Markov Switching Gaussian Mixture Linear Regression

By Zhenning Zhao*

Draft: March 29, 2024

This paper introduces a novel econometric model, the Markov Switching Gaussian Mixture Linear Regression, wherein both the probability model and the linear model are estimated simultaneously. A detailed discussion on the estimation procedure and the asymptotic standard errors of the estimators is provided, along with methods for predicting both in-sample and out-of-sample data. Finally, the model is applied to a MS-GMLR-VAR model on US aggregate data. The results indicate significant changes in the states of the world during periods of recession.

Ordinary Least Squares (OLS) regression stands as one of the most fundamental tools in the arsenal of econometricians for analyzing the relationship between any dependent and independent variables. For instance, given a pair of variables, as depicted in the left panel of Figure 1, OLS regression offers a straightforward way to determine their linear relationship. However, upon closer inspection of this figure, it becomes apparent that the relationship between these variables is not strictly linear. Rather, there appear to be two distinct linear patterns, with nature seemingly arbitrarily selecting one of them, as illustrated in the right panel of Figure 1. Indeed, this duality represents the true data generating process underlying this graph. I created these data by randomly assigning each data point to one of two possible states and subsequently determining the dependent variable's value based on the state and the independent variable.

* University of Texas at Austin, Economics Department. Please contact the author via email: znzhao@utexas.edu. The author appreciates the suggestions and comments provided by Oliver Coibion, Saroj Bhattarai and Stefano Eusepi.

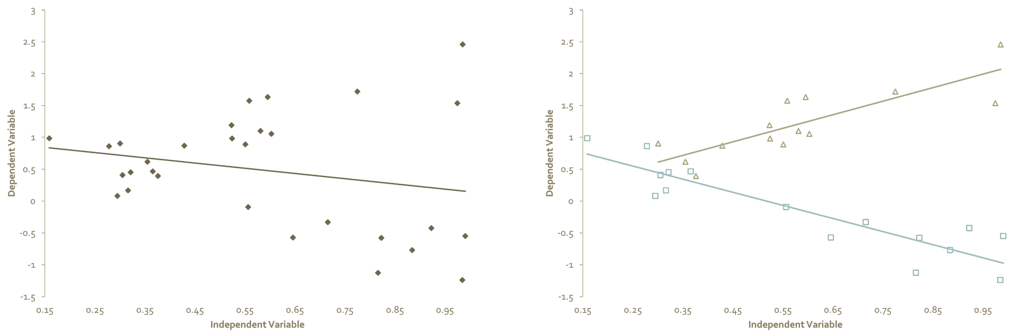

Figure 1. OLS Regression vs True Data Generating Process

Now, suppose we could directly observe the state of the world. In that case, effectively estimating the relationship between X and y within each state wouldn't be a significant challenge. We could simply conduct an OLS regression of y on X for each state, and voilà, we would have our solution. The truly interesting aspect arises from the fact that, as econometricians, we lack direct observation of the true states. This transforms the problem into a clustering task, especailly not a conventional task, as we need to cluster the relationship between X and y rather than the values of X and y themselves, which is a nontrivial challenge. This is precisely where Gaussian Mixture Linear Regression (GMLR) comes into play.

In cite1 and cite2, the authors developed the baseline algorithm for GMLR. "Gaussian" denotes the assumption of normally distributed error terms. "Mixture" indicates the presence of multiple hidden states. "Linear Regression" implies that the relationship between dependent and independent variables remains linear. The authors employed the EM algorithm and Maximum Likelihood Estimation (MLE) to accomplish this task and demonstrated that the algorithm yields consistent results.

While the baseline GMLR algorithm is powerful, it struggles with several issues. Firstly, it can only handle the case of one single dependent variable, yet in the real world, we often need to consider multiple variables simultaneously. This leads to correlations in the error terms for these variables, thereby inflating the number of parameters to be estimated. Secondly, in the previous papers implementing GMLR, normally people start with assuming an exogenous set of parameters controlling the probability of each data point belonging to each state. Subsequently, they run another linear regression of the posterior probability of the data being in a given state on the observables, aiming to elucidate which observable factors contribute to a data point being in a specific state. However, this approach introduces inconsistency into the model. If one assumes that the probability is exogenous, it implies that no observables should have an effect on the probabilities. So, how is it justified to assume in the second stage that observables can influence the probabilities?

