Uncertainty and Effectiveness of Monetary Policy: A Bayesian Markov Switching-VAR Analysis

Zulquar Nain, Bandi Kamaiah (2020)

A Replication by Ricardo Semião e Castro

2025-04-28

Table of contents

1 Introduction

Understanding how uncertainty affects monetary policy effectiveness is crucial as it directly impacts economic stability. The literature recognizes that monetary policy operates through various channels whose effectiveness varies with economic structure, meaning developed and developing countries experience different outcomes. Additionally, effectiveness changes over time, potentially due to varying uncertainty levels. Evidence largely suggests that increased uncertainty diminishes monetary policy effectiveness, though most research focuses on developed economies rather than developing ones.

The paper (Nain and Kamaiah 2020) makes an important contribution by employing a Markov Switching VAR model to endogenously identify regimes of high and low uncertainty in India's economy, filling a gap in literature that has predominantly focused on developed countries.

While their methodology, which includes the use of regime-dependent impulse response functions, is sound, some aspects could be improved. These include insufficient detail on the Gibbs sampling procedure (particularly regarding priors and conditional distributions), lack of posterior distribution diagnostics, limited sensitivity analysis for variable proxies, and minimal visual representation of data properties. In my replication, I will maintain fidelity to their approach while addressing these gaps, with particular focus on elaborating the Bayesian methodology and diagnostic procedures.

The code, data, and typesetting scripts can be found in my GitHub.

This section is about interpreting and commenting the work, which can be less relevant, although it motivates my changes to the original paper. Section 2 is about the methodology, where lies my main contribution of a more described sampler. Section 3 replicates the data analysis, which I value, but is less relevant in the econometric viewpoint. Section 4 shows the sampler results. Lastly, I conclude in section 5.

1.1 Overview of the Paper

1.1.1 Motivation and Literature Review

The authors start by describing the several channels through which monetary policy affects the economy, noting how the differences in their effectiveness are not yet fully understood, but how it is affected by the economic structure. This implies that developed countries will have different realities than developing ones. Furthermore, they cite (Sengupta 2014), who argues that the effectiveness of monetary policy also changes with time, which could be associated with changes in different uncertainties.

Additionally, the concept of uncertainty is discussed, being associated with forecast inability and risk, two hard-to-differentiate components. This is important to choose what uncertainty measure to use, with volatility being the preferred one.

The authors cite other papers, citations from monetary authorities, and stylized facts, to establish a consensus that more uncertainty diminishes the effectiveness of monetary policy, and hypotheses of the how.

Finally, they motivate their work by citing that much of this evidence is focused on developed countries, and that they will fill the hole by studying India. Additionally, the empirical literature uses VAR models with different measures of uncertainty, but few define endogenous regimes of high and low uncertainty, which the authors will do via the Markov Switching model.

1.1.2 Methodology

The authors defend the use of a Markov Switching VAR model, as its non-linear nature allows for capturing dynamic properties (time-varying cycles, breaks/jumps, etc.). The description of the model is in section Section 2.1.

For its estimation, the authors defend the use of Bayesian MCMC, as it is more flexible and provides full posterior distributions of all parameters, which is important to build the impulse response functions (IRFs). The authors follow the Gibbs sampling procedure outlined in (Balcilar et al. 2017), described in section Section 2.2.

Regime-dependent IRFs are calculated following (Ehrmann, Ellison, and Valla 2003), which assume regime switching does not occur beyond the shock horizon. Identification needs further assumptions, and the authors use the Cholesky decomposition in the order of OG, INF, and

IR (see data section), with the justification that real variables respond with a time lag to the monetary policy shocks which in turn respond contemporaneously to the real sector shocks.

1.1.3 Data and Data Analysis

The dataset used includes three variables from the EPW time-series database: Index of Industrial Production (IIP), Wholesale Price Index (WPI), and Weighted Call Money Rate (WCMR)[1]. All are with monthly frequency for the period of April 1991 to December 2016, including only the period after a structural reform and financial liberalization of the 90s.

IIP and WPI were seasonally adjusted via X-13 ARIMA-SEATS of U.S. Census Bureau. Additionally, they were transformed into growth rates (in percentage terms) to be used as the proxies for output growth (OG) and inflation (INF), respectively.

The authors discuss the changes in monetary policy in India, which ultimately led to the choice of Weighted Call Money Rate (WCMR) as the indicator of the policy variable.

The authors present summary statistics and tests in table 1. Specifically, they show evidence for: auto-correlation via Ljung-Box tests (for all variables, lags 1 and 4); non-normality via Jarque-Bera tests (for INF and OG); ARCH structure via LM statistics (for all the variables); and the coefficient of variation. Additionally, they plot the series and match their shifts to real-world macroeconomic events.

The authors run unit root tests, but run too many. Zivot-Andrews and Lee-Strazicich unit root tests were done, as they better account for structural breaks, something important in a regime switching context. They consider drift and drift + trend specifications, with lags chosen by BIC. In general, they find evidence that OG is stationary and WPI and WCMR aren't.

1.1.4 Findings and Conclusions

The authors successfully identify a high and low uncertainty differentiation of the regimes, as regime 2 coincides with several known crises and shocks. The plotted series against the regimes show signs of increased volatility and/or negative trends during the high uncertainty regime.

Both regimes are highly persistent, with $p_{11} = 0.922$ and $p_{22} = 0.914$. The long-run average probability of regime 1 is greater than 2's, which aligns with the probability for recessionary periods being lesser than expansionary periods.

The regime-dependent impulse response functions (RDIRFs) reveal different responses of the variables to a monetary policy shock, depending on the regime:

> We see that the responses of output growth and inflation facing a positive shock in monetary policy (tight monetary policy) are significant and in line with the conventional economic theory in both regimes. However, in the high uncertainty regime, responses are less compared to the responses in the low uncertainty regime. Output growth falls by at most approximately 0.61% in the low uncertainty regime, while in the high uncertainty regime it falls by at most approximately 0.30%. *[…]* A similar pattern is observed for the response of inflation following the shock in monetary policy. However, the difference between the peak response of inflation in low and high uncertainty regimes is less than output growth.

---

[1]The authors use several names for the variables, which was confusing. IIP is transformed into GIIP and OG (output growth), WPI is transformed into INF (inflation), and WCMR is also called IR (interest rate).

As a sensitivity analysis, the authors change the ordering to OG, IR, and INF. The general findings stay the same, with only intensities changing.

The authors show that these findings are in line with existing theoretical justifications for damped effects of monetary policy during periods of high uncertainty. They prescribe avoiding "wait and see" policy framing to avoid uncertainty in agents and more aggressive policy measures during high uncertainty periods (state-dependent policy). Additionally, more communication from the monetary authority is recommended to reduce uncertainty.

1.2 General Comments

In general, the paper is well-written and the methodology is sound. Specifically, they motivate their work and explain the results with very thorough references to the literature. The usage of RDIRFs is a good novelty, and the Bayesian approach helps this implementation.

Some points could be improved, in order of importance:

1. A better description of the Gibbs sampling procedure: the priors for each parameter, the conditional distributions used in each sampling step, and initial values, all should've been better described, specifically in mathematical terms. More details on the prior justifications should've been included, explaining their level of informativeness. This would help to show how robust the results are (or aren't) to the priors.
2. Diagnostics on the posterior distributions of the parameters: information on the effective sample size of the chains, convergence and dependence plots and diagnostics, etc. would help to show the quality of the Bayesian procedure.
3. Sensibility analysis in the chosen proxies for variables: there are other price and production indexes that could've been used to see if the results hold. The authors discuss the difficulty of monetary policy proxy selection, but could've considered other options too.
4. A more visual approach to the data analysis: histograms and ACF plots would've accompanied the Jarque-Bera and Ljung-Box tests well, showing more information about the data.
5. More detail on the minor modeling procedures: the parameters of the X13-ARIMA-SEATS de-seasonalization; results of the selection of p and m (the information criteria values associated with all the values considered); BIC criteria instead of AIC, given the small sample size.

