model's threshold parameters, the model is linear and inference is standard. We make two important departures from existing models in the literature. First, we allow the regimes to be identified via regime-dependence in the conditional variance in addition to regime-dependence in the conditional mean of the series. Such an extension is relevant for macroeconomic applications where series undergo different volatility regimes over time and hence information from regime-dependence in the second moment may be useful to exploit in order to identify and estimate more precisely the threshold parameters. The second novelty of our approach relative to existing models in the literature is the use of Bayesian inference on the VAR parameters while maintaining fast and efficient Bayesian estimation of the threshold without the need for a computationally expensive MCMC step. The remainder of this section provides a detailed technical description and justification of our modeling approach.

We first provide a brief discussion of the standard practice of estimation of threshold VAR (T-VAR) models in the literature before we outline how we depart from it and then establish the novel estimation methodology that we adopt. The univariate TAR model was first introduced by Tong (1977) and generalized in various directions by Tong and Lim (1980), Chan (1993), and Tong (2011), among others. Here we consider a multivariate generalization given by an $M \times 1$ T-VAR model of order p, characterized by k regimes:

$$y_t = \sum\nolimits_{i=1}^{k} \left(B_{0,i} + \sum\nolimits_{j=1}^{p} B_{j,i} y_{t-j} \right) \Psi_t^{(i)}(\gamma^0) + \Sigma^{1/2}\eta_t, \ \ \eta_t | \mathcal{F}_{t-1} \sim (0, I_M) \tag{3}$$

where the index $i = \{1, ..., k\}$ refers to each regime, $B_{0,i}$ is a vector of state-dependent intercepts, $B_{j,i}$ are state-dependent autoregressive matrices with all the roots of the associated characteristic polynomial outside the unit disk for each i, Σ is a positive definite covariance matrix and η_t are martingale difference innovations with $\mathcal{F}_t$ denoting the natural filtration of η_t containing information up to t. The choice of matrix square root to obtain $\Sigma^{1/2}$ will encode the identification restrictions that we use to identify the effects of monetary policy in our application; we discuss the details of our choice later. The parameter γ^0 is a $(k-1) \times 1$ threshold parameter vector which defines the regimes, with $\gamma_1^0 < \gamma_2^0 < \cdots < \gamma_{k-1}^0$, and $\Psi_t^{(i)}(\gamma^0)$ is an indicator function equal to one whenever a threshold condition associated with regime i is satisfied at period t. It is standard to assume that the regimes are driven by an underlying state variable s_t which is $\mathcal{F}_{t-1}$-measurable, which can be either internal or external to the model. The i^{th} regime is defined as all periods t such that $\Psi_t^{(i)}(\gamma^0) = \mathbb{I}(\gamma_{i-1}^0 < s_t \leq \gamma_i^0)$ (with $\Psi_t^{(1)}(\gamma^0) = \mathbb{I}(s_t \leq \gamma_1^0)$ and $\Psi_t^{(k)}(\gamma^0) = \mathbb{I}(s_t > \gamma_{k-1}^0)$ for the first and last regimes respectively), where $\mathbb{I}$ is the indicator function.

Next, we discuss the choice of the state variable s_t. Whenever s_t is a lagged variable from the vector y_t with lag $d \in \{1, ..., p\}$, the model is called a self-exciting T-VAR (SET-VAR) model. A SETAR model and its multivariate extensions have two important desirable properties: (i) the nonlinearity through the indicator functions makes the model piecewise linear, which facilitates simple estimation relative to more complex nonlinear models; (ii) while simple, the self-exciting mechanism can capture important nonlinearities that are particularly relevant in cyclical data and SETAR models can generate statistical phenomena ranging from jump resonance, nonlinear vibrations, jump cycles, harmonic distortions and even chaos (see Tong and Lim (1980) for a discussion and examples). Consistency and the resulting asymptotic distributions of the LS estimators in SETAR models are established in Chan (1993), and the associated limit theory in this literature is established by typically showing that the Markov chain defined by the companion form of the process is geometrically ergodic[4].

Letting $B_i = (B_{0,i}, B_{1,i} \ldots, B_{p,i})$ and $\beta_i = vec(B_i')$, conditional on the true threshold parameter γ^0, the estimation of the regime-dependent parameter vector β_i is standard. In particular, conditional on the true value of γ, the OLS estimator of β_i, for each regime i, is $\sqrt{n}-$consistent and asymptotically normal (see, e.g. Tong (2011)). Since γ is unknown in most empirical applications, a consistent estimator of γ is required for the estimation of β_i. This is typically done via a numerical minimization of the sum of squared residuals (SSR) as a function of γ (see Hansen (1997)). In practice, the vector $\beta = (\beta_1', \ldots, \beta_k')'$ is estimated via OLS for a grid of values of the threshold. Then, $\hat{\beta} = (\hat{\beta}_1', \ldots, \hat{\beta}_k')'$ is used to compute the residuals for all possible values of the grid for the threshold parameter γ and an estimator for γ is given by the value that attains the minimum SSR:

$$\hat{\gamma} = \arg\min_{\gamma} \sum\nolimits_{t=1}^{n} \hat{\varepsilon}_t' \hat{\varepsilon}_t = \arg\min_{\gamma} \left[\min_{\beta_1, \ldots, \beta_k} \sum\nolimits_{t=1}^{n} \varepsilon_t' \varepsilon_t \right]$$

where $\varepsilon_t = \left(y_t - \sum_{i=1}^{k} (I_M \otimes x_t') \beta_i \Psi_t^{(i)}(\gamma) \right)$, $\hat{\varepsilon}_t = \left(y_t - \sum_{i=1}^{k} (I_M \otimes x_t') \hat{\beta}_i \Psi_t^{(i)}(\gamma) \right)$ and $x_t = \left(1, y_{t-1}', ... y_{t-p}'\right)'$. The minimizer $\hat{\gamma}$ of the above minimization is equivalent to the estimator $\hat{\gamma}$ coming from joint minimization of the sum of squared residuals function:

$$\left(\hat{\gamma}, \hat{\beta}\right) = \arg\min_{\gamma, \beta} \sum\nolimits_{t=1}^{n} \varepsilon_t' \varepsilon_t.$$

The standard identification assumption in the literature is that for all regimes $i \in \{1, \ldots, k\}$, the following condition holds:

$$\forall i, j \in \{1, \ldots, k\}, \quad \beta_i \neq \beta_j \quad \text{when} \quad i \neq j; \tag{4}$$

in other words, at least one of the elements in the parameter vector β is required to differ across any pair of regimes. Super-consistency of $\hat{\gamma}$ to the true threshold value can be established under regularity conditions (e.g. Chan (1993)) with a faster rate of convergence to the true γ^0 (n instead of the usual parameteric $\sqrt{n}$). Inference in this model is typically conducted in two steps: (i) γ is estimated and the estimator $\hat{\gamma}$ is set as the threshold in the subsequent analysis, and (ii) conditional on $\hat{\gamma}$, inference on β_i is standard, consistent and

[4]This requires stability conditions on the autoregressive parameters across regimes, such as $\max_i \rho(F_i) < 1$, where $\rho(.)$ denotes the spectral radius and F_i is the companion matrix based on the autoregressive matrices $B_{j,i}$ in regime i.

asymptotically normal. The super-consistency for $\hat{\gamma}$ implies that estimation uncertainty of γ does not have a first-order effect on inference (e.g. limit distribution) for β_i and hence can be ignored, providing a justification for the two-step plug-in procedure described above and widely used in the literature. A similar two-step estimation method can be found in Samia and Chan (2011), where the objective function considered for γ is a likelihood function instead with error covariance constant across regimes. While in some papers the variance is allowed to be regime-specific (e.g. Chan (1993) and Tsay (1998)) in the second estimation stage, estimation of the threshold γ is identified only through regime-dependence in the conditional mean of the series. This could be a drawback of existing methods if additional information on the regimes contained in the second moments is ignored when estimating the threshold parameter γ in the first step. Since such additional information on the volatility of the series may be useful for identifying γ, particularly in macroeconomic data where we know that some regimes were characterized not just by mean but also by volatility changes, we extend the estimator of Samia and Chan (2011) by proposing a novel way to estimate γ by including the parameters in Σ in the regime-dependent parameter vector. Our approach is based on the use of a likelihood function, and, crucially, we allow for the variance parameters Σ to switch across regimes (that is, Σ_i may differ across i) in both estimation stages[5], enabling us to exploit additional information contained in the second moment of the series. Such an extension is economically relevant, since it allows us to identify regimes even when there may not be an associated break in the conditional mean but only in the conditional variance of the data. There is ample empirical evidence for the importance of allowing for the volatility of the series to change over time to properly capture the behaviour of the macroeconomy (see, e.g. Cogley and Sargent (2005), Primiceri (2005)).

The second novelty of our estimation procedure relative to existing frequentist and Bayesian approaches is that we allow for a Bayesian treatment of the model's autoregressive parameters and covariance matrices across regimes as well as for a prior distribution on the threshold parameter while maintaining computational efficiency. We achieve such computational gains and the proper Bayesian treatment for the regime-dependent VAR parameters by using a Bayesian point estimator for γ. Since such a point estimator converges at a faster rate than the VAR parameter estimates, the two-step procedure that we propose is well-justified, since the posterior uncertainty of γ does not affect the posterior of the VAR

[5] A semi-parametric equivalent to our parametric likelihood approach would amount to considering a GLS rather than an OLS objective function, i.e., minimization of the Mahalanobis instead of the Euclidean norm of the innovations in the first stage.

parameters for large samples[6]. In other words, letting $\theta_i = [\beta_i', vech\,(\Sigma_i)']'$ where $vech(.)$ denotes the half-vec operator, the difference between the scaled posterior centered around the true value θ_i^0 of the VAR parameters for regime i conditional on posterior values of the threshold γ and conditional on a super-consistent estimator $\hat{\gamma}$ satisfies:

$$p(\sqrt{n}(\theta_i - \theta_i^0)|\gamma, y_1, ..., y_n) - p(\sqrt{n}(\theta_i - \theta_i^0)|\hat{\gamma}, y_1, ..., y_n)) \to 0 \text{ as } n \to \infty.$$

This approach is in contrast to a fully Bayesian treatment of the threshold parameter γ, (see, e.g. Chen and Lee (1995), and Alessandri and Mumtaz (2017) and Alessandri and Mumtaz (2019) for applications), which requires approximating the posterior of γ through an expensive Metropolis step[7]. A Bayesian treatment of the VAR parameters is particularly relevant in the context of the SET-VAR model, since a large number of variables and lags can result in frequentist procedures overfitting, especially after splitting the observations of the sample into regimes, and a prior distribution can be extremely useful to penalize and regularize the estimation procedure.

We now turn to describing in detail the methodology we use in this paper. The VAR model with regime-dependent conditional mean and covariance is given by:

$$y_t = \sum\nolimits_{i=1}^{k} \left(B_{0,i} + \sum\nolimits_{j=1}^{p} B_{j,i} y_{t-j} + \Sigma_i^{1/2} \eta_t \right) \Psi_t^{(i)}(\gamma^0), \;\; \eta_t | \mathcal{F}_{t-1} \sim (0, I_M).$$

We assume a prior density for the VAR parameters $p\,(\beta_i, \Sigma_i)$ for each regime $i = 1, ..., k$ as well as a prior density for the threshold parameter $p\,(\gamma)$ independent from $p\,(\beta_i, \Sigma_i)$. For the sake of generality, we allow here for different priors $p(\beta_i, \Sigma_i)$ across regimes. In our empirical application, we use the same prior for all regimes to ensure that the uncovered differences across regimes are coming from the data rather than from prior beliefs. We consider a fine grid of N_γ equidistant points $\Gamma = \left(\underline{\gamma}, ..., \overline{\gamma}\right)$ for each element of γ, which corresponds to a discrete uniform prior for each element i: $p\,(\gamma_{ij}) = 1/N_\gamma$ for $\gamma_{ij} \in \Gamma$ for each gridpoint $j = 1, ..., N_\gamma$. Since we need distributional assumptions in order to write down a likelihood function, we proceed by making a Gaussianity assumption[8] on the innovations

[6]Alternatively, one can view γ as a hyperparameter, whose value is determined in a preliminary estimation step, which is a common approach in Bayesian inference (Giannone et al., 2015).

[7]Broemeling and Cook (1992) and Geweke and Terui (1993) provide earlier Bayesian treatment of γ in TAR models, obtaining a posterior through (numerical) integration.

[8]Such a distributional assumption is required for full information Bayesian estimation; however, posterior inference on the conditional mean parameters B_i continues to be valid for large samples even if the distribution of the innovations is non-Gaussian, as long as the first two conditional moments of the innovations are correctly specified (see, e.g. Petrova (2022)).