Thirdly, the baseline algorithm exclusively operates with cross-sectional data. However, in the real world, particularly in macroeconomics, we frequently encounter panels of time series data. Although the current Markov switching state-dependent vector autoregression model offers some assistance in addressing this issue, there lacks a unified framework that seamlessly bridges the gap between static and dynamic problems.

This paper aims to address these issues by introducing a generalized GLMR model that incorporates Markov switching properties. This model accommodates multiple dependent variables and their potential covariances. Leveraging Kim's filter as cited in reference 3, I ensure the model's compatibility with Markovian data generating processes. Key to this approach is utilizing smoothed probabilities for calculating the expectations of the log-likelihood function and prior probabilities for computing the inside likelihood. Following parameter estimation, I discuss asymptotic standard errors using the delta method. Finally, I employ the methodology outlined in reference 4 to generate consistent in-sample and out-of-sample predictions for this model.

A significant contribution of this paper is the development of a Python package[1] capable of running three variants of the GMLR models. The package encompasses

[1]The link to the Python package is: https://github.com/znzhao/GMLR.

codes for the baseline GMLR model with multiple variables, as well as the MS-GMLR model. Additionally, it features a data generator capable of producing sample test data following various models. Simulation tests demonstrate that the package consistently estimates true parameters. Moreover, for parameters unrelated to probability and variable changes, p-values are muted.

Finally, I applied the MS-GMLR model in a simple VAR framework for testing purposes. I aggregated monthly data on GDP growth, inflation, and the Federal Funds Rate from the St. Louis Federal Reserve Bank. Subsequently, I executed an MS-GMLR-VAR model with a lag of three months and two states. The results indicate that the VAR relationship among these variables varies during economic downturns. Specifically, the impulse response functions to interest rate shocks exhibit reduced intensity during recessionary periods. Furthermore, lower GDP growth rates and decreased inflation elevate the likelihood of entering a recessionary state, aligning intuitively with the definition of recessions.

In summary, this paper presents a generalized GLMR model that enriches GMLR's capabilities by integrating Markov switching properties and accommodating multiple dependent variables and time series data. Application of the MS-GMLR model in a VAR framework using real-world data demonstrates its effectiveness in capturing dynamic relationships among economic indicators, particularly during recessionary periods. This innovative approach significantly contributes to econometric analysis, offering valuable insights into economic phenomena. The remainder of the paper will proceed as follows: the next section provides a literature review, followed by Section 2, which discusses the main methodology details. Section 3 presents the model's behavior under simulations, while Section 4 applies the model and offers a concrete example in a VAR application. Finally, the last section concludes the paper.

I. Literature Review

TBD.

II. Methodology

A. Econometrics Model

The data focused by this paper will be generated according to the following structure. The relationship between sequences of dependent variable $\{y_t\}$ and an independent variable X_t under different states s across time t is defined by the equation:

$$y_t = \beta_s X_t + \epsilon_{s,t}, \tag{1}$$

where t is a specific point in time, s represents the state at that time, and the model encompasses all possible states $s \in \{1, 2, 3, ..., S\}$. Notice that y_t can be a vector for each period t, hence we have multiple dependent variables. In this model, β_s is the coefficient that measures the impact of the independent variable X_t on the dependent variable y_t in state s, and $\epsilon_{s,t}$ is the error term associated with state s at time t, which follows a normal distribution with mean 0 and variance Σ_s, i.e. $\epsilon_{s,t} \sim N(0, \Sigma_s)$.