In this replication, I intend to follow the paper as closely as possible, while also addressing some of the points above. My main focus will be on the methodology points 1 and 2.

2 Methodology

2.1 The MS-VAR Model

The authors defend the use of a Markov Switching VAR model, as its non-linear nature allows for capturing dynamic properties (time-varying cycles, breaks/jumps, etc.). I'll briefly describe it below.

$$X_t = \begin{cases} \alpha_1 + \beta_{11}X_{t-1} + \cdots + \beta_{p1}X_{t-p} + A_1 v_t & \text{if } S_t = 1 \\ \quad\quad\quad\quad \vdots \\ \alpha_m + \beta_{1m}X_{t-1} + \cdots + \beta_{pm}X_{t-p} + A_m v_t & \text{if } S_t = m \end{cases} \tag{1}$$
$$V_t \sim N(0, I_k)$$

Where:

- X_t are the K endogenous variables. Let $X := (X_t)_{t=1}^T$.
- $\Phi_i := [\alpha_i, \beta_{1i}, \ldots, \beta_{pi}]$ (each element a row), for all the m regimes i, are the $K \times (p+1)$ VAR(p) coefficients.
- $A_i v_t$ are the residuals, with A_i being a $K \times K$ matrix.
- S_t is an unobserved regime variable.

Consider the alternate formulation, that will be useful for prior formulation:

$$X_t = \begin{bmatrix} \mathbf{1}_k & X_{t-1} & \ldots & X_{t-p} \end{bmatrix} \cdot \Phi_i + A_i v_t, \quad \forall_{i \in \{1,\ldots,m\}} \tag{2}$$

Note that the variance-covariance matrix of the residuals is regime dependent:

$$\Sigma_i = E\left(A_i V_\nu V_\nu' A_i'\right) = A_i E\left(V_\nu V_\nu'\right) A_i' = A_i I_K A_i' = A_i A_i' \tag{3}$$

S_t is assumed to be independent of past X's, and, conditional on S_{t-1}, follows a hidden Markov process:

$$Pr(S_t = j | S_{t-1} = i) = p_{ij}, \quad \forall_t, \ \forall_{i,j \in \{1,\ldots,m\}}$$

The transition probabilities are grouped in the transition probability matrix P:

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1m} \\ p_{21} & p_{22} & \cdot\cdot & P_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ P_{m1} & P_{m2} & \cdot\cdot & P_{mm} \end{bmatrix} \tag{4}$$

Equations 1, 3, and 4 are the core model. All parameters are allowed to switch with the regime. Regime changes are treated as random events modelled by the Markov process.

In the paper, $K = 3$ (see data section), $p = 2$ (see model selection section), $m = 2$ (high and low uncertainty), and $T = 312$ time periods. Then, we have the following parameters:

- (Φ, Σ, P) is comprised of $m \cdot (k * (p+1) + K^2 + m) = 40$ parameters.
    - All the m Φ_i's, the $K \times (p+1)$ matrix of coefficients.
    - All the m $K \times K$ variance-covariance matrices $\Sigma = \{\Sigma_i : i = 1, \ldots, m\}$.
    - The $m \times m$ transition probability matrix P.

- We are also going the estimate the standard error matrix of Φ and the variable S (T entries).
    - If coefficients are taken as independent, Φ has $(1 + 2K) \cdot K = 21$ free entries. If they are taken to be independent across equations, it has $(1 + 2K)^2 \cdot K = 147$ free entries.

2.2 Estimation

The authors note that estimation of Markov switching models can be performed using Maximum Likelihood (ML) or the Expectation-Maximization (EM) algorithm. However, the ML approach becomes computationally intensive and challenging to implement as the number of parameters increases. The EM algorithm, while more tractable, doesn't provide standard errors directly, requiring additional computations. The authors choose the Bayesian Markov Chain Monte Carlo (MCMC) approach as it is more flexible and provides full posterior distributions of all parameters.

The authors follow the Gibbs sampling procedure outlined in (Balcilar et al. 2017). The procedure was poorly described (in both papers), below I give more details on the steps, priors, and initial values, with some personal choices.

2.2.1 Priors

First, note that the likelihood for an MS-VAR with Gaussian errors is the sum over all possible regime sequences $(S_1, \dots, S_T)$, given transitional probabilities:

$$\ell(\theta) = L(X|\theta) = \sum_{S_1=1}^{m} \dots \sum_{S_T=1}^{m} \left[\prod_{t=1}^{T} f(X_t \mid X_{t-1}, \dots, X_{t-p}, S_t, \theta) P(S_t \mid S_{t-1}, P) \right] \quad (5)$$

Where f is the Gaussian density, such that the likelihood has overall Gaussian functional form. Now, we can discuss the parameter's priors.

2.2.1.1 VAR coefficients (Φ_i)

For each regime, normal priors. I consider two sets of parameters, one empirically-driven, and Minnesota-type set, following (Litterman 1986).

The Minnesota parameters are as below. Other specifications could've been considered.

- Intercepts have 0 mean and $\sigma^2_{\alpha_k}$ variance. Let σ^2_α denote the aggregating vector.
- AR(1) coefficients have 1 mean, and 0 for all the others.
- AR(p) variances are $(\sigma^2_{\beta_{ii}}/p)^2$, and the related to equation i/variable j are $(\sigma^2_{\beta_{ii}} \theta \sigma_i / p\sigma_j)^2$. Let v_{ip} denote the row of such variance matrix.
- $0 < \theta < 1$ controls the relative tightness of the prior variance in the other lags, and σ_i is the ith diagonal element of Σ_i.
- Let V be the vector with all the elements in σ^2_α, v_{11}, v_{12}, v_{13}, v_{21}, ...
- I use $\theta = 0.1$, as the cross-correlation functions show little correlation between the variables (see data section). The values of $\sigma^2_{\beta_{ii}}$ are taken by the estimated variance of each VAR (OLS).

The emprical-driven one considers intercepts to be the unconditional averages of the series, and smaller AR(1) coefficients following the stationary evidence from the data section. The covariance is taken from the OLS estimates.

Then, vec(Φ_i) is distributed as a multivariate normal distribution, with averages (in order):

$$\mu_{\Phi_i} = \text{vec} \left(\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right), \quad \text{vec} \left(\begin{bmatrix} -0.8 & 0 & 7 \\ 0.7 & 0 & 0 \\ 0 & 0.7 & 0 \\ 0 & 0 & 0.7 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right) \quad \forall_{i \in \{1,...,m\}}$$

And variances (in order):

$$\Sigma_{\Phi_i}^2 = \text{diag}(V), \quad \text{block-diag}(Cov(\Phi^{OLS}, 3)) \quad \forall_{i \in \{1,...,m\}}$$

There could be added covariance terms to the diagonal structure of the Minnesota set. Also, block-diag$(., 3)$ creates a block diagonal structure with 3 row/column blocks.

I did not implemented the Minnesota prior as a hierarchical one, by dynamically updating the hyperparameters σ_i and σ_j. Instead, I considered them as fixed, taken from the OLS estimate of Σ. At the end, the empirical prior yielded better convergence results, so it was the one used.