$\eta_t|\mathcal{F}_{t-1} \sim \mathcal{N}(0, I_M)$. The log-posterior density of the model's parameters (except constants) is given by:

$$\ln(p(\beta, \Sigma, \gamma|y_1, ..., y_n)) = \ell(\beta, \Sigma, \gamma) + \sum\nolimits_{i=1}^{k} \omega_i \ln p(\beta_i, \Sigma_i) + \ln p(\gamma),$$

where the log-likelihood of the sample $\ell(\beta, \Sigma, \gamma)$ is given by the sum of the log-likelihoods across each regime:

$$\begin{aligned} \ell(\beta, \Sigma, \gamma) &= \sum\nolimits_{i=1}^{k} \ell_i(\beta, \Sigma, \gamma) \\ \ell_i(\beta, \Sigma, \gamma) &= -\frac{n_i}{2} \ln(2\pi) - \frac{n_i}{2} \ln\det(\Sigma_i) - \frac{1}{2} \sum\nolimits_{t=1}^{n} \varepsilon_{it}' \Sigma_i^{-1} \varepsilon_{it}, \end{aligned}$$

the innovations for each regime ε_{it} are defined as $\varepsilon_{it} = (y_t - (I_M \otimes x_t')\beta_i)\Psi_t^{(i)}(\gamma)$ and the weights ω_i depend on the contribution of each regime in the sample, satisfying $\omega_i = \frac{n_i}{n}$, with $\sum_{i=1}^{k} \omega_i = 1$, where n_i is the effective sample sizes in each regime i, $n_i = \sum_{t=1}^{n} \Psi_t^{(i)}(\gamma)$. The problem can be equivalently rewritten in a more compact way as a reweighting scheme of the likelihood of the observations $(y_1, \ldots, y_n)$ with flat (zero-one) weighting given by the regimes: for each regime $i \in \{1, ..., k\}$, observations that satisfy the threshold condition for the corresponding regime (i.e. $\Psi_t^{(i)}(\gamma^0) = \mathbb{I}(\gamma_{i-1}^0 < s_t \leq \gamma_i^0)$) are given weight one to evaluate the regime-specific likelihood $\ell_i(\beta, \Sigma, \gamma)$, with the remaining observations receiving weight zero. For each regime $i \in \{1, ..., k\}$, we denote the weights for the likelihood as

$$w_{t,i} = \mathbb{I}(\hat{\gamma}_{i-1} < s_t \leq \hat{\gamma}_i) \tag{5}$$

and further define the matrices $Y = (y_1, \ldots, y_n)'$, $X = (x_1', \ldots, x_n')'$ and $W_i = diag(w_{1,i}, \ldots, w_{n,i})$. The resulting "weighted" log-likelihood for each regime i of the sample $y = vec(Y)$ can be written as:

$$\ell_i(\beta, \Sigma, \gamma) \propto -\frac{tr(W_i)}{2} \ln(\det \Sigma_i) - \frac{1}{2}(y - (I_M \otimes X)\beta_i)' (\Sigma_i^{-1} \otimes W_i)(y - (I_M \otimes X)\beta_i),$$

where $tr(W_i) = n_i$ gives the "effective" regime sample sizes. Next, we consider joint maximization of the log-posterior density

$$\left(\hat{\gamma}, \hat{\beta}, \hat{\Sigma}\right) = \arg\max_{\gamma,\beta,\Sigma} \ln(p(\beta, \Sigma, \gamma|y_1, ..., y_n)),$$

where the maximizer $\hat{\gamma}$ can be equivalently obtained through

$$\begin{aligned}\hat{\gamma} &= \arg\max_{\gamma}\left[\max_{\beta_1,\ldots,\beta_k,\Sigma_1,\ldots,\Sigma_k} \ln(p\left(\beta,\Sigma,\gamma|y_1,...,y_n\right))\right] = \arg\max_{\gamma} \ln\left(p(\breve{\beta},\breve{\Sigma},\gamma|y_1,...,y_n)\right) \\ &= \arg\max_{\gamma\in\Gamma^{k-1}} \ell\left(\breve{\beta},\breve{\Sigma},\gamma\right) + \sum\nolimits_{i=1}^{k} \omega_i \ln p\left(\breve{\beta}_i,\breve{\Sigma}_i\right) \end{aligned} \tag{6}$$

where $\breve{\beta}$ and $\breve{\Sigma}$ are the posterior modes (the maximizers of the posterior density), as a function of the threshold parameter γ, Γ^{k-1} is the $(k-1)$ Cartesian product of the grid Γ and, in the last line, we have used the fact that our uniform prior for the threshold does not affect the maximizer $\hat{\gamma}$ other than through the set over which the maximization is performed $\gamma \in \Gamma^{k-1}$. It is straightforward to use our procedure with an informative continuous prior on γ. We choose the discrete uniform prior since: (i) it simplifies and streamlines the estimation through the grid search maximization, (ii) the threshold is a low-dimensional parameter which does not require penalization, and (iii) we prefer to let the data speak on the threshold values and choose not to impose any informative prior beliefs ex ante.

Next, we evaluate the corresponding log-posterior density at the posterior mode of the entire vector $\theta_i = [\beta_i', vech\,(\Sigma_i)']'$ over the Γ^{k-1}-dimensional grid of values for γ and estimate the threshold γ as the maximizer over the grid.

Conditional on the threshold estimator $\hat{\gamma}$, we proceed with standard Bayesian estimation for θ_i. Since the VAR model can be over-parameterized, especially when the number of variables M and the number of lags p is large and the sample size n is small, we follow a standard conjugate Bayesian methodology for the conditional inference on θ_i with standard Minnesota prior on B_i and Wishart prior on Σ_i^{-1} for each regime $i \in \{1,...,k\}$ of the form

$$\beta_i|\Sigma_i,\hat{\gamma} \sim \mathcal{N}\big(\beta_{0i},(\Sigma_i^{-1}\otimes\kappa_{0i})^{-1}\big), \quad \Sigma_i^{-1}|\hat{\gamma} \sim \mathcal{W}(\alpha_{0i},\lambda_{0i}) \tag{7}$$

where β_{0i} is a vector of prior means, κ_{0i} is a positive definite matrix controlled through a scalar overall shrinkage parameter, α_{0i} is the Wishart distribution scale parameter, and λ_{0i} is a positive definite matrix. While our methodology allows for the use of priors that differ across regimes, as mentioned before, for the empirical application of the paper, we set the same priors for each regime, since we do not want to impose regime-dependence ex ante.

In this way, our piecewise-linear Gaussian model with Normal-Wishart prior distribution for β_i and Σ_i^{-1} for each regime $i \in \{1,...,k\}$ yields a closed-form conjugate Normal-Wishart expression for the posterior density across each regime, conditional on the threshold γ of the form:

$$\beta_i|\Sigma_i,\hat{\gamma},X,Y \sim \mathcal{N}\big(\tilde{\beta}_i,(\Sigma_i^{-1}\otimes\tilde{\kappa}_i)^{-1}\big),\quad \Sigma_i^{-1}|\hat{\gamma}\sim\mathcal{W}(\tilde{\alpha}_i,\tilde{\lambda}_i), \tag{8}$$

where the posterior parameters $\tilde{\beta}_i$, $\tilde{\kappa}_i$, $\tilde{\alpha}_i$, $\tilde{\lambda}_i$ for each regime i are given by

$$\begin{aligned}
&\widetilde{\beta}_i = \left(I_M\otimes\widetilde{\kappa}_i^{-1}\right)\left[(I_M\otimes X'W_iX)\hat{\beta}_i+(I_M\otimes\kappa_{0i})\beta_{0i}\right],\\
&\widetilde{\kappa}_i=\kappa_{0i}+X'W_iX,\quad \widetilde{\alpha}_i=\alpha_{0i}+n_i,\ \widetilde{\lambda}_i=\lambda_{0i}+Y'W_iY+B_{0i}\kappa_{0i}B_{0i}'-\widetilde{B}_i\widetilde{\kappa}_i\widetilde{B}_i',
\end{aligned}$$

where $\hat{\beta}_i$ is the threshold OLS estimator for each regime i :

$$\hat{\beta}_i=(I_M\otimes X'W_iX)^{-1}(I_M\otimes X'W_i)y,$$

W_i is the diagonal matrix containing the zero-one weights for each regime defined in (5) through the estimated threshold $\hat{\gamma}$, $X=(x_1',...,x_T')'$ and $\widetilde{B}_i$ and B_{0i} satisfy $\tilde{\beta}_i := vec(\tilde{B}_i')$ and $\beta_{0i}:=vec(B_{0i}')$. The full details of our estimation algorithm can be found below.

3.1 Our Estimation Algorithm

Step 1. For each grid point in Γ^{k-1}, compute the posterior modes for β_i and Σ_i, which in our Normal-Wishart setup are given by $\breve{\beta}_i=\tilde{\beta}_i$ and $\breve{\Sigma}_i=\frac{\tilde{\lambda}_i}{(\tilde{\alpha}_i+M+1)}$.
Step 2. Evaluate the log-likelihood of the sample $\ell\left(\breve{\beta},\breve{\Sigma},\gamma\right)$ and the weighted prior density $\sum_{i=1}^{k}\omega_i\ln p\left(\breve{\beta}_i,\breve{\Sigma}_i\right)$ at the posterior modes $\breve{\beta}$ and $\breve{\Sigma}$ for each grid point in Γ^{k-1}.
Step 3. Numerically maximize the log-posterior $p\left(\breve{\beta},\breve{\Sigma},\gamma|y_1,...,y_n\right)$ with respect to γ over the $(k-1)$-dimensional grid and store the estimate $\hat{\gamma}$.
Step 4. Given $\hat{\gamma}$ from Step 3, make draws β_i and Σ_i from the posterior distribution in (8).

3.2 Monte Carlo Exercise

In this section, we design a small Monte Carlo exercise to study the properties of the proposed estimator and how it compares to existing threshold estimation approaches. In particular, we simulate data from four data generating processes (DGPs) with sample sizes $n\in\{200,1000\}$. In all cases, we generate observations from the following univariate process

$$y_t=\mu_t+\sigma_t\varepsilon_t,\ \varepsilon_t\sim\mathcal{N}\left(0,1\right),$$

with the following specifications: (i) DGP I: constant mean and constant volatility $\mu_t = 0$ and $\sigma = 1$ for all t; (ii) DGP II: regime-dependent mean and constant volatility $\mu_t = \mu_1 1\{y_{t-1} \leq \gamma\} + \mu_2 1\{y_{t-1} > \gamma\}$ with $\mu_1 = -1$, $\mu_2 = 1$ and $\sigma_t = 1$ for all t; (iii) DGP III: constant mean $\mu_t = 0$ for all t and regime-dependent volatility $\sigma_t = \sigma_1 1\{y_{t-1} \leq \gamma\} + \sigma_2 1\{y_{t-1} > \gamma\}$ with $\sigma_1 = 1$, $\sigma_2 = 2$; and (iv) DGP IV: regime-dependent mean as in (ii) and regime-dependent volatility as in (iii). For all the DGPs with regime-dependence, we set $\gamma = 0$. We estimate four models for each DGP: (i) a constant parameter model; (ii) a regime-dependent threshold model based on SSR minimisation with constant variance; (iii) a version of (ii) allowing for regime-dependence in the variance in the second stage (as in Chan (1993) and Tsay (1998)), and (iv) a regime-dependent threshold model based on likelihood maximisation proposed in the previous section where both mean and variance regime-dependence enters in both stages of estimation. In each case, we compare root mean square errors (RMSEs) for the estimated parameters of each specification and each DGP in Table 1 below.

Several conclusions emerge from the results. First, when there is no regime dependence as in DGP I, all approaches are valid for inference on the mean and variance parameters, as implied by RMSEs decreasing as the sample size increases; however, the constant parameter model is considerably more efficient, as expected in the absence of regimes. Next, when there is a switch in the mean, as in DGP II, the constant parameter model is inadequate and inconsistent not just for the mean but also the variance parameters. The model with SSR objective function delivers consistent estimator for the threshold parameter, as well as for the mean and variance parameters; and this is the case, whether or not regime-dependence is allowed in σ in the second stage, since the DGP has a constant variance. The likelihood-based approach which uses regime-dependence in σ to identify the threshold performs equally well for all parameters, leading to the conclusion that adding regime-dependence in σ in the first stage estimation does not distort inference on γ when such dependence is absent in the data. In DGP III, on the other hand, where only regime-dependence in the variance is present, both models without regime-dependence in σ in the first stage deliver inconsistent estimator for the threshold, as implied by the non-decreasing RMSEs. This in turn does not distort inference on the mean for this model (since the mean is constant in DGP III, so any sample split based on inconsistent threshold still delivers valid mean estimates). However, estimation of the variance is distorted if regime-dependence in σ is ignored, and even when it is allowed in the second stage, estimates across the two regimes are poor, since the threshold is not precisely estimated.

Average RMSE					
Constant Parameter Model					
	γ	μ_1	μ_2	σ_1	σ_2
DGP I, n=200	-	0.0572	0.0572	0.0398	0.0398
DGP I, n=1000	-	0.0254	0.0254	0.0177	0.0177
DGP II, n=200	-	0.9975	1.0025	0.4116	0.4116
DGP II, n=1000	-	0.9992	1.0008	0.4138	0.4138
DGP III, n=200	-	0.0899	0.0899	0.5725	0.4275
DGP III, n=1000	-	0.0402	0.0402	0.5796	0.4204
DGP IV, n=200	-	1.0940	0.9060	0.9010	0.1149
DGP IV, n=1000	-	1.0968	0.9032	0.9058	0.0947
SSR Estimation, $\sigma_1 = \sigma_2$ in Stage 2					
DGP I, n=200	-	0.3784	0.3872	0.0817	0.0817
DGP I, n=1000	-	0.2944	0.3046	0.0356	0.0356
DGP II, n=200	0.0219	0.0817	0.0802	0.0799	0.0799
DGP II, n=1000	0.0090	0.0362	0.0360	0.0357	0.0357
DGP III, n=200	0.4284	0.0920	0.1612	1.4702	0.4826
DGP III, n=1000	0.3643	0.0405	0.0725	1.4945	0.4945
DGP IV, n=200	0.9388	0.1151	0.4360	2.1042	1.1043
DGP IV, n=1000	0.9864	0.0530	0.4404	2.1285	1.1285
SSR Estimation, $\sigma_1 \neq \sigma_2$ in Stage 2					
DGP I, n=200	-	0.3784	0.3872	0.1515	0.1513
DGP I, n=1000	-	0.2944	0.3046	0.1042	0.1068
DGP II, n=200	0.0219	0.0817	0.0802	0.0560	0.0567
DGP II, n=1000	0.0090	0.0362	0.0360	0.0251	0.0257
DGP III, n=200	0.4284	0.0920	0.1612	0.1478	0.1624
DGP III, n=1000	0.3643	0.0405	0.0725	0.1172	0.0962
DGP IV, n=200	0.9388	0.1151	0.4360	0.0853	0.0936
DGP IV, n=1000	0.9864	0.0530	0.4404	0.0407	0.0428
Likelihood-based Estimation					
DGP I, n=200	-	0.2878	0.2908	0.2670	0.2759
DGP I, n=1000	-	0.2340	0.2394	0.2245	0.2256
DGP II, n=200	0.0221	0.0817	0.0802	0.0567	0.0573
DGP II, n=1000	0.0091	0.0362	0.0360	0.0253	0.0258
DGP III, n=200	0.0740	0.0840	0.1656	0.0599	0.1167
DGP III, n=1000	0.0176	0.0366	0.0717	0.0253	0.0506
DGP IV, n=200	0.0385	0.0861	0.1519	0.0607	0.1084
DGP IV, n=1000	0.0146	0.0380	0.0680	0.0266	0.0479

Table 1: RMSE Results

On the other hand, the likelihood-based model performs well and delivers precise estimates for all parameters, and crucially, it correctly identifies the value of the threshold. In the case of DGP III, the constant parameter model performs satisfactory for the mean, since the data have constant mean; however, estimation of the variance is inconsistent, as expected.

Finally, in DGP IV, where both the mean and the variance are subject to regime-dependence, the constant parameter model is unsurprisingly inadequate. Moreover, the model based on SSR minimisation cannot consistently estimate the threshold, suggested by the increasing RMSE over the sample size. Estimation of both the mean and variance parameters is contaminated by the imprecise threshold estimates, and we see increasing estimation errors as we increase the sample. The likelihood-based approach, by modelling the second moment regime-dependence in the first stage can consistently and precisely estimate the threshold, and consequently, delivers precise and consistent estimates for the mean and the variance parameters in the second stage.