The key innovation of this paper is to further elaborate the model on the dynamics of transitioning between states over time. The transition probabilities follows a traditional logit probability model. Specifically, it defines the probability of the t-th observation being in a particular state s, where s ranges from 1 to $S - 1$, with the formula:

$$P(z_t = s|P(z_{t-1})) = \frac{exp(\gamma_s X_t + \sum_{x=1}^{S} \eta_{s,x} \mathbb{1}(z_{t-1} = x))}{1 + \sum_{s=1}^{G-1} exp(\gamma_s X_t + \sum_{x=1}^{G-1} \eta_{s,x} \mathbb{1}(z_{t-1} = x))}. \tag{2}$$

Here, $P(z_t = s|P(z_{t-1}))$ represents the conditional probability that the t-th observation is in state s, given the state at $t - 1$. For observations transitioning into the final state $s = S$, the probability is distinctively given by:

$$(3) \qquad P(z_t = s|P(z_{t-1})) = \frac{1}{1 + \sum_{s=1}^{G-1} exp(\gamma_s X_t + \sum_{x=1}^{S-1} \eta_{s,x} \mathbb{1}(z_{t-1} = x))}.$$

Notice that to incorporate the Markovian properties of the hidden state, the model integrates a series of dummy variables representing the states from the previous period. If the parameters $\{\gamma_s\}$ are all set to zero, the model simplifies to the standard Markov Switching time series models. Suppose all parameters $\{\eta_{s,x}\}$ are zero, then the model will reduces to a cross-sectional model.

B. EM Algorithm

The key methodology of estimation depends on the Expectation-Maximization (EM) Algorithm, a fundamental statistical approach for finding maximum likelihood estimates in models with latent variables. Denote the vector of parameters to be estimated as $\theta = \{vec(\gamma_s), vec(\eta_{s,x}), vec(\beta_s), vec(\Sigma_s)\}$. The process starts with an initial guess of the parameters, setting the stage for the iterative procedure that seeks to refine these estimates to better fit the data.

The first major step within each iteration is the Expectation Step (E-step). During the E-step, the algorithm calculates the expected value of the log-likelihood function, with respect to the current estimate of the parameters. Due the the complexity of the model, finding the correct form of the expected log-likelihood function is the key to this paper and more will be discussed on this in the subsequent subsection.

Following the E-step is the Maximization Step (M-step), where the algorithm finds the parameters that maximize the expected log-likelihood function obtained in the E-step. This step updates the parameter estimates to improve the model fit.

The algorithm iterates between the E-step and M-step until the change in the log-likelihood function (or equivalently, the change in parameter estimates) between successive iterations falls below a predefined tolerance level. This iterative process ensures convergence to a set of parameter estimates that are theoretically guaranteed to approach a local maximum of the likelihood function.

Expectation Step. — In the Expectation Step, the algorithm computes the expected log-likelihood function based on previous parameter estimates. The Generalized Kim's Filter is used to calculate proper probabilities required to plugin to the expected log-likelihood function. Given the estimation of θ in interation $m-1$, the expected log-likelihood function $E\left[\log P(y|X,\theta)|X,y,\theta_{m-1}\right]$ is defined as

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{S}\left(\begin{array}{c}\overbrace{P(z_t=s|X_t,\{y_t,t=1,2,...,T\},\theta_{m-1})}^{\text{Smoothed Probability}}\times \\ \underbrace{\log\left(\underbrace{P(z_t=s|X_t,y_{t-1},\theta)}_{\text{Prior Probability}}\underbrace{P(y_t|X_t,y_{t-1},z_t=s,\theta)}_{\text{Normal PDF}}\right)}_{\text{Log-Likelihood Function}}\end{array}\right)$$

Notice that $P(z_t=s|X_t,\{y_t,t=1,2,...,T\},\theta_{m-1})$ denotes the probability that period t is in state s after observing all the observable data. By definition, this is the smoothed probabilities from the Kim's filter. On the other hand, $P(z_t=s|X_t,y_{t-1},\theta)$ denotes the probability that period t is in state s before period t data is observed. Hence, it denotes the prior probabilities that is given by equation (2) and (3). $P(y_t|X_t,y_{t-1},z_t=s,\theta)$ denotes the probability that suppose period t is in state s, the probability of observing y_t. This is the conditional probability that comes from the normal distribution of the errors $\epsilon_{s,t}$ in equation (1).