2.2.1.2 Transition probabilities (P)

Dirichlet priors with parameters 0.8 and 0.2 (respectively) following the authors argument on high persistence of regimes.

Let P_i denote P's i^{th} row, then:

$$P \sim Dir((0.8, 0.2))$$

2.2.1.3 Variance-covariance matrices (Σ_i)

For each regime, inverse-Wishart priors, following (Balcilar et al. 2017).

The scaling matrix is the sample covariance matrix of the data, which is similar to the covariance matrix of the OLS residuals. It was used very low degree of freedom, to be less informative. $K+1$ is the minimum for a proper prior. Larger values than that yielded poor convergence.

$$\Sigma_i \sim IW(K + 1, \widehat{Cov}(X))$$

2.2.2 Initial Values

I'll set P's initial value as $P_1 = (0.8, 0.2)$ and $P_2 = (0.2, 0.8)$. For the sampling of S, I use a $\mathbf{0.5}_{m \times m}$ matrix, to avoid a series dominated by the regime S_0, from it, S is sampled following the Markov Process. For Σ_i, the initial values are taken empirically from the covariance matrix of the OLS residuals. The other parameters do not need initialization.

2.2.3 Sampling Procedure

Let $t(S, i)$ return the time periods where $S_t = i$. Sampling steps:

1. **Initialization**: set starting values for S and Σ.
2. **Sampling Φ_i**: for each regime i, sample Φ_i from $Pr(\Phi_i | X, S)$.

   - Given S, we get a Gaussian + Gaussian conjugate distribution. Let $V = ((\Sigma_{\Phi_i}^2)^{-1} + X'_{t(S,i)} X_{t(S,i)})^{-1}$, the posterior is N with variance V and mean $V(\Sigma_{\Phi_i}^2 \mu_{\Phi_i} + X'_{t(S,i)} Y_{t(S,i)})$.
   - I added a stationarity check, that resamples the VAR coefficients if the roots of the characteristic polynomial are outside the unit circle.

3. **Sampling S**: for each time period t, draw the regime indicator S_t from $Pr(S_t | X, \Phi, \Sigma, S_{t-1}, P)$. Only accept draws where at least 5% of observations fall in the regimes associated with it.

   - The data, parameters, and Σ are used to find the likelihood of a each regime, given the historic values of regimes. This in turn is used to update the probabilities from P. Only then, S is sampled. The algorithm is implemented by backward sampling, and uses log-likelihood tricks to improve stability.

4. **Sampling P**: draw the transition probabilities P from $Pr(P|S)$.

   - Because of the independence from the past of the Markov process, P's rows conditional posterior only depends on S, and can be sampled from the Dirichlet distribution. Let $c = \sum_{t=2}^{T} I\{S_t = S_{t-1}\}$, the posterior is $Dir((0.8 + c, 0.2 + (T - 1 - c)))$.

5. **Sampling Σ_i**: for each regime i, draw the variance-covariance matrix Σ_i from $Pr(\Sigma_i | X, \Phi, S, P)$.

   - Given the parameters, we get the Gaussian + inverse-Wishart conjugate. Let e_i be the vector of residuals of the regime, the posterior is IW with DFs $(K+1) + \sum_{t=1}^{T} I\{S_t = i\}$ and scaling matrix $\widehat{Cov}(X) + \sum_{t \in t(S,i)} e_{it} e'_{it}$

6. **Repeat steps 2-5**: the authors iterate until convergence to obtain the joint posterior distribution of all parameters. In practice, it is done a loop with $n_{sims} = 60,000$ MCMC iterations, with $30,000$ burn-in iterations.

   - Due to computational constraints, I used $n_{sims} = 1,000$ and $n_{burnin} = 200$, which implied a much poorer sampling. The diagnosis analysis helps to understand the extent of the problem.

In pseudo-code terms, the algorithm is described below. The functions with the same name can be seen in the file sampling.R.

2.3 Regime Dependent IRFs

With a non-linear model, it is hard to calculate and interpret impulse response functions (IRFs). The authors follow (Ehrmann, Ellison, and Valla 2003) to calculate regime-dependent IRFs, that assume regime switching does not occur beyond the shock horizon. The effect of a one standard deviation shock to variable k ($v_{k,t}$) at time t on X at horizon $t + h$ ($h \geq 0$) is given by:

$$\left. \frac{\partial E_t X_{t+h}}{\partial V_{k,t}} \right|_{S_t = \cdots = S_{t+h} = i} = \rho_{ki,h}$$

Algorithm 1 Gibbs Sampling Procedure for MS-VAR Model

1: **Container:** create an empty container $R_{1:n_{sims}}$.
2: **Initialization:** Set starting values for P, S, and Σ.
3: **for** $s \leftarrow 1$ **to** n_{sims} **do**
4:      **for** $i \leftarrow 1$ **to** m **do**   ▷ Sample Φ
5:          $\Phi_i \leftarrow$ sample_Phii(X, S)
6:      **end for**
7:      **for** $t \leftarrow 1$ **to** T **do**   ▷ Sample S
8:          $S_t \leftarrow$ sample_St$(X, \Phi, P, \Sigma, S_{t-1})$
         Only accept draws where at least 5% of observations fall in the associated regimes.
9:      **end for**
10:     $P \leftarrow$ sample_P(S) ▷ Sample P
11:     **for** $i \leftarrow 1$ **to** m **do**   ▷ Sample Σ
12:         $\Sigma_i \leftarrow$ sample_Sigmai(X, Φ, S, P)
13:     **end for**
14:     $R_s \leftarrow (\Phi, S, P, \Sigma)$
15: **end for**
16: $R_{(n_{burnin}+1):n_{sims}}$

Where $\rho_{ki,h} = (\rho_{ki,1}, ..., \rho_{ki,m})$ - in the paper, $\theta_{ki,h}$ - is the vector of responses of X to the shock, conditional on regime i.

As always in the VAR literature, we can estimate Σ_i, but not A_i. The authors use the Cholesky decomposition in the order of OG, INF, and IR.

3 Data

The original dataset used data of from the EPWRF time-series database. Its access is paid, so I used an alternative source from the Research Bank of India. The same variables are used, but the time series spans 2005-04 to 2017-02. Other sources were available, such as World Bank data, but the quality was poor.

3.1 Series Analysis

I plot the historic values 1, CCFs 2, and histograms 3 of all the variables, as discussed in the general comments sections[2]. The historic values and CCFs are useful to motivate the empirical prior parameters, the drifts $(-0.3, 0, 8)$, dependence coef$(AR(1)) = 0.7$, and low cross-correlation of the series. The distributions of INF and OG seem to be leptokurtic, and WCMR is obviously asymmetric. I defend that this is a useful addition to the replication.

3.2 Summary Statistics

The summary statistics tables and unit root tests are shown in the tables below.

