A Nonparametric Endogenous Switching Model with an Application to Macroeconomics

Robert MacDonald*
University of California, Irvine

Abstract

This paper proposes a regime-switching linear model with time-varying transition probabilities, endogenous switching, and a nonparametric error distribution. The last two qualities are achieved by letting the conditional mean of the normalized observation errors be a potentially nonlinear function of the errors in the state equation. We demonstrate that this specification permits a very flexible marginal distribution for the observation error. A Markov Chain Monte Carlo algorithm for sampling from the posterior distribution of parameters is developed. A simulation study demonstrates that existing parametric switching models yield biased parameter estimates when the data is generated by a model with nonlinear endogenous switching. We apply the model to US quarterly output growth. The proposed model is shown to fit the data better than parametric switching models.

JEL classification: C14, C15, C34, C32, C51, C58, E32
Keywords: Nonparametric Methods, Markov Chain Monte Carlo, Regime Switching, Structural Change, Macroeconometrics, Business Cycle

---

*Department of Economics, University of California-Irvine, 3151 Social Sciences Plaza, Irvine, CA 92697-5100; email: rmacdon1@uci.edu.
I would like to thank my advisor, Ivan Jeliazkov, for his invaluable feedback. I would also like to thank the participants in the NBER-NSF Seminar on Bayesian Inference in Economics and Statistics and the California Econometrics Conference for their helpful comments and advice.

1 Introduction

Until the past decade, time series econometrics has focused primarily on parametric models. This was true of both linear vector autoregressions (VARs) (Sims, 1980; Litterman, 1986; Primiceri, 2005; Koop and Korobilis, 2013) and mixture models (Beaudry and Koop, 1993; Sims and Zha, 2006; Uribe and Lopes, 2020). Early work in nonparametric time series models focused on approximating nonlinear conditional mean functions in either univariate or small multivariate processes (Auestad and Tjøstheim, 1990; Härdle et al., 1998; Hamilton, 2001). A good deal of the more recent work has focused on Dirichlet process mixture models (DPMs). DPMs have been used to model the error distribution of asset returns in stochastic volatility models (Jensen and Maheu, 2010; Delatola and Griffin, 2011). Shahbaba (2009) used the DPM to identify regimes in U.S. real GDP growth, allowing the number of regimes to be selected by the model. Kalli and Griffin (2018) use a DPM to flexibly model VAR processes. Nonparamteric VARs are an active area of research. Jeliazkov (2013) modeled the conditional mean of each dependent variable as a sum of nonlinear univariate functions of explanatory variables. Huber and Rossini (2021) model the conditional mean of a VAR process using Bayesian additive regression trees. A consistent finding in the nonparametric VAR papers is that relaxing the assumption of linearity leads to better forecasting performance.

One of the primary methodologies for introducing nonlinearity to time series econometrics has been the class of models known as Markov-switching models (MSMs). MSMs are an extension of the hidden Markov model (Baum and Petrie, 1966) to the case of a continuously-distributed dependent variable. They allow model parameters to switch between different regimes. MSMs were introduced by Goldfeld and Quandt (1973) and popularized by Hamilton (1989), who used a model of US Gross National Product growth as an alternative method for dating business cycle turning points. Since then, models with Markov-switching have been applied extensively to business cycles (Albert and Chib, 1993; Boldin, 1996; Ghysels et al., 1997; Chauvet and Hamilton, 2006) as well as financial data (Vigfusson, 1997; Haas et al. 2004; Guidolin and Timmermann, 2005). MSMs have expanded to include models with time-varying transition probabilities (TVTP) (Diebold et al. 1994; Filardo, 1994; Filardo and Gordon, 1998) as well as state space models (Kim, 1994; Chauvet, 1998; Kim and Nelson, 1999). A more recent class of models allows for endogenous switching (Chib and Dueker, 2004; Kim et al, 2008; Hwu et al., 2019; Kang and Kim, 2020). These models allows the innovations in the equations governing regime transitions to be correlated with innovations in the observation equation. Flexible error distributions are almost entirely missing in the MSM literature. A Normal error distribution for the observation equation is assumed in virtually all models. Two notable exceptions are Dueker (1997) and Hwu (2018). The former modeled stock returns using a student's t-distribution where the degrees of freedom switch between different regimes. Hwu (2018) is the only MSM we have found where the observation errors have a nonparametric distribution. He develops a switching mean model where the error distribution is generated by a Dirichlet process. Hwu (2018) assumes that switching is exogenous and transition probabilities are constant.