To summarise, this simple simulation exercise demonstrates clearly the advantages of our proposed approach. The standard approach based on SSR minimisation fails in the cases when there is regime-dependence in the variance (DGP III and IV), whether or not such regime dependence in the variance is modelled in the second stage. This is the case, since it does not use second moment information to identify the threshold parameter; and, consequently, conditional on the imprecisely estimated threshold parameter, delivers poor estimates for the mean and variance parameters in the second stage. This problem is not resolved when we increase the sample size, implying that the threshold parameter estimates cannot be recovered in these DGPs. On the other hand, the likelihood-based approach, by allowing additionally for variance regimes in the first stage estimation, performs very well whether or not there is regime-dependence in either the mean or variance parameters (DGPs I through IV) and delivers precise estimates for all parameters with estimation errors decreasing with the sample size.

4 Inflation and the Effects of Monetary Policy

We apply the SET-VAR methodology outlined in the previous section to U.S. data in order to study the effects of monetary policy at different inflation levels. Given the recent inflation experience not only in the U.S. but also across the world, an important question is whether policymakers' decisions have the same effect when inflation is around the 2 percent target as when inflation is much higher. The most widely used models in the literature to allow for possible structural changes in the evolution of the economy are time-varying parameter (TVP) VAR models. While TVP-VAR models are extremely flexible, an important drawback is that all of the model's parameters are allowed to change at every point in time. This lack of parsimony leads to two serious issues that hinder the practical implementation of TVP-VAR models: (i) they are subject to the curse of dimensionality, and so the widely used state space approaches (see, e.g. Cogley and Sargent (2005), Primiceri (2005)) come with large computational costs that grow quickly with the number of parameters and lags, and (ii) if the true parameters switch only across a finite number of macroeconomic regimes, allowing parameter changes at each period is unnecessary and can result in a loss of efficiency; for example, in the TVP-VAR setup, Petrova (2019) obtains a nonparametric consistency rate for the time-varying parameters while the SET-VAR approach obtains the standard parametric $\sqrt{n}-$consistency rate.

Alternative approaches that focus on a small number of distinct regimes (Hamilton, 1989; Sims and Zha, 2006) typically model the regimes as a function of an unobservable variable that follows a discrete Markov chain. Our choice to model the regime directly as a function of an observable variable facilitates a more transparent understanding of what drives the different regimes, allows one to focus on the specific nonlinearities characterized by our choice of state variable, and is computationally substantially less demanding.

Compared to both TVP- and Markov-switching VAR models, our approach provides parsimony, interpretability, and computational ease. Another class of models related to ours consists of smooth transition VARs, where VAR parameters are a convex combination of two sets of VAR parameters and the weights are governed by a smooth function of an observable variable (Auerbach and Gorodnichenko, 2012). Finally, Mavroeidis (2021) and Aruoba et al. (2022) develop VAR models with occasionally binding constraints that also speak to nonlinear dynamics of macroeconomic outcomes, albeit different nonlinearities from those we are focusing on.

4.1 Data, Priors, and Our State Variable

We use monthly U.S. data starting in January 1970 through December 2007 on the Federal funds rate (FFR), the unemployment rate, and inflation (computed as the year-on-year growth of the consumer price index). All data series are from the Federal Reserve Bank of St. Louis. In Appendix C we show that our findings are robust to including the BAA spread as an additional observable along the lines of Caldara and Herbst (2019). In addition, we use a proxy for the unobserved monetary policy shock to identify the effects of exogenous changes in monetary policy. In particular, we use the updated version of the Romer & Romer's monetary policy shock (Romer and Romer (2004); Wieland and Yang (2020)). We choose the Romer & Romer instrument because it allows us to use data from the 1970s and 1980s to infer about the effects of monetary policy shocks. Alternative instruments based on high-frequency changes in asset prices around monetary policy decisions (Gertler and Karadi, 2015) are generally only available for much shorter and more recent sample periods. Recent work by Aruoba and Drechsel (2022) uses machine learning techniques to incorporate textual data and nonlinearities to generalize the Romer & Romer approach, but unfortunately for our purposes, their sample does not contain the high inflation episodes of the 1970s. The downside of the Romer & Romer proxy is that the sample ends in 2007 because there is no unique way to extend the measure during periods where the effective lower bound on nominal interest rates binds. In Appendix D we present results for one possible extension that incorporates the effective lower bound - results are similar to our benchmark analysis presented here.[9]

In order to apply the SET-VAR approach, we require a suitable choice for the state variable that drives the regimes. Given that central banks consider inflation to be the relevant macroeconomic variable to target, and hence to determine monetary policy choices, we consider it to be the natural candidate. For measurability (so that the RHS of the model is not random given $\mathcal{F}_{t-1}$), we use inflation from the previous period; that is, we set $d = 1$. In the notation of Section 3, this means that $s_t = \pi_{t-1}$[10]. In theoretical macroeconomic models, a one-period lag of the inflation rate is often an important state variable.

For the estimation of the model, we use a specification with 12 lags and three regimes, and impose a flat prior on the threshold vector γ in the first estimation step. In the notation of Section 3, we set $\underline{\gamma}$ and $\overline{\gamma}$ to be the minimum and maximum observed values of the state variable in the sample, respectively. More details about the construction of the grid Γ can be found in Appendix A. For the VAR parameters, we use a loose Minnesota-style prior with overall shrinkage $\lambda = 1$ to ensure flexibility. Since the variables included in our SET-VAR do not exhibit a clear stochastic trend, we follow standard practice (Bańbura et al. (2010), Kilian and Lütkepohl (2017)) and center the coefficient on the first lag of each variable at zero. We further impose the condition that, in each regime, the companion form of the SET-VAR only has eigenvalues less than one in complex modulus. The prior for the Wishart parameters is set following the automatic rule in Kadiyala and Karlsson (1997). Importantly, we impose the same prior in all regimes; that is, our priors are not regime-specific, and hence, the estimated threshold $\hat{\gamma}$ is not directly affected by our choice of VAR priors. This is not a

[9] The extended sample ends in 2017, because there is no public access to more recent Tealbook forecasts prepared by Board staff.

[10] Since year-on-year inflation is a very persistent series, different values of the lag d deliver very similar results. We have performed robustness checks with respect to d; these additional results can be found in Figure A-2 in the Appendix. The Appendix also contains an additional set of empirical results with the addition of BAA spread to the model.

necessary feature of the methodology outlined, but rather a choice, in order to avoid imposing any prior beliefs about the different regimes ex ante.

With the above choice of state variable, lag order, and number of regimes, the SET-VAR model (3) becomes:

$$y_t = \sum_{i=1}^{3} \left(B_{0,i} + \sum_{j=1}^{12} B_{j,i} y_{t-j} + \Sigma_i^{1/2} \eta_t \right) \mathbb{I}(\gamma_{i-1} < \pi_{t-1} \leq \gamma_i). \tag{9}$$

We consider the model (9) and estimate the threshold parameter $\gamma = (\gamma_1, \gamma_2)'$ using our novel Bayesian approach, allowing, in addition, for Σ_i to differ across regimes, as explained in Section 3. This yields threshold estimates $\hat{\gamma} = (5.49, 11.02)$ and the resulting regimes are defined accordingly:[11]

- Low regime: $\pi_{t-1} \leq 5.49$ (74.3% of the sample);
- Medium regime: $5.49 < \pi_{t-1} \leq 11.02$ (19.6% of the sample);
- High regime: $\pi_{t-1} > 11.02$ (6.1% of the sample).

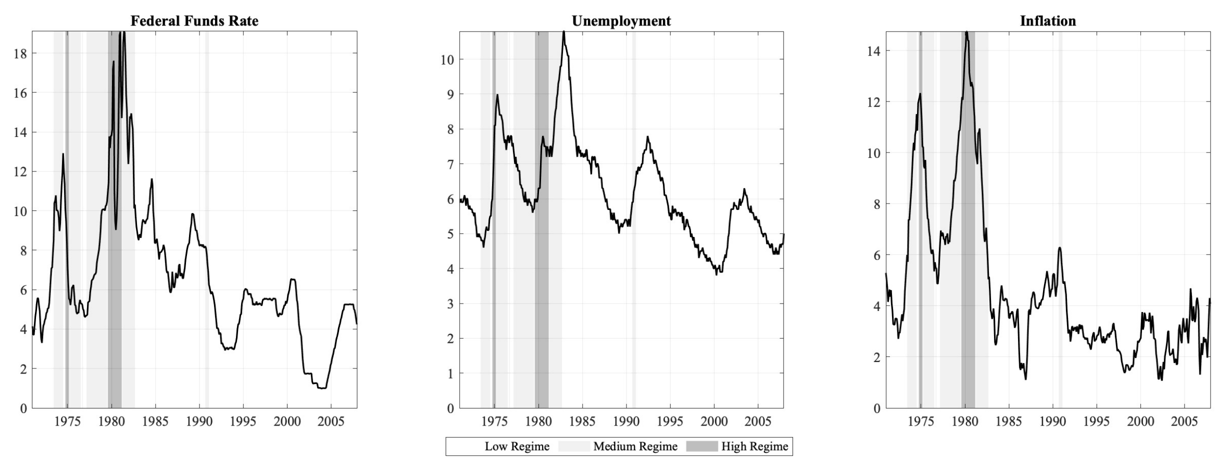

Figure 2: Macroeconomic data in our VAR. Light and dark grey areas denote the medium- and high-inflation regimes respectively.

Figure 2 displays our raw macroeconomic data against the estimated regimes. It is clear that the high regime (i.e. the regimes in place whenever inflation is higher than 11.02 percent) represents periods characterized by outliers, that is, observations that are not necessarily

[11]Figure A-1 in the Appendix displays the posterior objective function against the two-dimensional grid for γ, which we maximize to obtain the threshold estimates.

representative of the vast majority of the sample. Seventy-five percent of our sample is represented by what we label the low regime. We interpret the third regime, which almost exclusively covers the short era of non-borrowed reserves targeting by the Federal Reserve, as effectively removing outliers from our estimation. Only having two regimes, and thus not removing this outlier period, contaminates the estimates for the first two regimes and leads to drastically different results, as we show in Appendix E. This provides a strong justification for the inclusion of the third regime.

In Figure 3, we perform the following exercise: we estimate as an alternative a fully time-varying parameter (TVP) VAR model using the same lag order and priors, using the quasi-Bayesian kernel approach of Petrova (2019) and display the TVP model-implied long-run means and variances of the series against the identified regimes of our SET-VAR model.[12] As discussed, this TVP model is more flexible, since it allows the VAR parameters to change in each period, but on the downside, it is also a highly parameterized model with a slower nonparametric rate of convergence, and crucially, it can be unnecessarily overparameterized, especially if it is applied to a setup where the parameters are only a subject to a small number of regimes changes. It also makes it harder to interpret what drives the resulting nonlinearities. We conduct this exercise to investigate whether the estimated long-run means and variances of the flexible TVP model display regime-dependence along the lines of our identified regimes. As is clear from Figure 3, not only the unconditional means but also the unconditional variances of the series change with the SET-VAR-identified inflation regimes, providing a justification for our selection of lagged inflation as the state variable, as well as for our modeling choice to use regime-dependence in the second moments to identify the threshold. Furthermore, this figure provides some evidence that the assumption of three regimes is reasonable for our data set.

A central theme of our findings is that our approach endogenously filters out outliers or unusual time periods: those with unusually high inflation rates. There is always a tension about what to do with somewhat unusual periods in empirical analyses. On the one hand, behavior during those periods might not be representative of most of the sample, resulting in estimation bias; but on the other hand, periods of high volatility (which turn out to be periods of high inflation in our sample) help tightly pin down estimates if the underlying relationships are unchanged. The results in this paper lend credibility to the former view, while, at the same time, providing a data-driven methodology that can help researchers

[12] We define the unconditional or long-run moments as the moments associated with the parameters in place at each point in time, assuming no further parameter changes, as is common in the literature on time-varying parameter VARs (Cogley and Sargent, 2005; Primiceri, 2005).

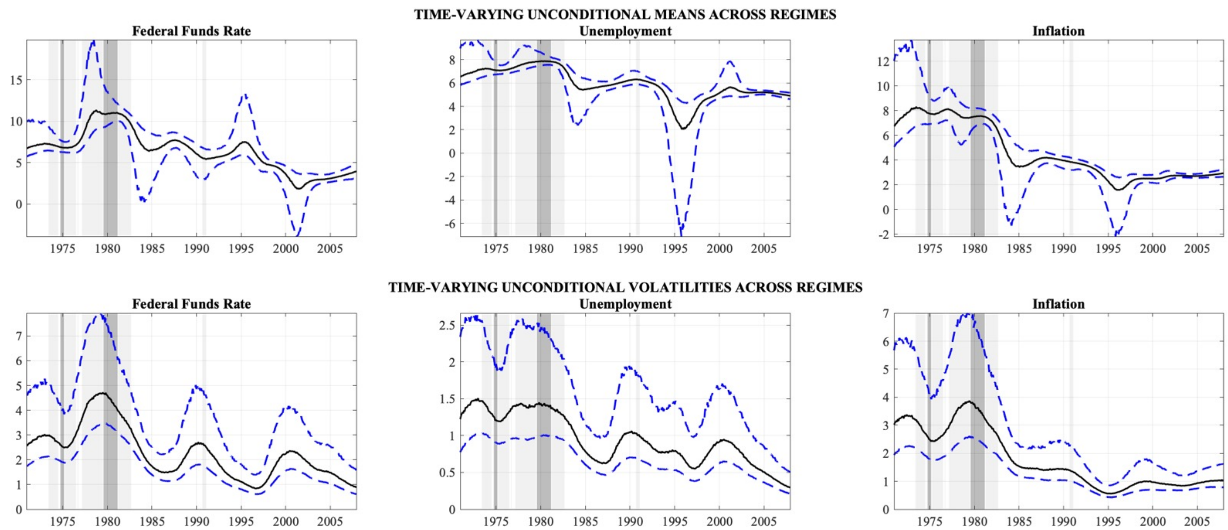

Figure 3: Time-varying LR means and variances against estimated regimes

decide whether outliers are contaminating results.