View Table 1, it is useful to compare to the original paper's tables. We can see that the series actually are quite different. There are larger averages, and less volatility in their series. Larger

---

[2]These were plotted with my own package varr.
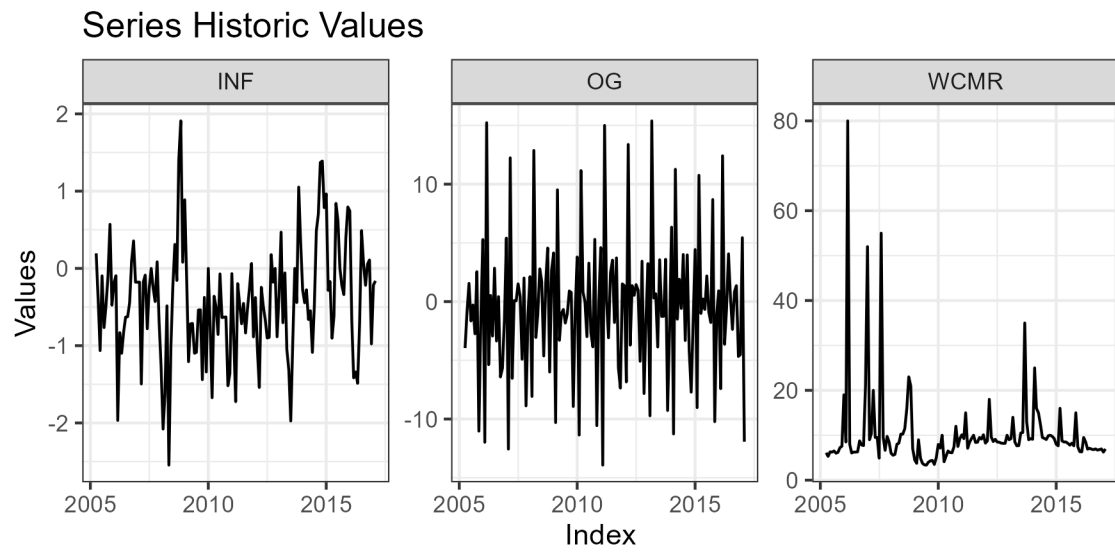
Figure 1: Historic values of the variables
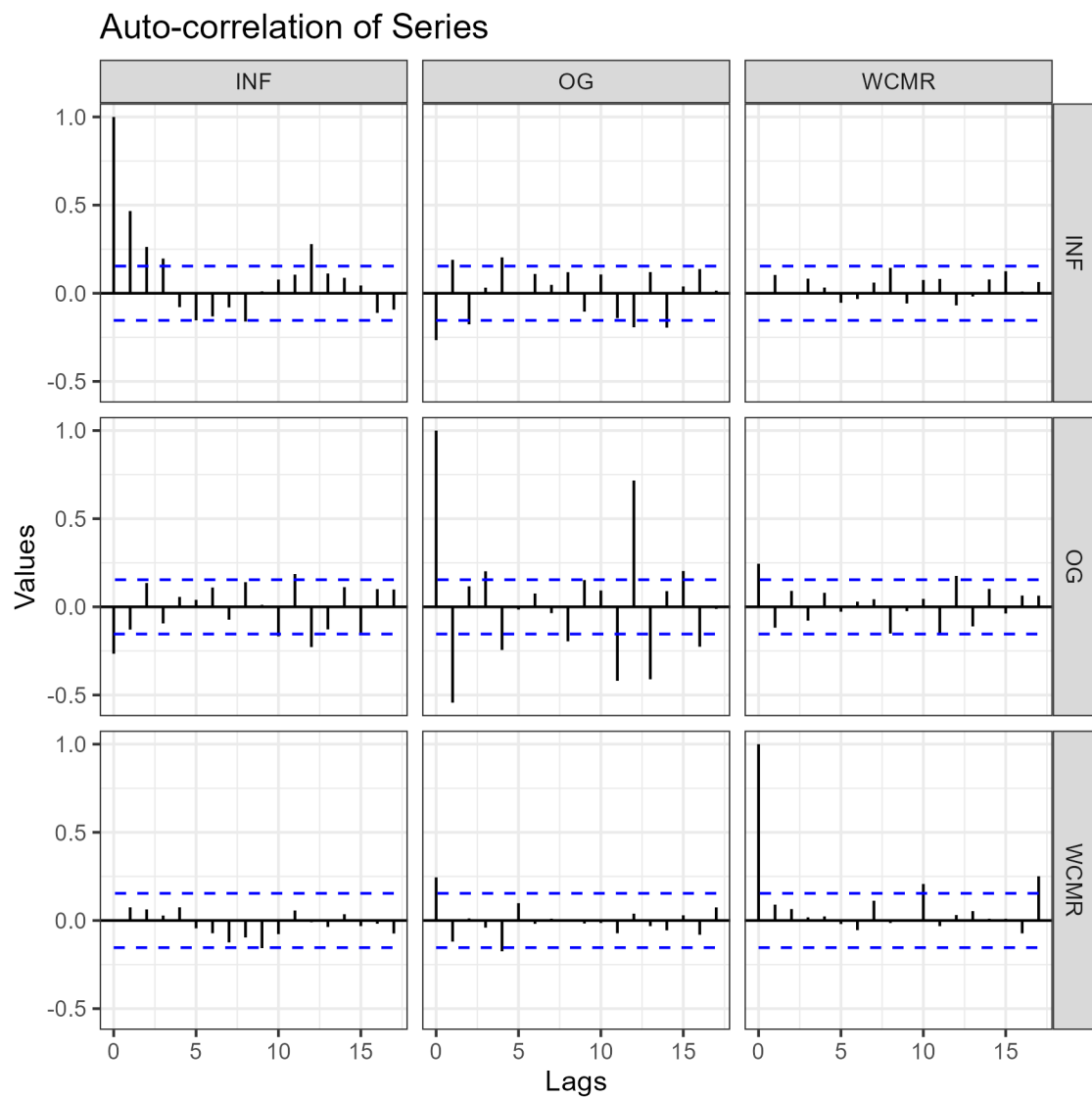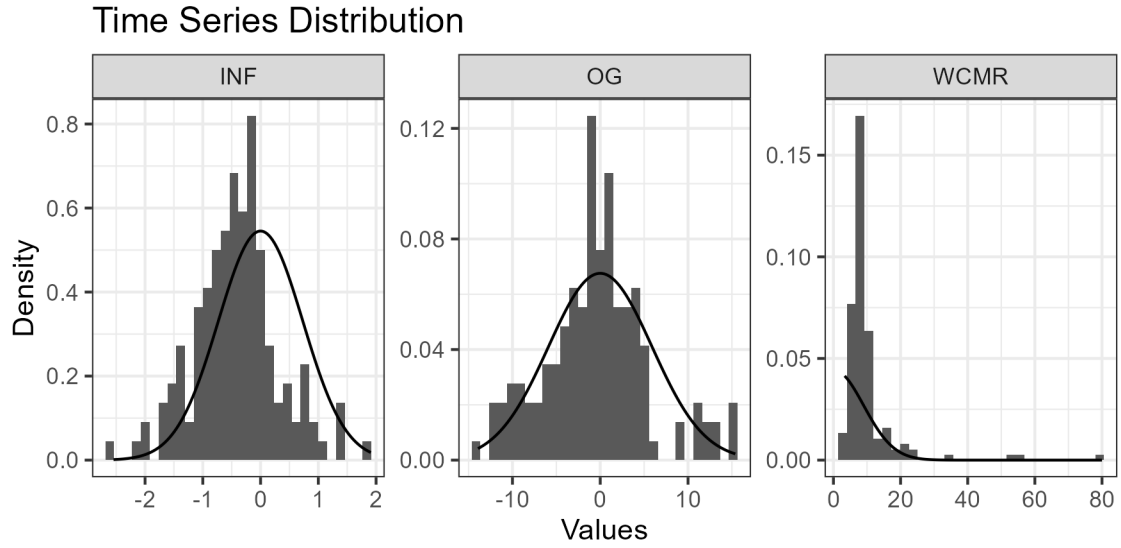

Figure 2: ACFs of the variables

Figure 3: ACFs of the variables

evidence for dependence and ARCH volatility. With the smaller window, this information suggest a harder-to-model dataset for this replication.