The crucial aspect now lies in computing the prior and smoothed probabilities period by period. Drawing inspiration from Kim's filter proposed in citation [3], I have generalized this idea to obtain both probabilities. The Generalized Kim's

Filter consists of two parts. In the first step, we iterate over periods in a forward manner. The forward iteration of the Generalized Kim's Filter computes the prior (or predicted) probability of being in state s before observing y_t according to equations (2) and (3). However, since the true state of the last period is not observable to us, we cannot directly use these two equations. Instead, I use the posterior probabilities that we calculate below for the last period to determine the prior probabilities for the current period.

$$P(z_t|X_t, y_{t-1}, \theta) = \begin{cases} \frac{exp(\gamma_s X_t + \sum_{x=1}^{S} \eta_{s,x} P(z_{t-1}=x|X_{t-1}, y_{t-1}, \theta))}{1+\sum_{s=1}^{G-1} exp(\gamma_s X_t + \sum_{x=1}^{G-1} \eta_{s,x} P(z_{t-1}=x|X_{t-1}, y_{t-1}, \theta))} & \text{if } s < S \\ \frac{1}{1+\sum_{s=1}^{G-1} exp(\gamma_s X_t + \sum_{x=1}^{G-1} \eta_{s,x} P(z_{t-1}=x|X_{t-1}, y_{t-1}, \theta))} & \text{if } s = S \end{cases}$$

Then because we directly observe the dependent variable in period t, we can update the posterior (or filtered) probability of being in state s after observing y_t. The posterior probability $P(z_t = s|X_t, y_t, \theta)$ is updated using Bayes' rule:

$$P(z_t = s|X_t, y_t, \theta) = \frac{P(z_t = s|X_t, y_{t-1}, \theta)P(y_t|X_t, z_t = s, \theta)}{\sum_s P(z_t = s|X_t, y_{t-1}, \theta)P(y_t|X_t, z_t = s, \theta)}$$

Notice that the denominator represents the total probability of observing y_t in period t, while the numerator denotes the probability of observing y_t while the current period is in state s. After calculating the prior and posterior probabilities for perod t, the algorithm iterate forwards and calculate the next period prior and postiors.

After we run the filter forwards for all t, we run the backward iteration of the Generalized Kim's Filter to compute the smoothed probability of being in state s after observing the entire sequence of data $\{y_t, t = 1, 2, ..., T\}$. Suppose we have already calculated the $t+1$ period smoothed probability $P(z_{t+1} = s|\{X_t\}, \{y_t\}, \theta)$. We first calculate the state transition probability matrix P_t^{TX} specific to period t, which represents the probabilities of transitioning from one state to another, given the information of the independent variables, because now the independent variables will affect the transitioning probabilities. The transitioning probabilities is an $S \times S$ matrix, with row i column j of it being the probability of moving from state i to state j from period t to $t+1$. Denote this matrix as P_t^{TX}.

The smoothed probability for period t is then updated using the transition probabilities and the $t+1$ period smoothed probabilities, again according to Bayes' Rule:

$$P(z_t = s|X_t, y_t, \theta) \sum_i \frac{P_{i,j,t}^{TX} P(z_{t+1} = j|\{X_t\}, \{y_t\}, \theta)}{\sum_j P_{i,j,t}^{TX} P(z_{t+1} = j|\{X_t\}, \{y_t\}, \theta)}$$

This step is the same as in the usual Kim's filter. The key difference is that now the transition matrix is time varying, and it depends on the value of the time-specific independent variables. The backward process initializes by setting the last period's filtered probability as equal to the last period's smoothed probability. Iterating backwards will result in a sequence of smoothing probabilities that can be used in calculating the expected log-likelihood function as we discussed.

Maximization Step. — The EM algorithm is essentially a MLE algorithm, so the second step of the algorithm is maximizing the expected log-likelihood function that we calculated previously. This estimator, denoted as $\hat{\theta}_t$, is obtained by solving the following optimization problem:

$$\hat{\theta}_t = \text{argmax}_\theta \; E\left[\log P(y|X, \theta)|X, y, \theta_{m-1}\right] + \alpha||\theta_{t-1}||_p$$

subject to the condition that the variance-covariance matrices for the errors in state s, denoted as Σ_s, are positive definite for all $s = 1, 2, \ldots, S$. The term $\alpha||\theta_{t-1}||_p$ represents L^p regularization, which reduces to the traditional MLE when $\alpha = 0$.