4.2 The Effects of Monetary Policy Shocks

In this section, we consider how the transmission and effectiveness of monetary policy can differ when the economy is in different inflation regimes. We estimate the structural SET-VAR model with the Romer & Romer instrument as a proxy for the policy shock. Formally, our identification strategy involves a Cholesky factorization of the regime-dependent covariance matrix, ordering the Romer & Romer proxy first in the vector of observables, following Plagborg-Møller and Wolf (2021). We normalize the Cholesky factor to have ones on the main diagonal; then the first column of the resulting matrix, associated with the policy instrument, yields the effect of the monetary shock on impact on all variables, up to scaling. Given that impulse responses are identified only up to scale in our setting, we further normalize the impact vectors to ensure better comparability across inflation regimes. Namely, we analyze a monetary policy shock that causes an immediate increase of 50 basis points in the FFR in every regime. We first look at what happens if we disregard any nonlinearities and estimate instead a linear Bayesian VAR (using the same priors as in the regime-dependent case). The posterior median and 68 percent posterior bands of the resulting impulse response functions to a 50 basis points monetary policy shock are displayed over a horizon of 90 months in Figure 4. From the figure, we find that there is no statistically meaningful movement in unemployment and inflation in response to a monetary policy shock if we focus on the linear model, a result that few policymakers would take at face value. Many papers in the VAR literature that study monetary policy shocks and their effects find similar inconclusive or even counter-intuitive evidence, which often also depends on the exact sample used (Bu et al., 2020; Ramey, 2016). One compelling explanation for such discrepancies in the empirical results could be that if the true underlying effects of policy shocks were in fact regime-dependent, then fitting a linear model on differing samples that could contain various proportions of each separate inflation regime would result to large variations in the empirical findings.

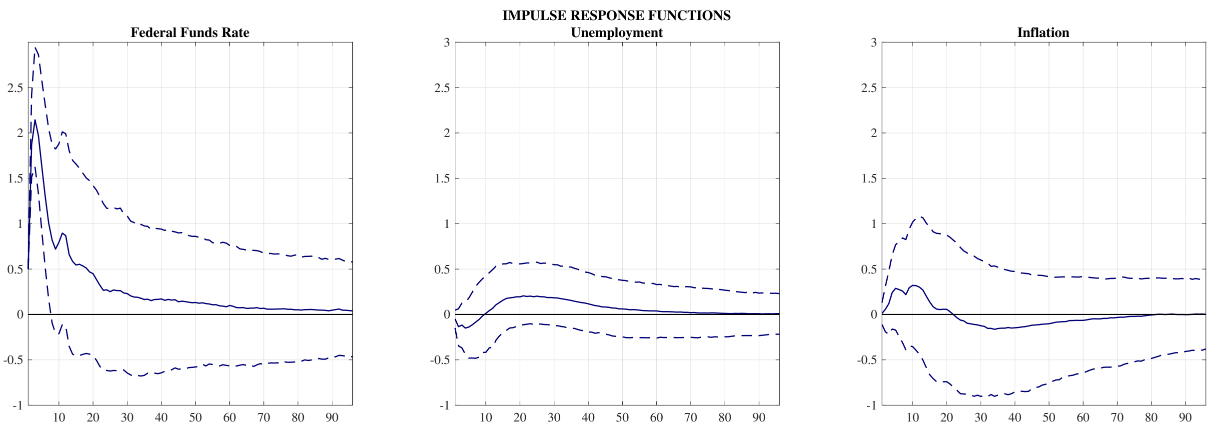

Figure 4: Fixed-coefficient impulse responses - posterior median and 68 percent posterior bands.

Next, we demonstrate how these problems can be mitigated by explicitly allowing for regime-dependence associated with the underlying level of inflation in the economy. While in the linear model, there is a unique notion of an IRF, in the nonlinear model, due to the nonlinearities present, there are multiple ways to define an IRF. In particular, we compute two types of IRFs. The first type is computed as in the linear model, *conditional* on a particular regime staying in place. This way of reporting results is common in the nonlinear VAR literature (Cogley and Sargent, 2005). The IRFs computed this way represent the responses to a policy shock, given that: (i) inflation at horizon 0 is within each regime (far from the threshold values), (ii) there are no other shocks, and (iii) the policy shock is sufficiently small, so that inflation remains in the same regime in all future horizons considered. Figure 5 presents the IRFs conditional on staying in each regime[13]. The left column displays the response when inflation in the last period is less than 5.5 percent. There

[13]The impulse responses of the Romer & Romer instrument can be found in Figure A-3.

is a persistent increase in nominal rates, but it is not associated with any significant increase in the unemployment rate. On the other hand, inflation falls after around two years, but the effect is relatively small and short-lived, with no impact after 40 months. When inflation is between 5.5 percent and 11 percent, we find a response of the short-term interest rates similar to that in the low-inflation case, but this interest rate change is associated with a substantial increase in the unemployment rate as well as a more pronounced, but also more delayed, fall in inflation. In fact, the initial increase in inflation, which is present, but not significant in the low-inflation regime, is now significant, but short-lived. In this medium-inflation regime, the responses of unemployment and inflation are much longer-lived and the monetary policy shock still has effects after 96 months (8 years). These long-lived responses are an outcome of the larger persistence of the series in this regime, which we discuss further in Section 3.3. The response of inflation can be explained by price-setters changing prices more quickly in high-inflation environments, a common result in the literature on state-dependent pricing (Golosov and Lucas, 2007). Finally, for completeness, the last column of Figure 5 displays the IRFs in the high-inflation regime, where we find that the response of the nominal interest rate is much shorter-lived and leads to a small, but significant, decrease in the unemployment rate and a small, but significant increase in inflation. Both of these counter-intuitive movements are short-lived. As mentioned above, our preferred interpretation is that the third regime effectively serves as an endogenous removal of periods characterised by unusual policy and outlier observations.

To summarize, once we allow for dependence on the level of inflation, we find that monetary policy has meaningful effects on inflation except when inflation is very high. Additionally, it also has substantial effects on labor market outcomes when inflation is not too low (higher than 5.5 percent). These results have important implications for policymakers, since the timeline and tradeoffs they face are different depending on the underlying level of inflation. The lack of impact on labor market outcomes when inflation is not high can, at least partially, explain the recent U.S. experience whereby inflation was brought down without a substantial increase in the unemployment rate (recall that this episode is not in our sample; so our model can produce "out-of-sample" conjectures for this period, since it has not been fitted to those observations). Importantly, not only do the effects of policy surprises differ on impact but they also have very different dynamics and transmission across regimes, with the total effects being much larger but also taking much longer to realize when inflation is in the medium regime, implying that after inflation increases beyond the threshold, the timing of policy surprises' impact on macroeconomic outcomes changes and policy effects

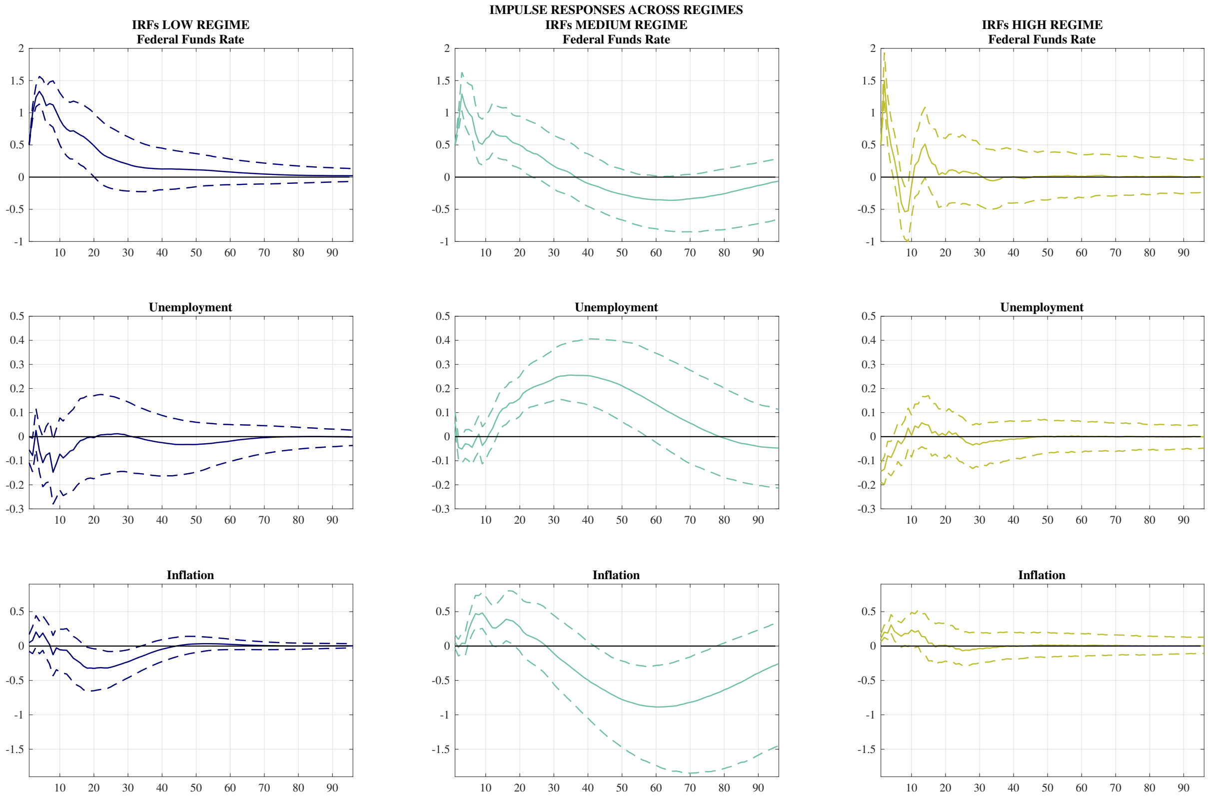

Figure 5: Regime-dependent impulse responses - posterior median and 68 percent posterior bands.

take considerably longer to realize.

In addition, we compute time-specific IRFs, which highlight further the nonlinearities of the model. In particular, different starting points, as well as the size and sign of the policy shock, could result in different responses, since all these three considerations may switch the model to different regimes at different horizons. We compute these period-specific IRFs to a δ that may embed the sign and size of the shock as the difference between the $h-$step ahead forecast implied by the model when there is a single policy shock at period t and no other shocks and the $h-$step ahead forecast in the absence of any shocks:

$$\begin{aligned} IRF_t^{\delta}(h) &= \mathbb{E}\left(y_{t+h}|\mathcal{F}_{t-1}, \eta_{1,t} = \delta, \eta_{j,t+h} = 0 \text{ for } \forall j \neq 1 \cup \forall h \geq 1\right) \\ &\quad -\mathbb{E}\left(y_{t+h}|\mathcal{F}_{t-1}, \eta_{j,t+h} = 0 \text{ for } \forall\ j \text{ and } h\right). \end{aligned}$$

The algorithm to compute these period-specific shocks can be found below.

Period-Specific IRF Algorithm

Step 1. Given the posterior of B and Σ in (8) and the chosen period of interest t, determine the regime i at t and set $i_t^{no\_shock} = i_t^{shock} = i$. Set $\hat{y}_{t-1:t-p}^{shock} = \hat{y}_{t-1:t-p}^{no\_shock} = y_{t-1:t-p}$.

For each posterior draw k for B^k and Σ^k iterate between Steps 2-4 below.

Step 2. For $h = 1$, given the initial period regime $i_t^{shock} = i_t^{no\_shock}$, determine the regime-specific coefficients and compute the one-step ahead projections:

$$
\begin{aligned}
\hat{y}_{t+1}^{shock} &= B_{0,i_t^{shock}}^k + \sum\nolimits_{j=1}^{p} B_{j,i_t^{shock}}^k \hat{y}_{t+1-j}^{shock} + \Sigma_{i_t^{shock}}^{k,1/2} [\delta, 0, ..., 0]' \\
\hat{y}_{t+1}^{no\_shock} &= B_{0,i_t^{no\_shock}}^k + \sum\nolimits_{j=1}^{p} B_{j,i_t^{no\_shock}}^k \hat{y}_{t+1-j}^{no\_shock},
\end{aligned}
$$

where δ contains the size and sign of shock (ordered first in our application) and $\Sigma_{i_t^{shock}}^{1/2}$ encodes our (regime-specific) identification scheme.

Step 3. For $h = 2, ..., H$, given the previous period regimes i_{t+h-1}^{shock} and $i_{t+h-1}^{no\_shock}$, determine the regime-specific coefficients and compute the projections:

$$
\begin{aligned}
\hat{y}_{t+h}^{shock} &= B_{0,i_{t+h-1}^{shock}}^k + \sum\nolimits_{j=1}^{p} B_{j,i_{t+h-1}^{shock}}^k \hat{y}_{t+h-j}^{shock} \\
\hat{y}_{t+h}^{no\_shock} &= B_{0,i_{t+h-1}^{no\_shock}}^k + \sum\nolimits_{j=1}^{p} B_{j,i_{t+h-1}^{no\_shock}}^k \hat{y}_{t+h-j}^{no\_shock},
\end{aligned}
$$

and determine the regimes i_{t+h}^{shock} and $i_{t+h}^{no\_shock}$ at $t + h$.

Step 4. Given $\hat{y}_{t+h}^{shock}$ and $\hat{y}_{t+h}^{no\_shock}$, compute and store the corresponding posterior draw for the IRF as $IRF_t^{\delta}(h) = \hat{y}_{t+h}^{shock} - \hat{y}_{t+h}^{no\_shock}$.

Figure 6 contain the IRFs for selected periods[14]. From Figure 6, it is clear that as the economy switches between regimes, the effects that a policy shock has can be very different. We find very similar responses in 1973 (and before) as in 1984 (upto 1991) and any period after 1995; since for all these periods, given that the shock is not large enough to cause a switch, the dynamics is governed by the coefficients in the low inflation regime. On the other hands, in periods such as 1978 and 1984, we have dynamics guided by the high inflation regime, and in addition, we have a large uncertainty around the responses, since switching is allowed from one horizon to another for each posterior draw. Finally, 1991 (as well as some periods in the late 1970s and early 1980s) is a period of medium regime inflation, and this results in unemployment and inflation responses that are considerably more persistent than

[14] We found no serious asymmetries between the responses of positive and negative, as well as big and small shocks, so we focus on the normalized 50 basis points responses. In the case of very large shocks, such asymmetries are expected to be exacerbated since large shocks have the power to switch regimes.