Table 1: Summary statistics

Variables/Statistics	GIIP	INF	CMR
Mean	-0.510	-0.415	10.145
Prob.(Mean=0)	0.304	0.000	0.000
Std Dev	5.904	0.732	8.985
Minimum	-13.924	-2.546	3.350
Maximum	15.389	1.909	80.000
Skew	0.319	0.201	5.122
Kurtosis	3.495	3.716	34.235
Jarque-Bera	3.886	4.026	6438.319
Prob.(JB=0)	0.143	0.134	0.000
LJung-Box(Q1)	43.096	31.765	1.181
Prob	0.000	0.000	0.277
LJung-Box(Q4)	60.042	48.575	1.928
Prob	0.000	0.000	0.749
ARCH(1)	9.869	3.179	0.004
Prob	0.002	0.075	0.951
ARCH(4)	23.030	5.151	0.097
Prob	0.000	0.272	0.999
COV	11.584	1.765	0.886

As discussed, too many unit root tests were done, Id've chosen the ADF test and the structural break varieties (given the topic), but did all besides Ng-Perron for completeness. See Table 2 and Table 3. We can see greater evidence for stationarity, but, not on the structural break ones, which could imply a different relevance of breaks in my dataset.

Table 2: Unit root tests

	Test-stat			Critical Values		
Tests/Variables	OG	INF	WCMR	1%	5%	10%
ADF (C)	-13.537	-5.705	-7.479	-3.454	-2.871	-2.572
ADF (C+T)	-13.649	-5.789	-7.557	-3.992	-3.426	-3.136
DFGLS (C)	-0.556	-2.244	-1.567	-2.573	-1.942	-1.616
DFGLS (C+T)	-3.000	-2.886	-2.170	-3.470	-2.910	-2.606
PP (C)	-23.827	-7.218	-10.852	-3.451	-2.871	-2.572
PP (C+T)	-24.227	-7.294	-10.920	-3.988	-3.424	-3.135
KPSS (C)	0.107	0.291	0.223	0.739	0.463	0.347
KPSS (C+T)	0.045	0.100	0.098	0.216	0.146	0.119

Table 3: Unit root tests with structural breaks

	ZA test			LS test		
Variables	Statistics	P-value	Break Date	Statistics	Break Date 1	Break Date 2
GIIP	-3.86	0	2009:01	-9.696	2010:07	2015:02
INF	-4.805	0	2009:02	-7.199	2008:05	2009:03
CMR	-3.075	0.001	2007:09	-11.404	2007:04	2011:01

4 Estimation and Model Results

The results from the sampling and derived statistics are shown in this section. First note that:

- The different dataset and new modeling section imply that the results won't be too comparable. That's ok, altering the poorly described methodology was part of my goal.
- That said, the quality of the results can be poor. The low number of iterations and overall poorer quality of the data poses a problem.

4.1 Diagnosis

The traceplots, ACFs, and histograms of the posterior vs. prior distributions, for the most relevant parameters[3], can be seen in the appendix. The red line denotes the burn-in period. The histograms are from the posterior, and black lines are the priors. Comments are below.

- For P, we can see how the sampling goes away from 0.5 quickly for regime 2, and doesn't for regime 1. The final distribution is even more persistent than the prior.
- For S, I present the average duration of regimes, which shows a multimodal distribution, in line with the difference in the regimes. Looking back, this graph would've been more interesting if separated by regimes.
- For Φ, it is interesting to see how the matching of priors varies across regimes. $MCMR$ presents overall less dependence.

---

[3]The Phi parameters more related to $MCMR$.

The effective sample size, for the 5 lowest, and 5 greater, are shown in Table 4 [4]. The standard error of Φ has very little effective sample, such that the standard errors of the IRFs are what will be most biased by the low number of simulations. At the same time, the Φ coefficients have good sample sizes.

Table 4: Effective sample size

	$se(\Phi)$ 27	$se(\Phi)$ 214	$se(\Phi)$ 221	Φ 221	Φ 113	Φ 110	Φ 210	Φ 18	Φ 28
IACT	15.87	15.87	15.87	15.78	0.85	0.83	0.79	0.77	0.77
Neff	23.45	31.3	31.3	31.3	999	999	999	1092.17	1223.96

4.2 Model Results

Now, I'll talk about the actual MS-VAR model results. Lets start with table 5 (Table 5), which is very confusing, with information criteria mixed with data from P and S. I've chosen to separate it. The likelihood of the model is vastly smaller than the original paper's values. But, information criteria aren't too comparable between largely different models.

Table 5: Effective sample size

	MS(2)-VAR(2)	Linear VAR(1)
Log-Likelihood	-895.029	-1052.692
AIC	1870.058	6.482
BIC	1988.572	6.732
HQ	1918.216	6.584
log(FPE)	6.587	6.482

Then, the main difference can be seen in lower probability of regime 1 in Table 6. Going against the prior and literature on persistent regimes, the first regime has a $(0.5, 0.5)$ row. Regime 2 has the expected result. This has the effect of larger unconditional probability for R2, more observations, and larger duration. R1 mainly has only 1-month spells.

Table 6: Effective sample size

	Regime 1	Regime 2	Probability	Observations	Duration(Months)
Regime 1	0.441	0.558	0.154	22	1.69
Regime 2	0.129	0.870	0.846	121	8.64

To further analyze the differences, see Figure 4 with the smoothed probabilities, and Figure 5 with the series on top of it. I've removed the superfluous second panel of smoothed probability. We can see that the probability of regime 1 spikes in some specific months, but never in big spells. Specially, regime 2 dominates during 2008–2012 and 2015 onwards. With very little regime 1 periods, it is hard to see much of a pattern. While this could just signify that there aren't strong results in the data, it is probably a bad result, coming from the limitations of the present work. The regime interpretation from the original paper does not hold here.
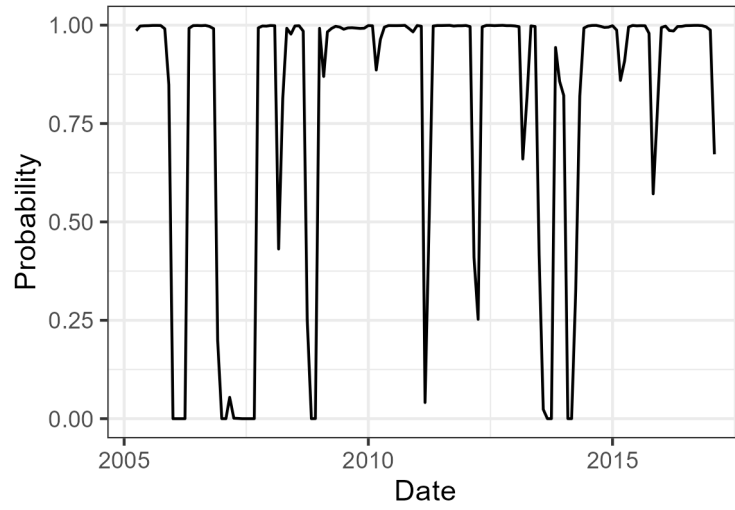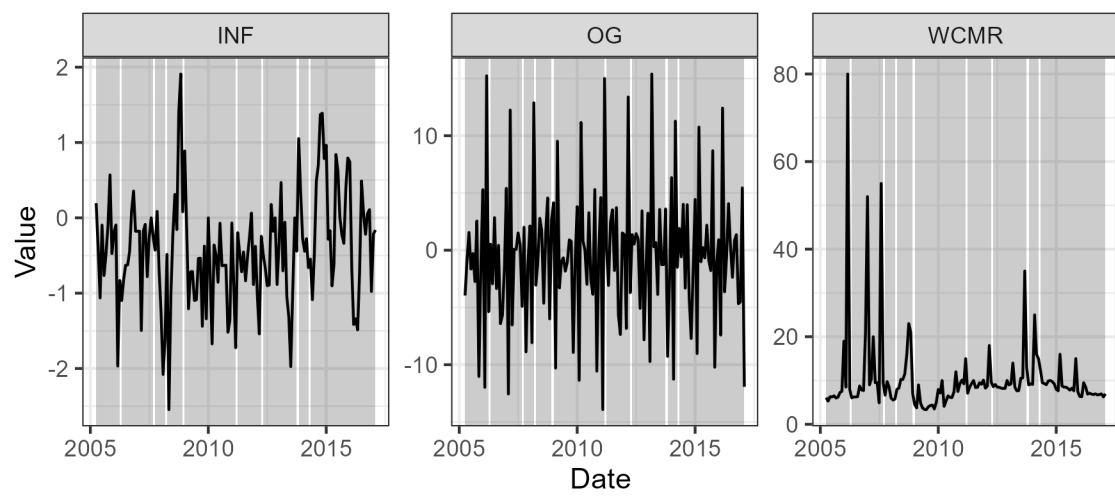
Figure 4: Smoothed Probability of Regime 2