In addition to parameter estimation, hyperparameter selection is crucial in model optimization. One such hyperparameter is the number of states in the model. The model with the optimal number of states is typically chosen based on a criterion such as the Bayesian Information Criterion (BIC), defined as:

$$BIC = k\log(T) - 2\log(\hat{L})$$

where k is the number of parameters in the model, T is the number of data points, and $\hat{L}$ is the maximized value of the likelihood function. The model offering the lowest BIC is selected as it indicates the best trade-off between goodness of fit and model complexity.

C. Asymptotic Distribution

Suppose we are dealing with only one dependent variable. In this case, obtaining the standard error of the estimators is straightforward. The asymptotic estimation variance $Var(\hat{\theta})$ can be described by the Law of Large Numbers, following the general rule of MLE estimators. It follows:

$$\frac{1}{T-k}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\partial E\left[\log P(y|X,\theta)|X,y,\theta_{m-1}\right]}{\partial\theta}^{T}\frac{\partial E\left[\log P(y|X,\theta)|X,y,\theta_{m-1}\right]}{\partial\theta}\right)^{-1}$$

However, when dealing with multiple dependent variable variance-covariance matrices Σ_s, certain steps are involved due to the existence of the constraint that the variance-covariance matrix must be positive definite. To address this constraint, we adjust the parameters to be estimated. Instead of directly estimating the variance-covariance matrix, we first use LU decompositions for each of the positive definite matrices:

$$\Sigma_s = U_s U_s^T$$

We estimate $\{U_s\}$ instead of $\{\Sigma_s\}$ because implementing the constraints is simpler in this form. After obtaining the estimations for $\{U_s\}$, the standard error of Σ_s is obtained using the delta method:

$$Var(\hat{\Sigma}_s) = \left(\frac{\partial \Sigma_s}{\partial U}\right)^T Var(\hat{U}) \left(\frac{\partial \Sigma_s}{\partial U}\right)$$

According to the Law of Large Numbers and the rules of general MLE, these should be the asymptotic standard errors for the estimators.

D. Model Predictions

Given the complexity of the model structure, both in-sample and out-of-sample predictions for Markov Switching models are challenging. Drawing inspiration from the idea presented in reference cite 4, this paper treats the prediction task as another form of estimation. The basic idea is that the optimal predictor should maximize the likelihood for each period given all the available information up to that point.

Based on this principle, in-sample prediction, denoted as $\tilde{y}_t$, involves determining the value of y that maximizes the expected log probability of observing y given the input X_t and the estimated parameters $\hat{\theta}$. Mathematically, this is represented as:

$$\tilde{y}_t = \text{argmax}_y \sum_{s=1}^{S} \left(\begin{array}{c} \overbrace{P(z_t = s | X_t, \{y_t\}, \hat{\theta})}^{\text{Smoothed Probability}} \\ \log(P(z_t = s)P(y|X_t, z_t = s, \hat{\theta})) \end{array} \right)$$

It's important to note that smoothed probabilities are utilized in this calculation, considering that for in-sample predictions, we possess information up until period T.

Similarly, for out-of-sample prediction denoted as $\tilde{y}_{T+h}$, the procedure is analogous but is based on the input X_{T+h} and the estimated parameters $\hat{\theta}$. Mathematically, this is represented as:

$$\tilde{y}_{T+h} = \text{argmax}_y \sum_{s=1}^{S} \left(\begin{array}{c} \overbrace{P(z_{T+h}=s|X_{T+h}, y_{T+h-1}, \hat{\theta})}^{\text{Prior Probability}} \\ \log(P(z_{T+h}=s)P(y|X_{T+h}, z_{T+h}=s, \hat{\theta})) \end{array} \right)$$

Here, since we lack full information for period $T+h$ at period T, prior probabilities are employed to compute out-of-sample predictions.