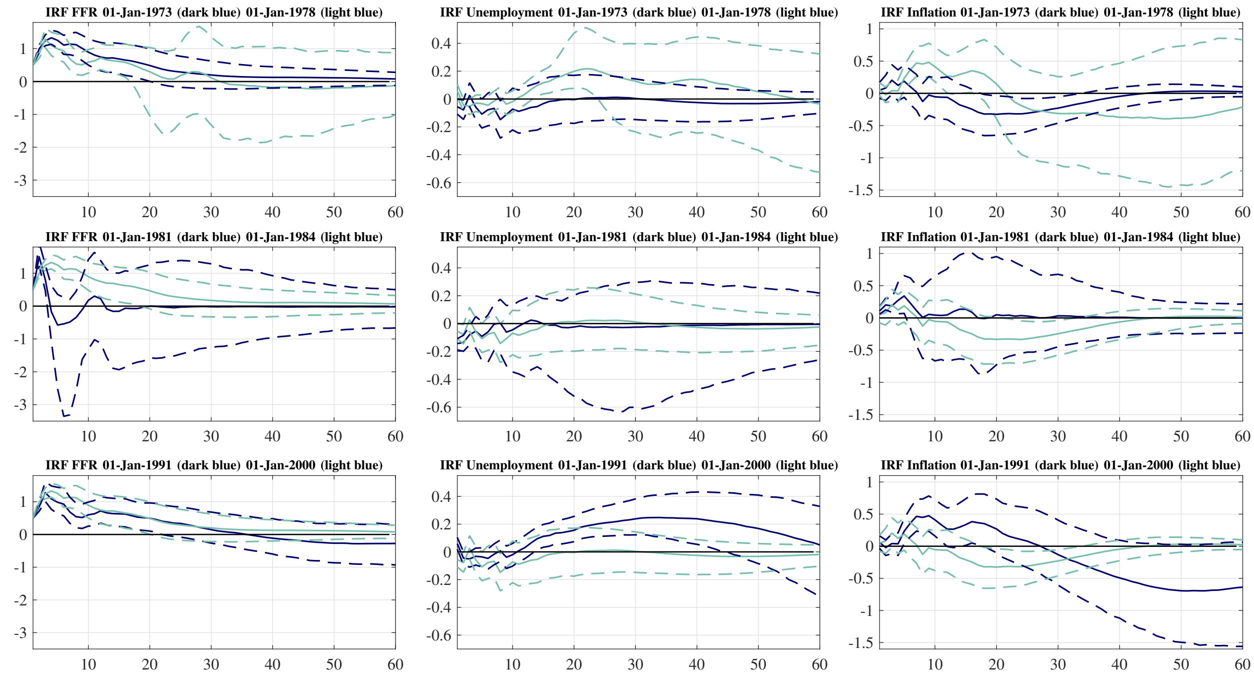

Figure 6: Impulse responses for selected periods - posterior median and 68 percent posterior bands.

responses from the low inflation regime that takes place after 1991.

4.3 How Different Are the Regimes?

We now provide some reduced-form analysis based on our estimated regime-dependent VAR model. This analysis is useful to assess whether there are further differences in the economic environment across regimes, but also to provide a background for the uncovered regime-dependence in the monetary policy effects in Section 4.2. Given the stationarity restriction that we impose in each regime, we can estimate regime-specific unconditional first and second moments. In this section, we focus on the correlation between unemployment and inflation and persistence of variables across regimes; we provide evidence for other first and second moments in Appendix G.

Figure 7 displays the posterior median and the posterior 25^{th} and 75^{th} percentiles for the correlation between inflation and unemployment across the three inflation regimes. While the long-run correlation between inflation and unemployment does not have a structural (Phillips curve slope) interpretation, it does measure the unconditional reduced-form relationship between the two, which can provide a summary of the inflation-unemployment relationship. From the figure, we find that this correlation in the low-inflation regime has a negative sign, as expected from a New Keynesian model, but it becomes considerably stronger as the economy moves into the medium-inflation regime. The stronger inverse relation in the medium regime is consistent with our finding of policy shocks having larger effects on unemployment in that regime. Finally, in the high regime, the correlation between inflation and unemployment switches sign and becomes positive, consistent with the periods of stagflation in that regime, implying that in high-inflation settings, inflation is actually harmful rather than beneficial to employment.

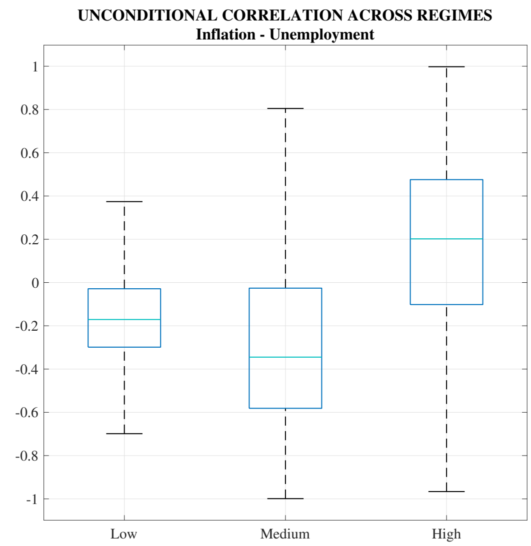

Figure 7: Unconditional correlations if each regime was in place indefinitely. Central lines in the boxes are the posterior medians, and the edges are the posterior 25^{th} and 75^{th} percentiles. The whiskers extend to most extreme data points not considered outliers.

Since we found different transmission of the policy shock, with some effects lasting considerably longer while others disappearing within a few months, we investigate whether variables have different persistence across inflation regimes (see Appendix G for details on how this measure is computed).s

Figure 8 presents the posterior medians and the posterior 25^{th} and 75^{th} percentiles for the persistence measure of inflation, unemployment, and FFR across the three regimes over horizons of 1, 6, 12, and 48 months. It is clear from the figure that there is considerable regime-dependence in the persistence of the series. The estimated persistence of inflation varies dramatically across regimes and is low when inflation is low. When the economy is in a medium inflation regime, the persistence of inflation is dramatically higher even at very distant horizons. This finding has important policy implications, since it is evident that inflation persistence depends on the level of inflation and that whenever inflation finds itself above the estimated threshold of around 5.5 percent, its dynamics slows, making inflation a more persistent process and hence considerably altering the lags required to achieve policy objectives. The same is true for unemployment, which also has implications for how effective policy is and how long it takes to achieve policy targets related to unemployment across regimes.

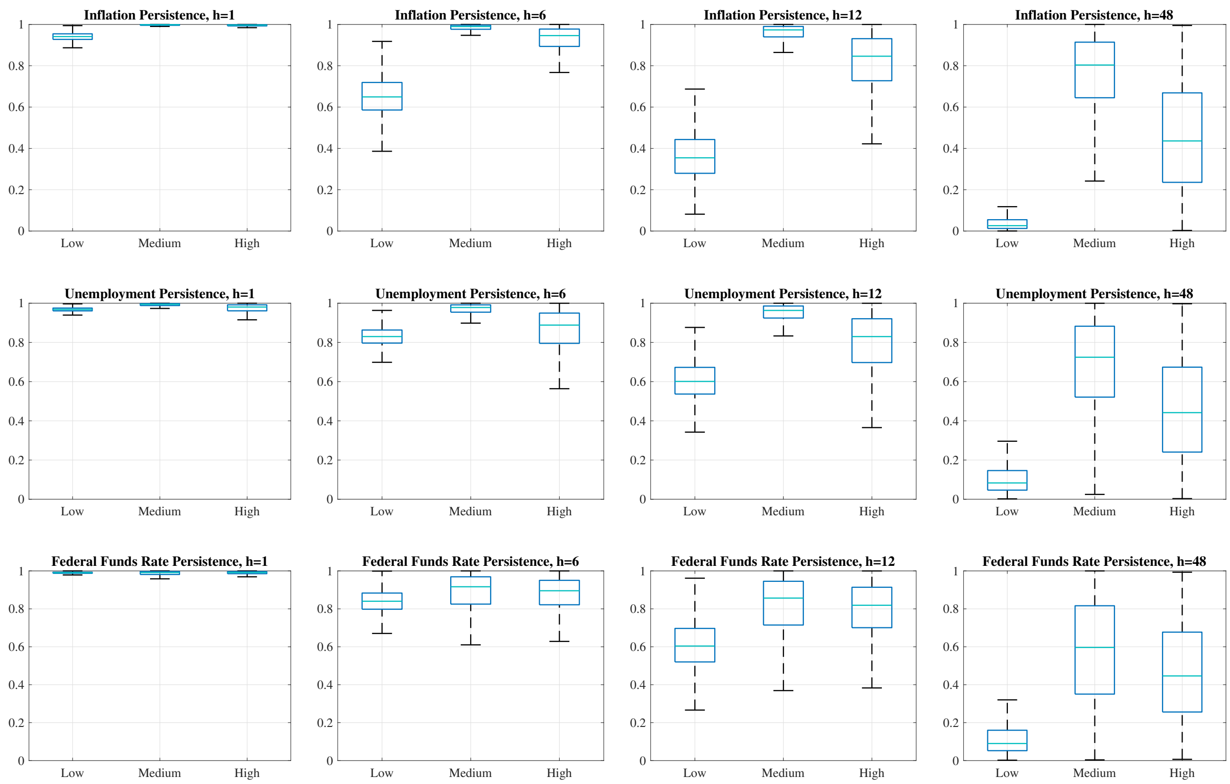

Figure 8: Persistence of our variables if each regime was in place indefinitely. Central lines in the boxes are the posterior medians, and the edges are the posterior 25$^{\text{th}}$ and 75$^{\text{th}}$ percentiles. The whiskers extend to most extreme data points not considered outliers.

5 Conclusions

In this paper, we build a self-exciting threshold Bayesian VAR model to investigate whether monetary policy depends on inflation levels. The econometric contribution of the paper is twofold. First, we allow for regime-dependence in the variance of the series for full likelihood-based identification of the threshold parameter, which is particularly relevant for macroeconomic series that have been documented to undergo volatility regimes over time. Second, we combine two-step frequentist estimation with Bayesian regularization using priors on the VAR parameters, resulting in a parsimonious nonlinear time-series model. This method simplifies and accelerates estimation compared to fully Bayesian threshold parameter treatments, while maintaining interpretability relative to Markov-switching models, where regimes are driven by unobserved latent processes.

Using our self-exciting threshold Bayesian VAR, we find that the effects of monetary policy vary substantially with the underlying level of inflation in the economy. For most of our post-WWII sample, inflation has been less than 5.5 percent in the U.S., which our model identifies as a period during which monetary policy has no meaningful effects on labor markets. Even though our sample ends in 2007 due to the availability of the instrument series used to identify the effects of monetary policy, our results are consistent with the recent "soft landing" of the U.S. economy. On the other hand, when inflation is between 5.5 and 11 percent, the effects of monetary policy are larger and longer-lasting, since variables are much more persistent in this regime, and the effects of monetary policy on unemployment are sizable and significant.

References

Alessandri, Piergiorgio, and Haroon Mumtaz. 2017. "Financial conditions and density forecasts for US output and inflation." *Review of Economic Dynamics* 24 66–78.

Alessandri, Piergiorgio, and Haroon Mumtaz. 2019. "Financial regimes and uncertainty shocks." *Journal of Monetary Economics* 101 31–46.

Aruoba, S Borağan, and Thomas Drechsel. 2022. *Identifying Monetary Policy Shocks: A Natural Language Approach.* Centre for Economic Policy Research.

Aruoba, S. Borağan, Marko Mlikota, Frank Schorfheide, and Sergio Villalvazo. 2022. "SVARs with occasionally-binding constraints." *Journal of Econometrics* 231 (2): 477–499, Special Issue: The Econometrics of Macroeconomic and Financial Data.

Ascari, Guido, and Timo Haber. 2021. "Non-Linearities, State-Dependent Prices and the Transmission Mechanism of Monetary Policy." *The Economic Journal* 132 (641): 37–57. 10.1093/ej/ueab049.

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

Bańbura, Marta, Domenico Giannone, and Lucrezia Reichlin. 2010. "Large Bayesian vector auto regressions." *Journal of applied Econometrics* 25 (1): 71–92.

Broemeling, Lyle D., and Peyton Cook. 1992. "Bayesian analysis of threshold autoregressions." *Communications in Statistics-Theory and Methods* 21 (9): 2459–2482.

Bu, Chunya, John Rogers, and Wenbin Wu. 2020. "Forward-Looking Monetary Policy and the Transmission of Conventional Monetary Policy Shocks." Finance and Economics Discussion Series 2020-014, Board of Governors of the Federal Reserve System (U.S.).

Caldara, Dario, and Edward Herbst. 2019. "Monetary policy, real activity, and credit spreads: Evidence from Bayesian proxy SVARs." *American Economic Journal: Macroeconomics* 11 (1): 157–192.

Canova, Fabio, and Fernando J. Pérez Forero. 2024. "Does the transmission of monetary policy shocks change when inflation is high?".

Chan, Kung-Sik. 1993. "Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model." *The Annals of Statistics* 520–533.

Chen, Cathy W.S., and Jack C. Lee. 1995. "Bayesian inference of threshold autoregressive models." *Journal of Time Aeries Analysis* 16 (5): 483–492.

Christiano, Lawrence J., Martin Eichenbaum, and Charles L. Evans. 1999. "Monetary policy shocks: What have we learned and to what end?" Volume 1. of Handbook of Macroeconomics 65–148, Elsevier.

Cogley, Timothy, Giorgio E. Primiceri, and Thomas J. Sargent. 2010. "Inflation-gap persistence in the US." *American Economic Journal: Macroeconomics* 2 (1): 43–69.

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

Gertler, Mark, and Peter Karadi. 2015. "Monetary policy surprises, credit costs, and economic activity." *American Economic Journal: Macroeconomics* 7 (1): 44–76.

Geweke, John, and Nobuhiko Terui. 1993. "Bayesian threshold autoregressive models for nonlinear time series." *Journal of Time Series Analysis* 14 (5): 441–454.

Giannone, Domenico, Michele Lenza, and Giorgio E. Primiceri. 2015. "Prior Selection for Vector Autoregressions." *The Review of Economics and Statistics* 97 (2): 436–451.

Golosov, Mikhail, and Robert E. Jr. Lucas. 2007. "Menu Costs and Phillips Curves." *Journal of Political Economy* 115 (2): 171–199.

Hamilton, James D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57 (2): 357–384, `http://www.jstor.org/stable/1912559`.

Hansen, Bruce. 1997. "Inference in TAR Models." *Studies in Nonlinear Dynamics and Econometrics* 2 (1): 1–16.

Kadiyala, K. Rao, and Sune Karlsson. 1997. "Numerical methods for estimation and inference in Bayesian VAR-models." *Journal of Applied Econometrics* 12 (2): 99–132.

Kilian, Lutz, and Helmut Lütkepohl. 2017. *Structural vector autoregressive analysis.* Cambridge University Press.