Figure 5: Series and Smoothed Probability of Regime 2

Finally, we can interpret the RDIRFs. As the authors did, Figure 6 has the order OG, INF, and MCMR, while Figure 7 has the order OG, MCMR, and INF. The regime interpretation is not the same, so there is not a parallel do draw there. Here we can see another main differnce of the replication, with strongly different IRFs. It is interesting that the effect of MCMR on itself and inflation has different signs at each regime, with unusual interpretations for regime 2 of positive effect on inflation. This should be taken with a grain of salt.

The sensitivity analysis is very different, with a strange alternating pattern, showing some instability in the sample. The authors found very similar results when comparing with the first ordering.

Figure 6: Impulse Response to Interest Rate (WCMR) Shock

---
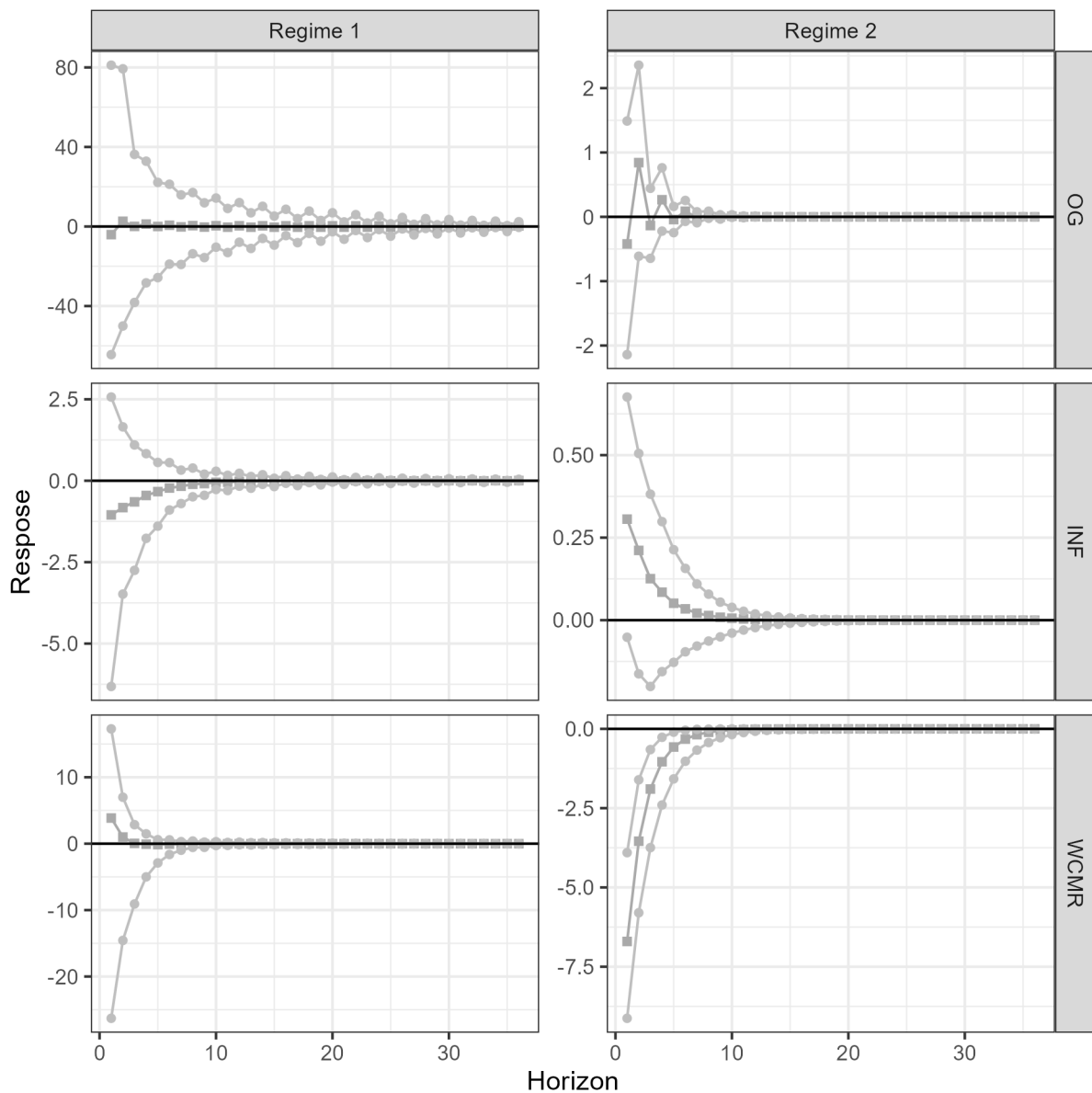
[4]The numbering in the column names relates to the $vec(.)$ representation of the variables, with the regimes concatenated as columns. Overall, not very useful.