III. Model Simulation

TBD.

IV. Application

Finally, I will demonstrate an application of the MS-GMLR model using real-world data. I modify the model slightly to transform it into a vector autoregressive (VAR) version. Assuming the variables of interest can be organized into an array $\{y_t\}$, the model is represented as:

$$y_t = \sum_{l=1}^{L} A_{s,l} y_{t-l} + \epsilon_{s,t}, \quad \text{for } s \in \{1, 2, ..., S\}$$

Here, L denotes the total number of lags incorporated into the model. It's essential to note that in the MS-GMLR model, these lagged terms also influence the probability of the relationship transitioning into a specific state.

For simplicity, let's consider output growth, inflation rate based on CPI, and the Federal Funds Rate as the three variables included in the VAR model. All data is sourced from FED St. Louis, covering the period from January 1982 to December 2019 at a monthly frequency. As an illustrative example, I select only two states and set the maximum number of lags to three. The impulse response functions (IRFs) of the two states can be derived by iterating the model with three different types of shocks, and the standard error of the IRF can be obtained

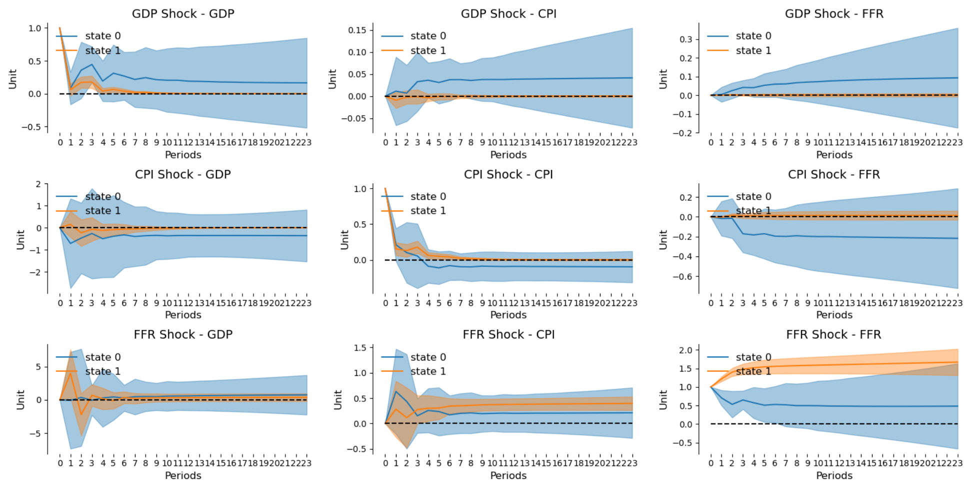

Figure 2. Impluse Response Functions of Two States

using the delta method.

Figure 2 presents the IRFs across two distinct states, revealing notable differences in their behavior. From the analysis of the graph, several key observations emerge. Firstly, in state 0, the Federal Funds Rate shock exhibits significantly less persistence compared to state 1. This suggests that the impact of monetary policy adjustments on the economy and financial markets is more transient in state 0. Secondly, the responses of all three variables—GDP, inflation, and the FFR—to shocks in GDP and inflation are considerably more pronounced during state 0 than in state 1. This indicates a heightened sensitivity of these economic indicators to shocks in state 0, reflecting potentially different underlying economic conditions or policy environments between the two states.

The table 1 presented below illustrates the unique insights offered by the MS-GMLR model. A primary observation is the propensity for the system to persist in the same state from one period to the next when initiated from a specific state initially. However, a more intriguing finding emerges when examining the relationship between economic indicators and state transitions. Specifically, the data suggests that during periods of declining inflation and GDP growth rates, there