Mavroeidis, Sophocles. 2021. "Identification at the Zero Lower Bound." *Econometrica* 89 (6): 2855–2885.

Petrova, Katerina. 2019. "A quasi-Bayesian local likelihood approach to time varying parameter VAR models." *Journal of Econometrics* 212 (1): 286–306.

Petrova, Katerina. 2022. "Asymptotically valid Bayesian inference in the presence of distributional misspecification in VAR models." *Journal of Econometrics* 230 (1): 154–182.

Plagborg-Møller, Mikkel, and Christian K. Wolf. 2021. "Local projections and VARs estimate the same impulse responses." *Econometrica* 89 (2): 955–980.

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

Ramey, V.A. 2016. "Macroeconomic Shocks and Their Propagation." In *Handbook of Macroeconomics*, edited by Taylor, J. B., and Harald Uhlig Volume 2. of Handbook of Macroeconomics, Chap. 0 71–162, Elsevier.

Romer, Christina D., and David H. Romer. 2004. "A new measure of monetary shocks: Derivation and implications." *American Economic Review* 94 (4): 1055–1084.

Samia, Noelle I., and Kung-Sik Chan. 2011. "Maximum likelihood estimation of a generalized threshold stochastic regression model." *Biometrika* 98 (2): 433–448.

Sims, Christopher A. 2003. "Implications of rational inattention." *Journal of Monetary Economics* 50 (3): 665–690, Swiss National Bank/Study Center Gerzensee Conference on Monetary Policy under Incomplete Information.

Sims, Christopher A., and Tao Zha. 2006. "Were There Regime Switches in U.S. Monetary Policy?." *American Economic Review* 96 (1): 54–81.

Tong, Howell. 1977. "Discussion of a paper by A.J. Lawrance and N.T. Kottegoda." *Journal of the Royal Statistical Society: Series A* 140 (1): 34–35.

Tong, Howell. 2011. "Threshold models in time series analysis—30 years on." *Statistics and its Interface* 4 (2): 107–118.

Tong, Howell, and Keng S. Lim. 1980. "Threshold autoregression, limit cycles and cyclical data." *Journal of the Royal Statistical Society: Series B (Methodological)* 42 (3): 245–268.

Tsay, Ruey S. 1998. "Testing and modeling multivariate threshold models." *Journal of the American Statistical Association* 93 (443): 1188–1202.

Weber, Michael, Bernardo Candia, Tiziano Ropele et al. 2023. "Tell me something I don't already know: Learning in low and high-inflation settings." *NBER Working Paper* 31485.

Wieland, Johannes F, and Mu-Jeung Yang. 2020. "Financial dampening." *Journal of Money, Credit and Banking* 52 (1): 79–113.

Wu, Jing Cynthia, and Fan Dora Xia. 2016. "Measuring the macroeconomic impact of monetary policy at the zero lower bound." *Journal of Money, Credit and Banking* 48 (2-3): 253–291.

Online Appendix for
"Monetary Policy across Inflation Regimes"

A Grid Construction

For the estimation of the threshold vector γ, we consider a fine grid of N_γ equidistant points $\Gamma = (\underline{\gamma}, \ldots, \overline{\gamma})$ for all the elements in γ, and perform a grid search over Γ^{k-1}, which denotes the $(k-1)$ Cartesian product of Γ; in our case $k = 3$ and $\gamma = [\gamma_1,\ \gamma_2]'$. In particular, to consider all of the possible observed values of the state variable, we fix $\underline{\gamma} = \min\{s_t\}$ and $\overline{\gamma} = \max\{s_t\}$. In the sample we analyze, $\min\{s_t\} = 1.07$ and $\max\{s_t\} = 14.76$. The number of grid points N_γ is chosen to accommodate the trade-off between fineness of the grid and computational efficiency. For our application, we choose $N_\gamma = 500$, which implies increments in the state variable of approximately 0.03 across grid points. The posterior evaluation is performed imposing the condition that: (i) $\gamma_2 > \gamma_1$ always, and (ii) the distance across the points over which we compute the posterior is constant. The grid search works as follows:

1. For a grid point g_i in Γ, set $\gamma_1 = g_i$;

2. For each grid point $\gamma_2 = \{g_i, g_{i+1}, \ldots, \overline{\gamma}\}$, evaluate the posterior (equation 6);

3. Set $g_i = g_{i+1}$ and repeat 1-2 until $\gamma_1 = \gamma_2 = \overline{\gamma}$;

4. Search for the maximum value of the posterior over the two-dimensional grid Γ^{k-1}. The corresponding value of the threshold vector is $\hat{\gamma}$.

Figure A-1 displays the posterior objective function against the two-dimensional grid.

B Additional Results

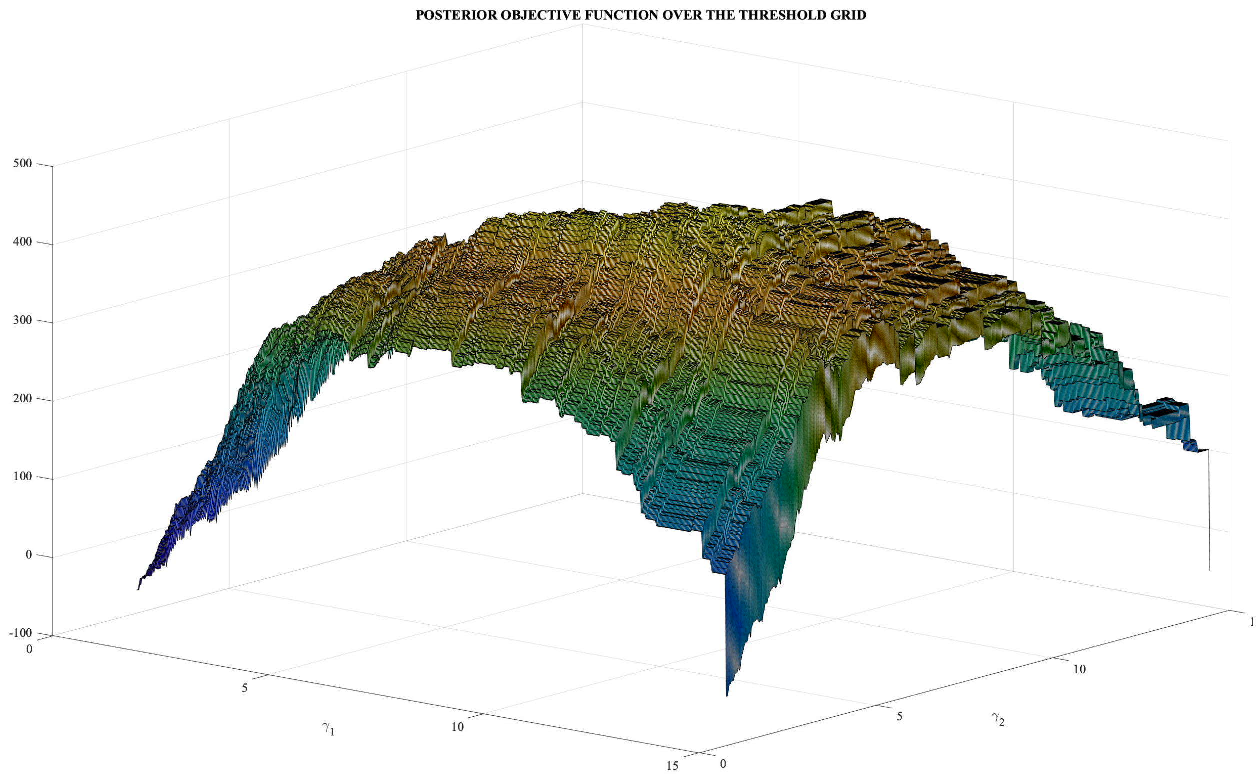

Figure A-1: Objective function

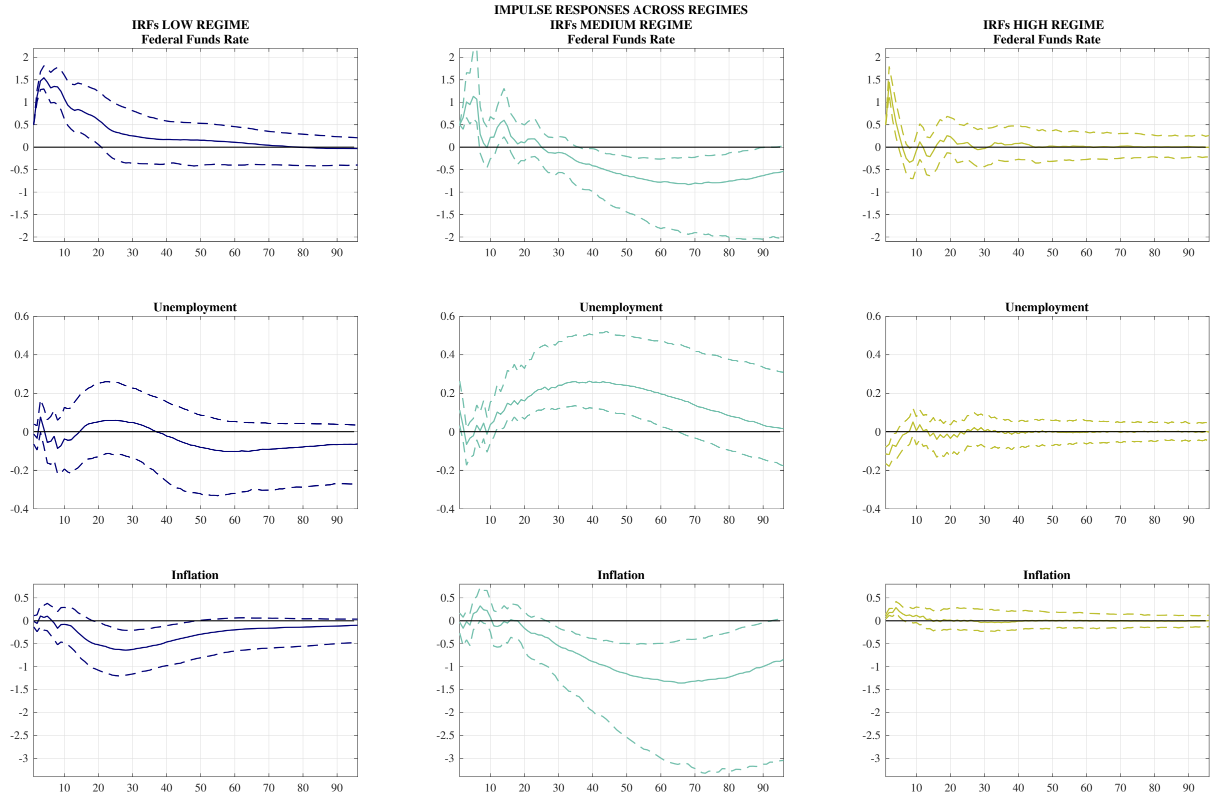

Figure A-2: IRFs with $s_t = \pi_{t-6}$

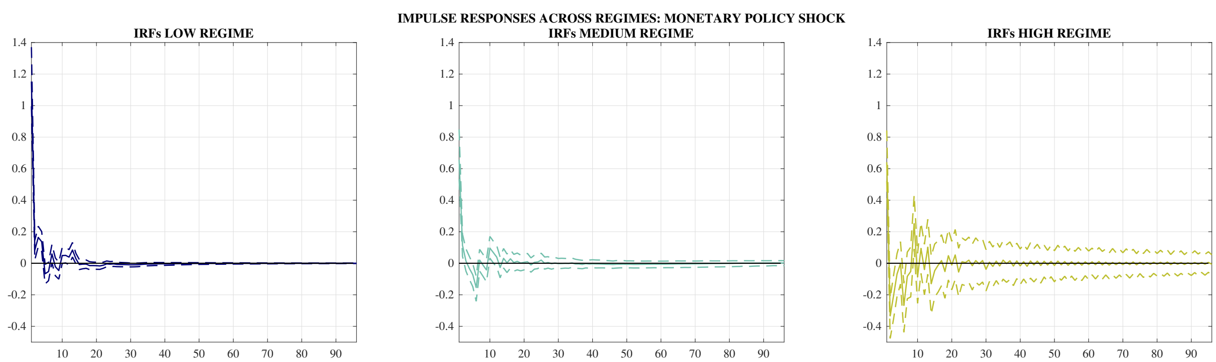

Figure A-3: IRFs of the proxy

C Model with BAA Spread

As an additional exercise, we estimate a larger SET-VAR(12) with three regimes that also includes a measure of financial conditions. In particular, we follow Caldara and Herbst (2019) and use the BAA spread. This is the Moody's seasoned BAA corporate bond yield relative to the yield on ten-year treasury constant maturity. Using the same state variable and methodology as in the main specification of the paper, we find the threshold estimates to be $\hat{\gamma} = (5.26, 11.12)$, which are in line with our main results. The corresponding IRFs can be found in Figure A-5.