Figure 7: Impulse Response to Interest Rate (WCMR) Shock
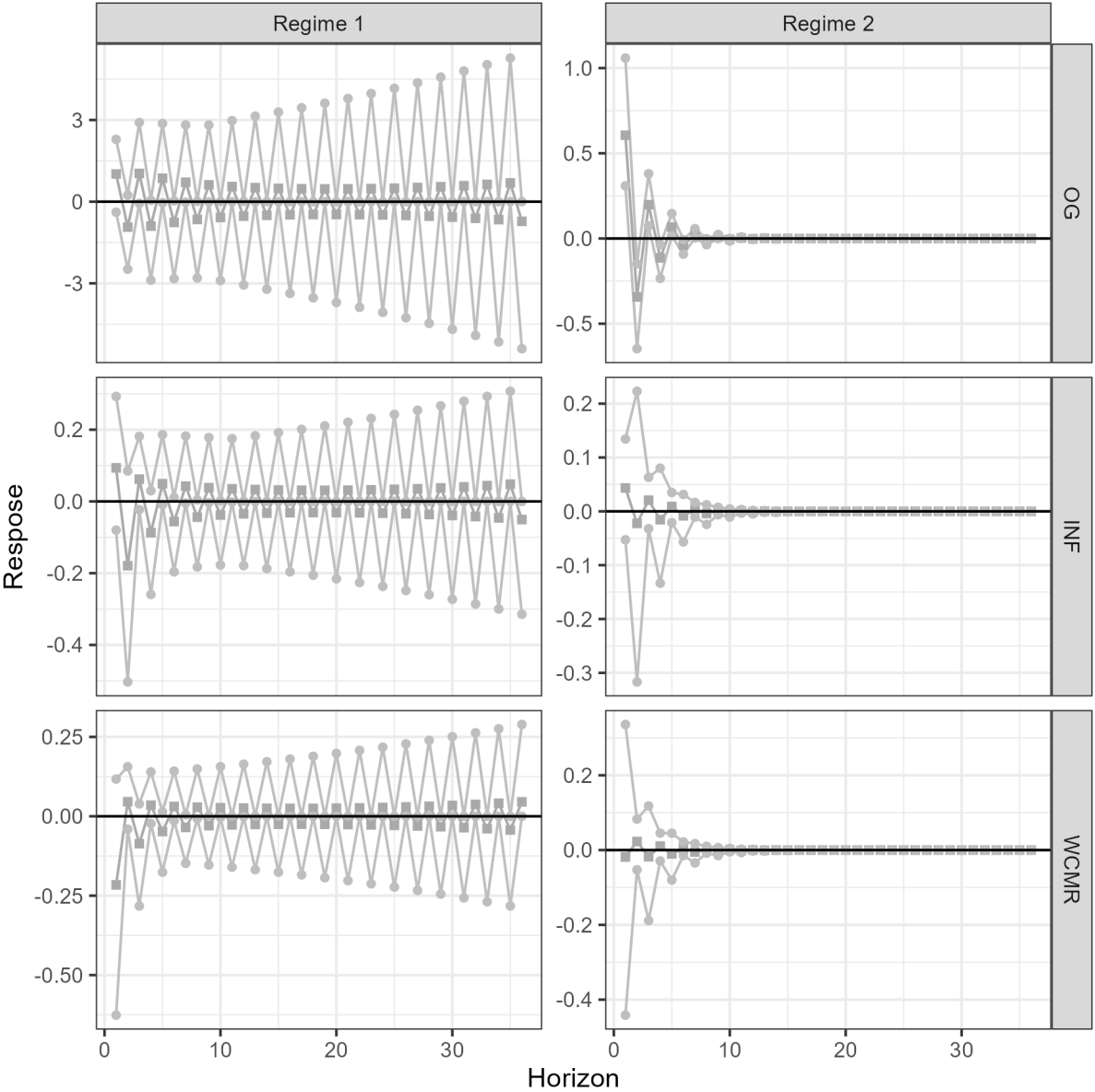
5 Conclusion

While the results were disappointing, I believe most of the issues stem from computational and time constraints on the number of simulations. The different results for S are probably not due to this but could be attributed to the smaller dataset.

In general, I am very pleased with the setup created. The method, even if not perfect, is much better described. The matrix notation for Φ yielded a highly generalizable implementation.

The coding implementation is another point of great effort. The matrix notation for Φ resulted in a highly generalizable implementation, making the code readily adaptable to any K and p, and almost readily adaptable to any m. There are 1,197 lines of code, with 33 functions defined. The only statistical packages used were for distributions, `urca` for tests, and `coda` for $N\_\{eff\}$'. LLMs were used to discuss some implementations of the Bayesian algorithms, but the code is fully original.

In summary, while the results could have been better, and there might be flaws in the methodology, I am very content with the overall implementation.

References

Balcilar, Mehmet, Riza Demirer, Rangan Gupta, and Reneé van Eyden. 2017. "The Impact of US Policy Uncertainty on the Monetary Effectiveness in the Euro Area." *Journal of Policy Modeling* 39 (6): 1052–64. https://doi.org/https://doi.org/10.1016/j.jpolmod.2017.09.002.

Ehrmann, Michael, Martin Ellison, and Natacha Valla. 2003. "Regime-Dependent Impulse Response Functions in a Markov-Switching Vector Autoregression Model." *Economics Letters* 78 (3): 295–99. https://doi.org/https://doi.org/10.1016/S0165-1765(02)00256-2.

Litterman, Robert B. 1986. "Forecasting with Bayesian Vector Autoregressions: Five Years of Experience." *Journal of Business & Economic Statistics* 4 (1): 25–38. http://www.jstor.org/stable/1391384.

Nain, Zulquar, and Bandi Kamaiah. 2020. "Uncertainty and Effectiveness of Monetary Policy: A Bayesian Markov Switching-VAR Analysis." *Journal of Central Banking Theory and Practice* 9 (s1): 237–65. https://doi.org/10.2478/jcbtp-2020-0030.

Sengupta, Nandini. 2014. "Changes in Transmission Channels of Monetary Policy in India." *Economic and Political Weekly* 49 (December): 62–71.