Table 1—Logit Model Estimation for State 0

Variables	Parameters	Standard Error	P Value
L1.GDP	-0.0933	0.0491	0.0573
L1.CPI	-0.4736	0.2233	0.0339
L1.FFR	0.3323	0.7772	0.6690
L2.GDP	0.0540	0.0380	0.1550
L2.CPI	0.1610	0.1764	0.3616
L2.FFR	-0.5533	1.0566	0.6005
L3.GDP	-0.0178	0.0359	0.6202
L3.CPI	-0.1635	0.1748	0.3494
L3.FFR	0.7174	0.7245	0.3221
P(L.state 0)	2.4723	0.7047	0.0005
Const	-3.0649	0.6282	0.0000

is an increased likelihood of transitioning into state 0. This correlation strongly suggests that state 0 represents a recessionary phase in the economic cycle. When combined with Figure 2 which displays the impulse response functions, we can infer that economic responses are more pronounced during recessions, with larger movements observed in all variables and a less persistent effect of monetary policy shocks. The MS-GMLR model stands out in explicitly elucidating these relationships, establishing a direct connection between observable economic indicators and the probabilities of transitioning into specific states.

Further insights are gained by examining the smoothed probability history for state 0, as depicted in Figure 3. Prior to 1994, the states exhibit extreme volatility, making it challenging to draw definitive conclusions about the determinants of the world's economic state based solely on data from this period. However, the situation markedly changes after the Federal Reserve's decision to enhance its communication with the public. Post-1994, there is a noticeable stabilization and persistence in the economic state, suggesting that the Fed's communication strategy has had a stabilizing effect on the economy.

State 0 appears to capture periods of economic downturn, as it closely aligns with major recessions, including the 1997 Asian Financial Crisis, the 2001 dot-

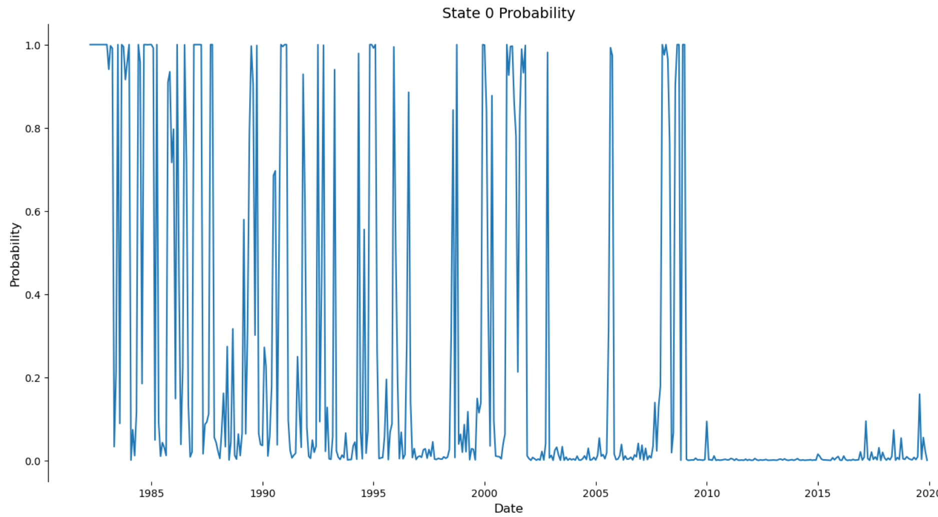

Figure 3. Smoothed Probability History for State 0

com bubble burst, and the 2008 Financial Crisis. This association further reinforces the interpretation of state 0 as indicative of recessionary periods within the economy. The analysis clearly demonstrates the significant impact of the Federal Reserve's public communication on economic stability. By making its policies and outlook more transparent, the Fed has contributed to reducing volatility and uncertainty in the economic landscape, as evidenced by the post-1994 data. The less volatile behavior observed after the initiation of public communication underscores the critical role that central bank transparency and communication play in stabilizing the economy and financial markets.

V. Conclusion

In conclusion, the Markov Switching Gaussian Mixture Linear Regression (MS-GMLR) model emerges as a powerful tool for analyzing time series data characterized by complex interdependencies. Through its comprehensive framework, the MS-GMLR model enables researchers to delve into intricate patterns within empirical data, particularly evident in its application to Vector Autoregression

(VAR) analysis. One notable finding from this empirical exploration is the indication that economic states may undergo transitions during recessionary phases. This insight underscores the model's capacity to uncover nuanced shifts in the economic landscape, offering valuable insights for policymakers and analysts alike.