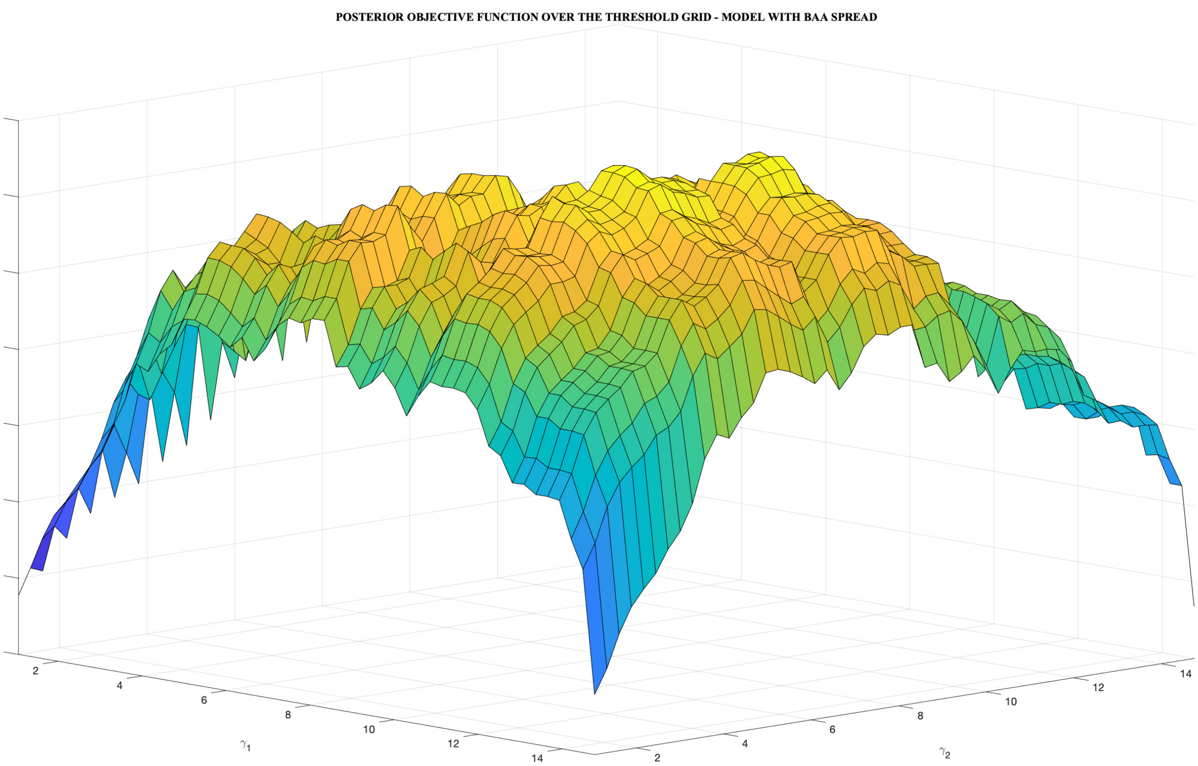

Figure A-4: Objective function

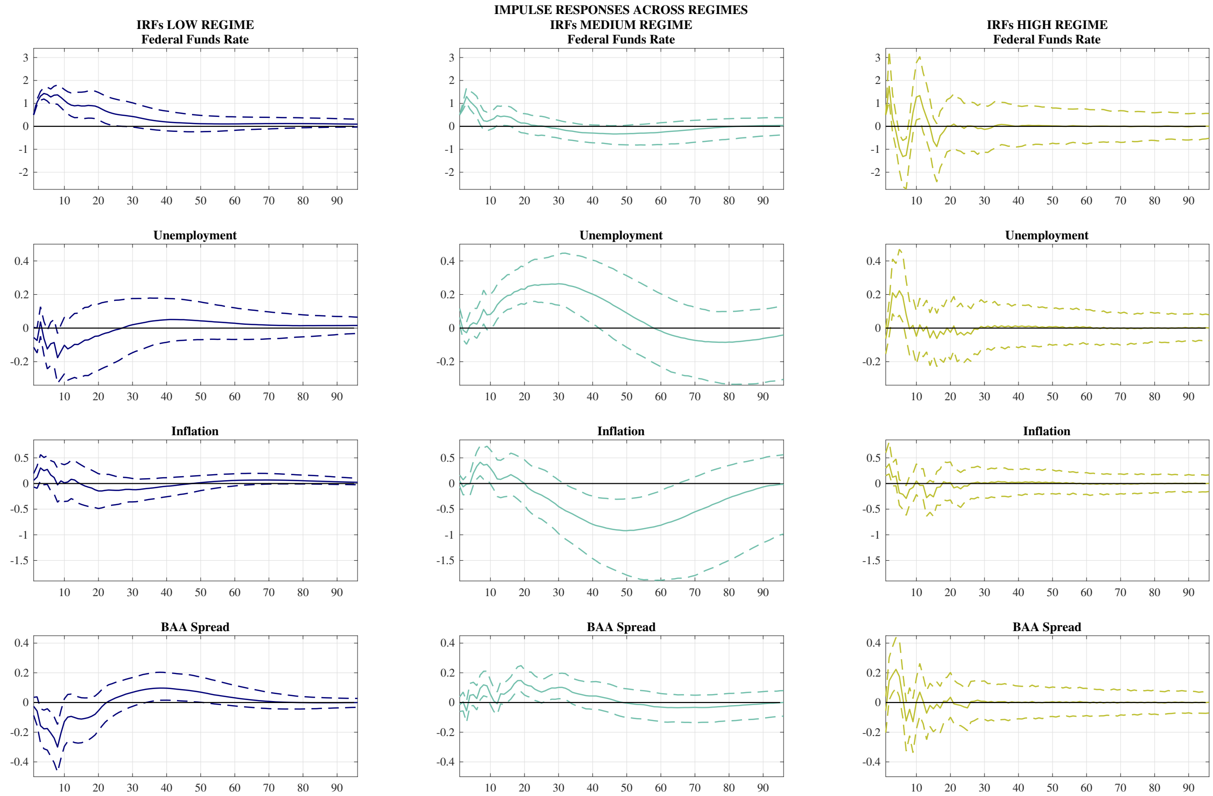

Figure A-5: Regime-dependent impulse responses - posterior median and 68 percent posterior bands.

D Model Using Data Up to 2017

We also estimate a SET-VAR(12) with three regimes including data up to December 2017. To account for the zero-lower bound (ZLB, or effective lower bound) period in this sample, we use a measure of the shadow interest rate constructed by Wu and Xia (2016) instead of the federal funds rate when available. This series is available starting in 1990 and updated to 2023. For the years prior to 1990, we use the federal funds rate as in the main specification in the paper. Due to the Romer & Romer proxy series ending in 2007, we employ a similar measure provided by Silvia Miranda-Agrippino. Then, using the same state variable and methodology as in the paper, we find the threshold estimates for this longer sample to be $\hat{\gamma} = (5.95, 11.03)$, very close to those in the paper. The resulting IRFs can be found in Figure A-7.

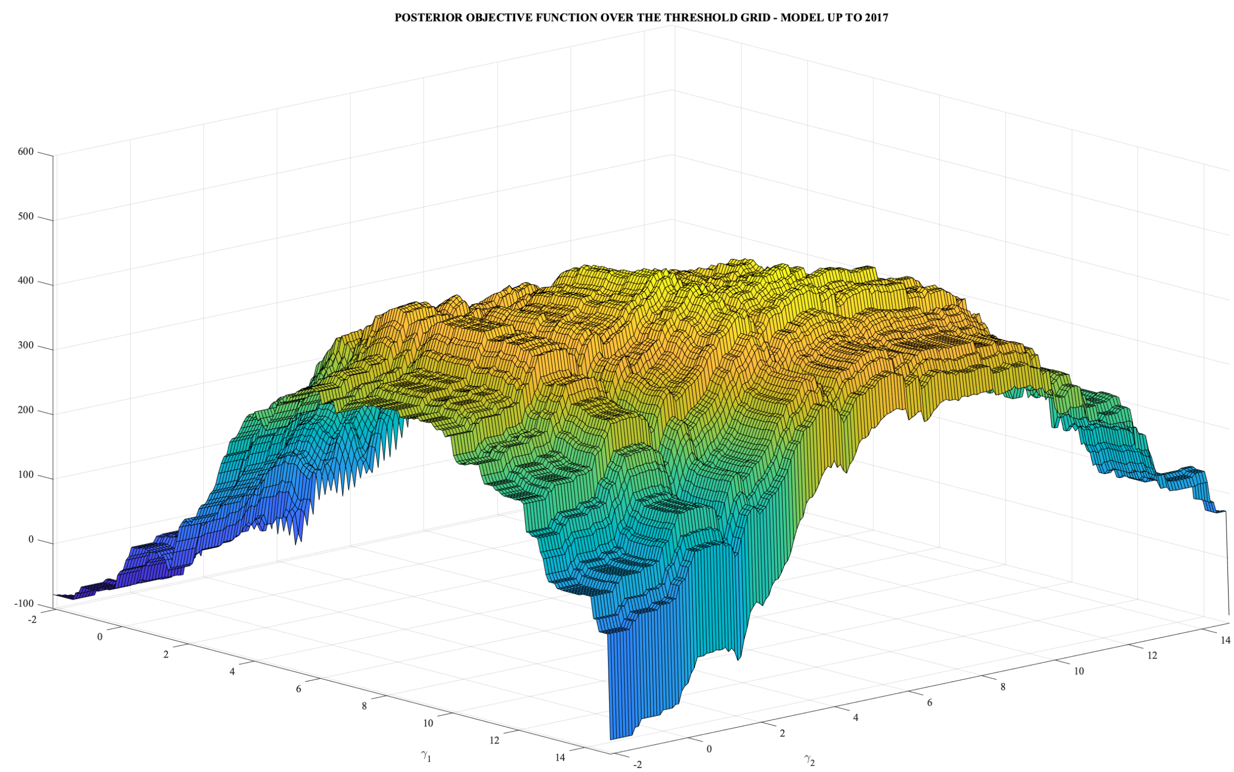

Figure A-6: Objective function

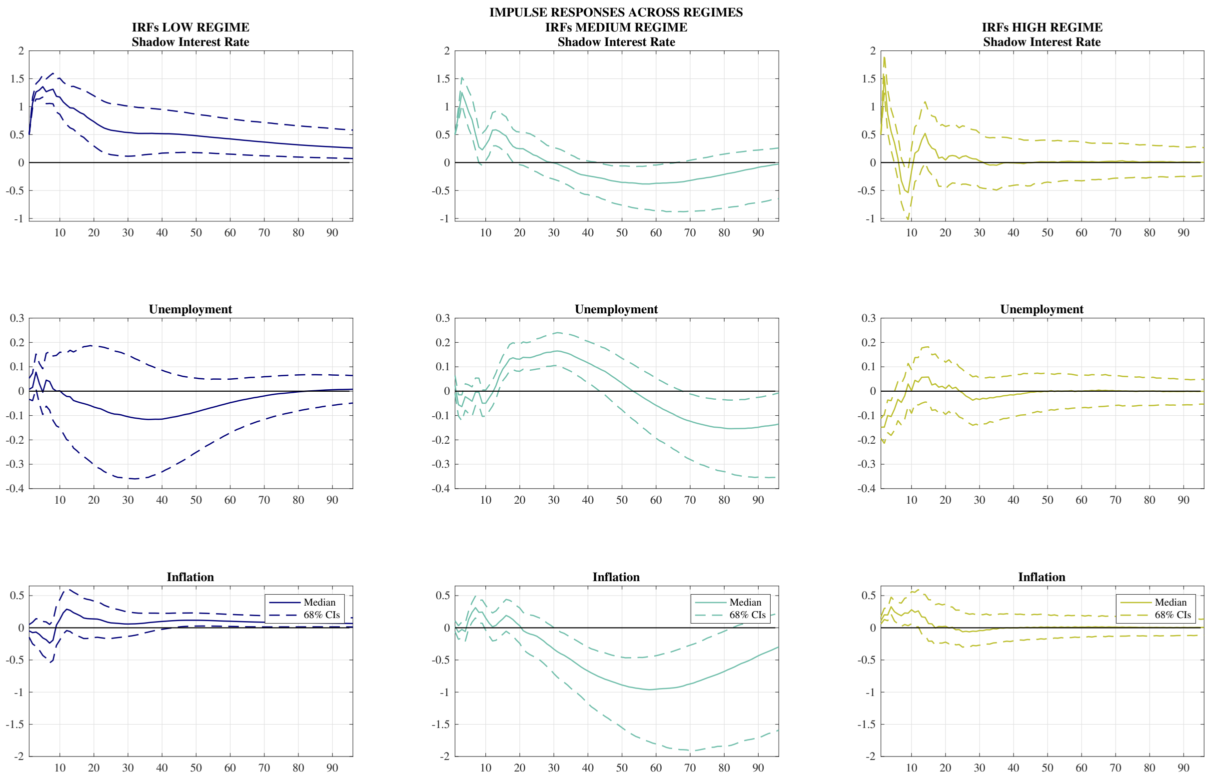

Figure A-7: Regime-dependent impulse responses - posterior median and 68 percent posterior bands.

E Model Imposing Two Regimes

To understand the value of the third regime, we estimate a SET-VAR(12) with two regimes, using the same data, state variable and priors as before. The threshold is now considerably different $\hat{\gamma} = 8.26$, in-between the two values we find when including a third regime. The associated IRFs are plotted in Figure A-8 and are also different. The implied IRFs for the high regime look like a mix between the medium and high regimes of the main specification while the IRFs for the low regime look like a mix of the low and medium regimes, which is undesirably since the unusual periods of the third regime contaminate the estimates of the model with two regimes; we use this as a justification for the inclusion of a third regime.

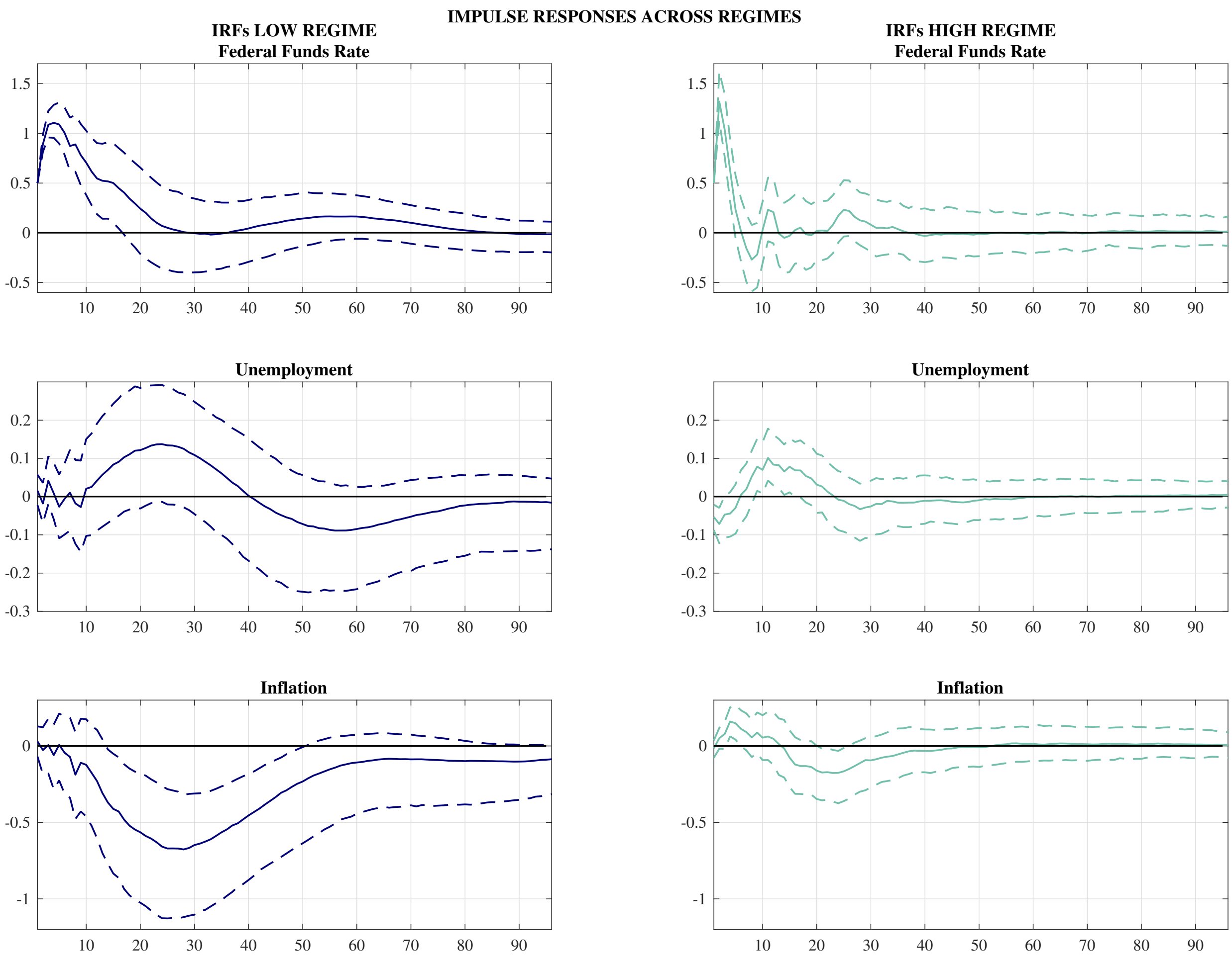

Figure A-8: Regime-dependent impulse responses - posterior median and 68 percent posterior bands.