A Appendix: Sampling Diagnosis
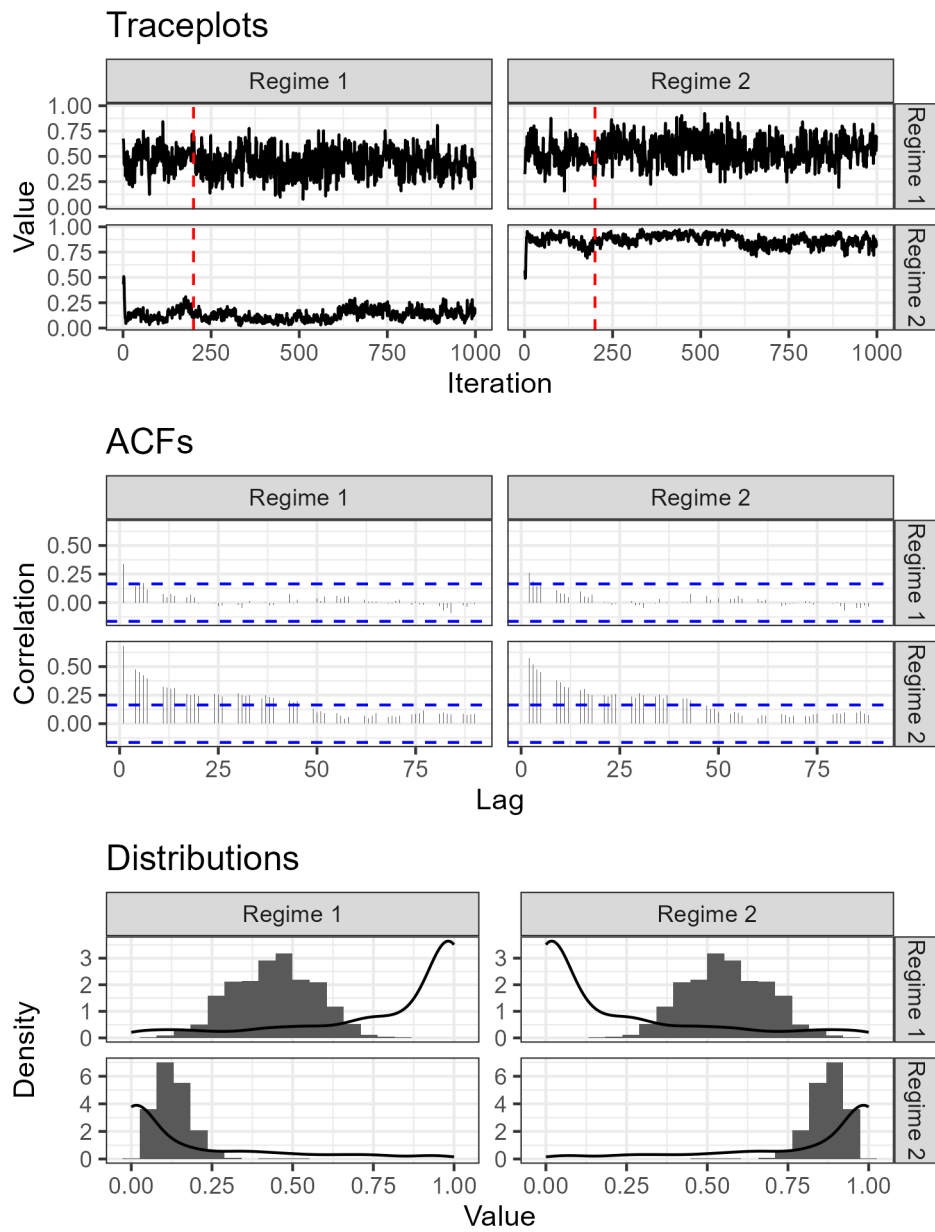
Figure 8: Diagnosis plots for P
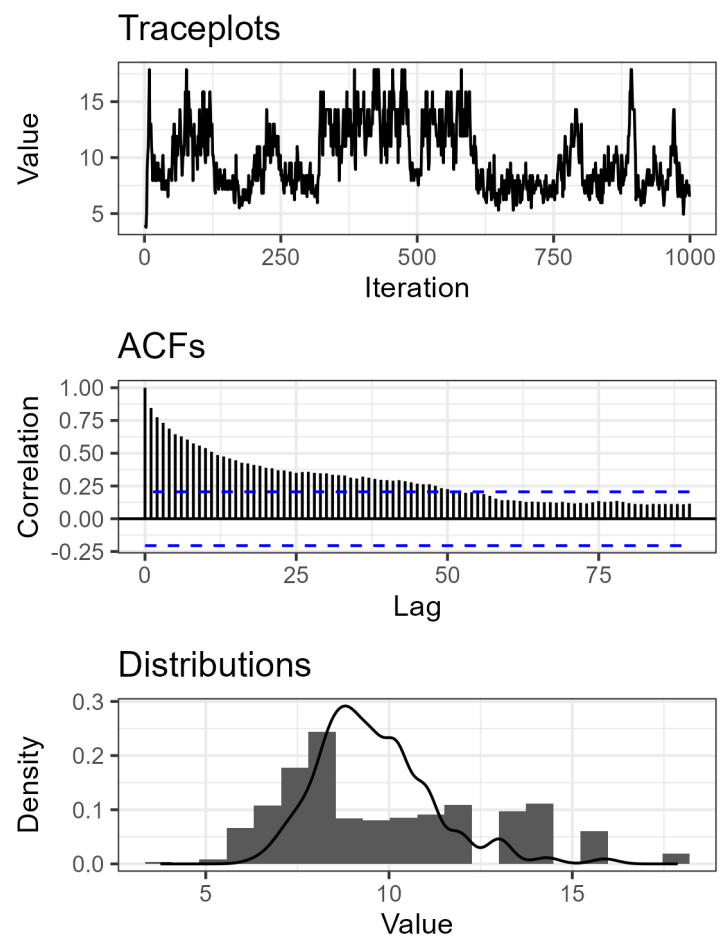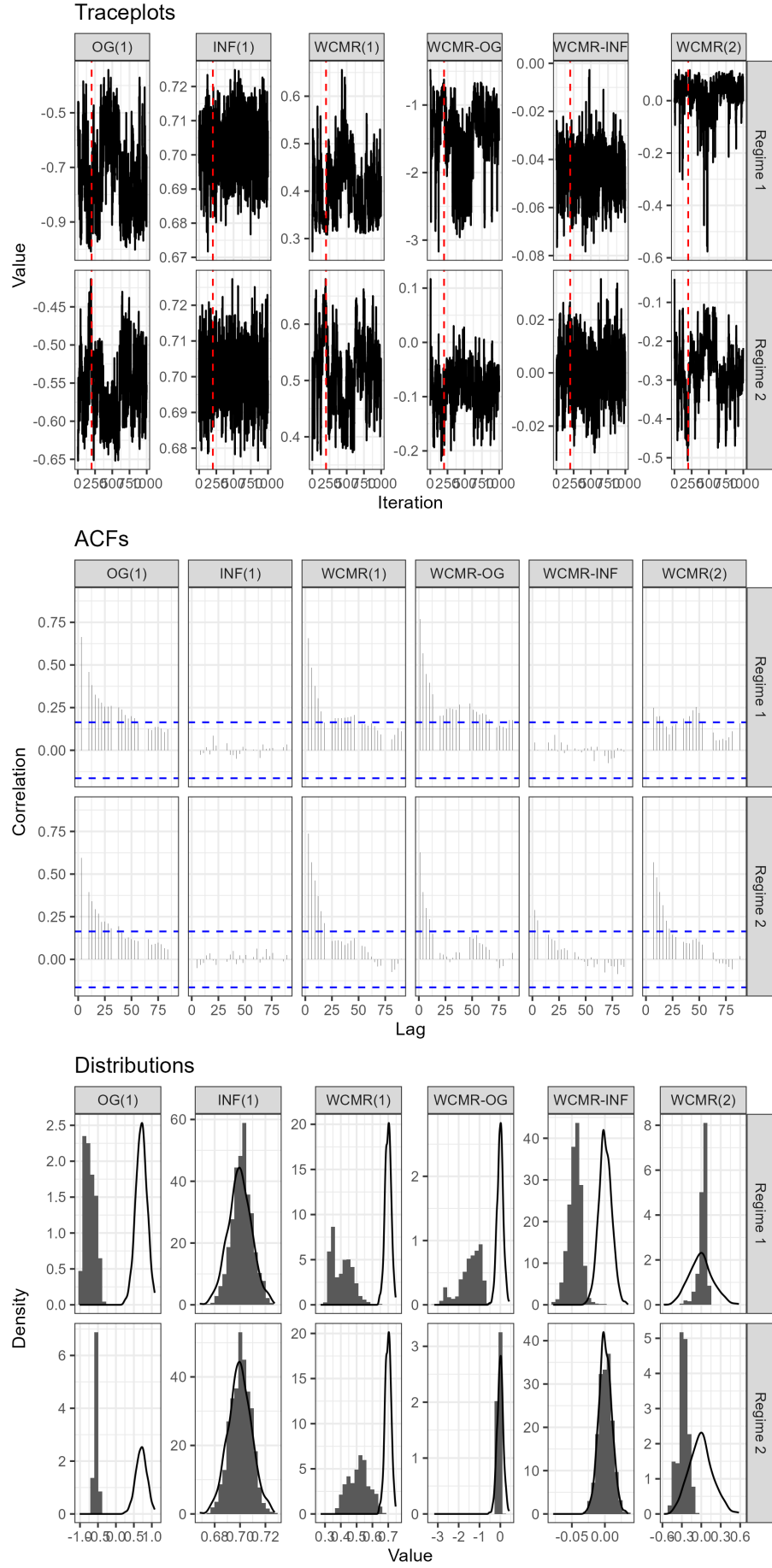
Figure 9: Diagnosis plots for S

Figure 10: Diagnosis plots for *Phi*

Figure 11: Diagnosis plots for *Sigma*