F SSR Based Threshold Estimation - Constant Variance

To address the relevance of incorporating regime-dependence in the variance of the model for the threshold estimation, we perform an additional exercise in which we follow the standard approach in the literature for estimating our model. That is, we estimate the thresholds from the SSR objective function as the first estimation step, and then estimate the VAR imposing constant variance across regimes for the second step. Following the same approach of the main specification, we evaluate the SSR at the posterior modes for the regime-dependent β_i's over the threshold grid and look for its minimum. The implied thresholds in this case are $\hat{\gamma} = (6.97, 11.28)$, the first being quite different from what we find using our proposed estimator. The associated IRFs are plotted in Figure A-9.

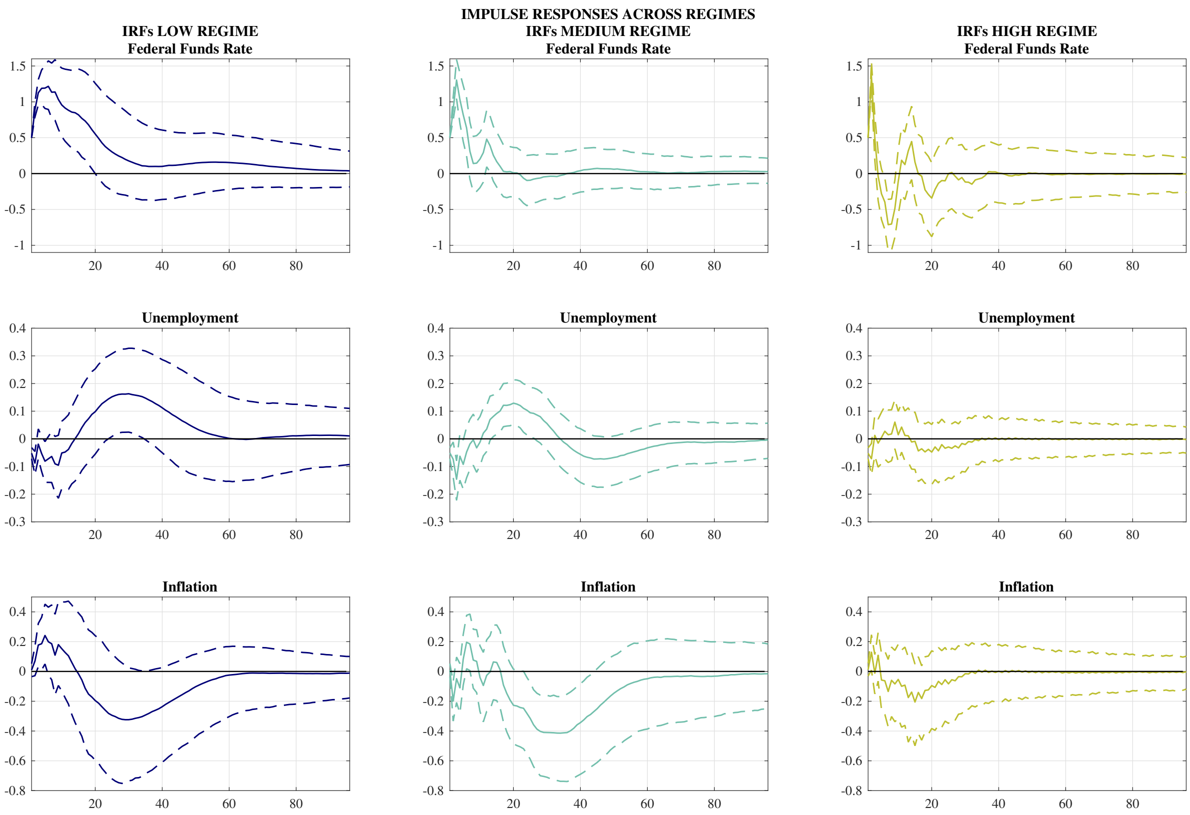

Figure A-9: Regime-dependent impulse responses - posterior median and 68 percent posterior bands.

The difference we observe in the IRFs could be not just due to the constant variance, but also from the different estimated thresholds. To further investigate the implications of the former, we also show in figure A-10 the regime-specific unconditional correlations, as we do for the main specification. Since the model now has constant covariance matrix and this matters for the computation of the correlations, one of our main results on the trade-off between inflation and unemployment changes. We use this as a justification for the likelihood-based model with regime-dependence in the covariance.

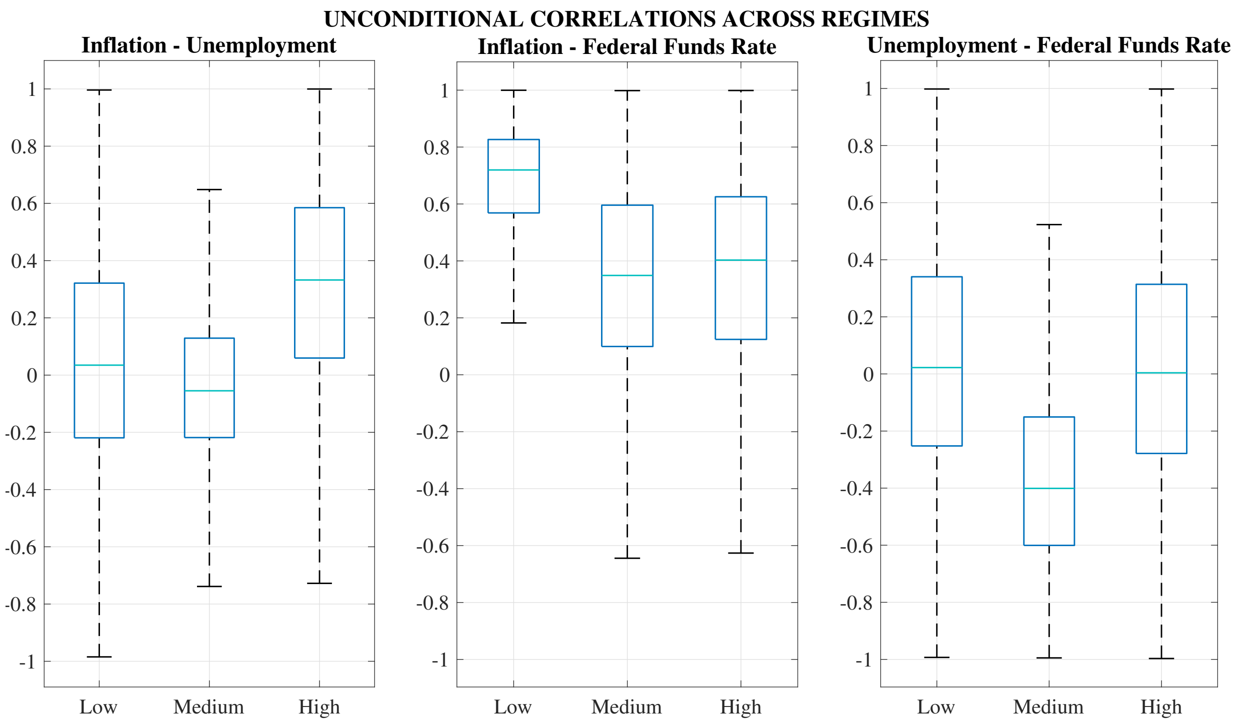

Figure A-10: Unconditional correlations if each regime was in place indefinitely. Central lines in the boxes are the posterior medians, and the edges are the posterior 25^{th} and 75^{th} percentiles. The whiskers extend to most extreme data points not considered outliers.

G Further Evidence on Differences Across Regimes

For the results presented in this section, we work with the companion form of model (9) :

$$z_t = \sum_{i=1}^{k} (\mu_i + A_i z_{t-1} + \varepsilon_{i,t}) \mathbb{I}(\hat{\gamma}_{i-1} < \pi_{t-1} \leq \hat{\gamma}_i), \tag{A-1}$$

where

$$z_t := \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p-1} \end{bmatrix}, \quad \mu_i := \begin{bmatrix} B_{0,i} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad A_i := \begin{bmatrix} B_{1,i} & B_{2,i} & \dots & B_{p,i} \\ I_M & 0 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & & I_M & 0 \end{bmatrix}, \quad \varepsilon_{i,t} := \begin{bmatrix} \Sigma_i^{1/2}\eta_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Given the stability condition $\max_{1\leq i\leq k} \rho\left(A_i\right) < 1$ where $\rho\left(.\right)$ denotes the spectral radius, we can compute the implied regime-specific unconditional means τ_i and unconditional variances U_i for each variable from the vector MA(∞) representation (assuming each regime remains in place indefinitely) as

$$\tau_i = (I - A_i)^{-1}\mu_i$$

$$U_i = \sum\nolimits_{j=0}^{\infty} A_i^j \Omega_i (A_i^j)',$$

where $\Omega_i = \mathbb{E}[\varepsilon_{i,t}\varepsilon_{i,t}']$. Figure A-11 displays the estimated posterior densities for τ_i and U_i for each regime.

The results in Figure A-11 confirm the existence of differences across regimes. This is evident not only when looking at the long-run trends but also when looking at the volatility of the series, which is a further justification for our modeling choice to allow for the threshold estimation to also depend on the second moment of the variables included in the model.

To investigate the persistence across regimes, we compute the regime-specific persistence h steps ahead for each variable, using the measure proposed by Cogley et al. (2010):

$$R^2_{i,h,k} = 1 - \frac{e_k'(\sum_{j=0}^{h-1} A_i^j \Omega_i (A_i^j)')e_k}{e_k'(\sum_{j=0}^{\infty} A_i^j \Omega_i (A_i^j)')e_k} \tag{A-2}$$

where e_k is the k^{th} standard basis vector for $\mathbb{R}^{Mp}$, which selects the k^{th} variable in the system. This measure accounts for the fraction of the total variance of each variable explained by past shocks at different horizons. It takes values between 0 and 1, with numbers closer to 1 indicating higher model-implied autocorrelation for the variable, suggesting that its dependence on past shocks dies out more slowly, and, hence, implying a more persistent variable.

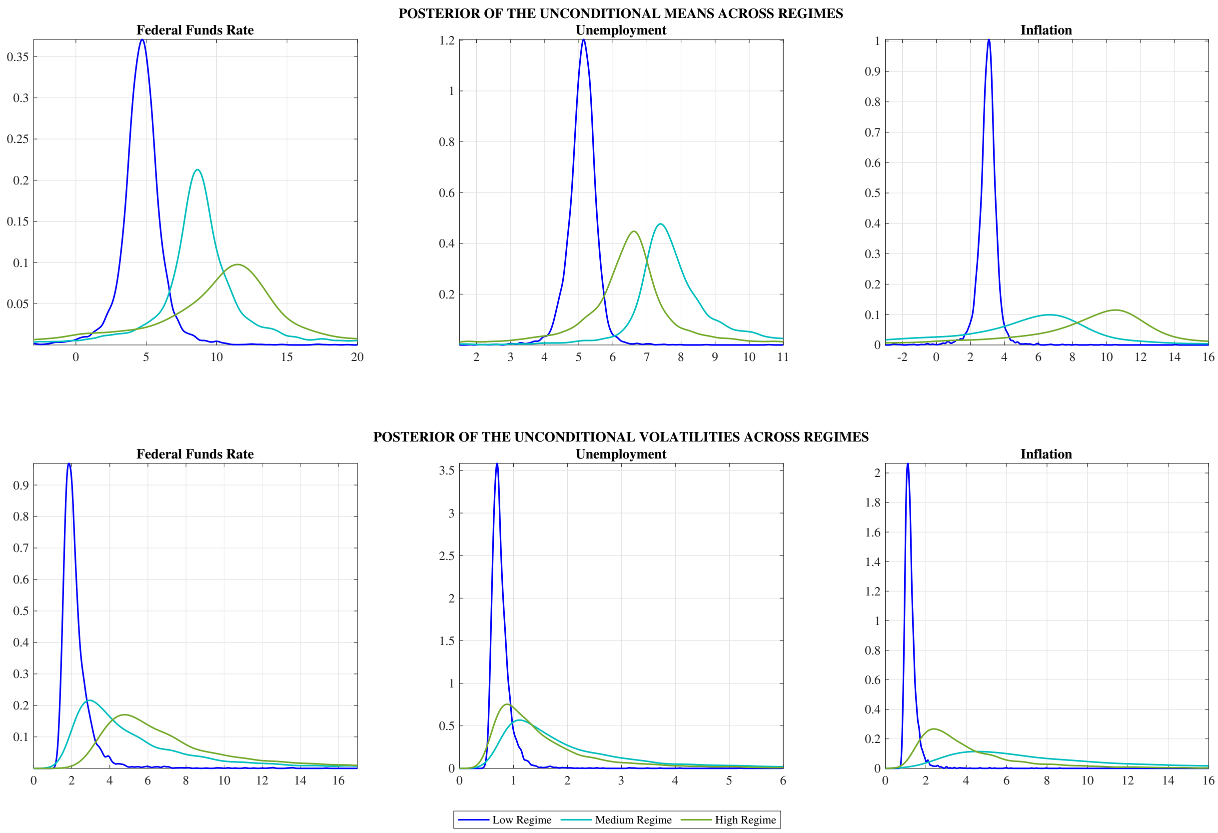

Figure A-11: Posterior distribution of unconditional means and standard deviations if each regime was in place indefinitely.