This paper proposes a regime-switching linear model with TVTP, endogenous switching, and a nonparametric error distribution. Both of these qualities are achieved by letting the conditional mean of the normalized observation errors be a potentially nonlinear function of the errors in the state equation. Our model differs from Hwu (2018) both in the formulation of the error distribution and robustness to endogeneity and TVTP.

The rest of the paper is organized as follows. Section 2 outlines the proposed model. Section 3 describes how samples from the posterior distribution of model parameters are simulated using Markov Chain Monte Carlo (MCMC) methods. Section 4 reports the results of a simulation study. Section 5 describes model comparison using Bayes factors. Section 6 applies the model to US output growth data. Section 7 concludes.

2 The Proposed Model

2.1 Model Setup

Consider the model

$$y_t = x_t'\beta_{s_t} + \sigma_{s_t}\varepsilon_t, \tag{1}$$

$$s_t = 1\{s_t^* > 0\}, \tag{2}$$

$$s_t^* = z_t'\delta_{s_{t-1}} + \eta_t, \tag{3}$$

$$\varepsilon_t \sim N(g(\eta_t), 1), \;\; \eta_t \sim N(0, 1). \tag{4}$$

When $g(\eta_t)$ is linear, $g(\eta_t) = \rho\eta_t$, this model is observationally equivalent to the endogenous switching model of Kim et al. (2008). This can be seen by rewriting the model as

$$y_t = x_t'\beta_{s_t} + \tilde{\sigma}_{s_t}\tilde{\varepsilon}_t, \tag{5}$$

$$\tilde{\sigma}_{s_t} = \sigma_{s_t}\sqrt{1+\rho^2}, \tag{6}$$

$$\tilde{\varepsilon}_t = \frac{\varepsilon_t}{\sqrt{1+\rho^2}}, \tag{7}$$

$$\begin{bmatrix} \tilde{\varepsilon}_t \\ \eta_t \end{bmatrix} \sim N(0_2, \Omega), \tag{8}$$

$$\Omega = \begin{bmatrix} 1 & \tilde{\rho} \\ \tilde{\rho} & 1 \end{bmatrix}, \tag{9}$$

$$\tilde{\rho} = \frac{\rho}{\sqrt{1+\rho^2}}. \tag{10}$$

The formula for $\tilde{\rho}$ guarantees that Ω is positive definite.

To estimate the model, $g(\eta_t)$ is approximated nonparametrically. Let $g(\eta_t) \approx \hat{g}(\eta_t) = \sum_{n=1}^{p} \rho_n b_n(\eta_t) = \rho' b_t$, where $\{b_n(\eta_t)\}$ are basis functions. The basis functions are normalized to equal 0 at the origin. Without this normalization, the intercept would not be jointly identified with $\hat{g}(\eta_t)$. This was a natural choice of normalization because it means there is no impact from endogeneity when $\eta_t = 0$, just as in a parametric enodogenous switching model. The two types of approximations that we considered were polynomial series and regression splines. We only report results for regression splines because they consistently performed comparably to or better than series regression. The polynomial series specification worked well when $g(\eta_t)$ was also polynomial, as would be expected, but less so for other types of functions. Note that when a linear spline is used to approximate $g(\eta_t)$, the joint distribution of ε_t and η_t becomes a mixture of disjoint truncated Normal distributions. For any other approximation, the joint distribution is nonstandard.

2.2 The Implications of $g(\eta_t)$ for the Marginal Distribution of ε_t

The general form of $g(\eta_t)$ allows for great flexibility in the marginal distribution of ε_t. Figure 1 contains plots of $f(\varepsilon_t)$ under various conditional mean functions. It demonstrates that we can induce skewness (1.a and 1.b), excess kurtosis (1.c), and bimodality (1.d) using simple functional forms for $g(\eta_t)$. The reader will observe that neither the unconditional mean nor the unconditional variance are constant with respect to $g(\eta_t)$. This is in contrast to the various types of parametric endogenous switching models (Chib and Dueker, 2004; Kim et al., 2008; Hwu et al., 2019). The existing literature models (ε_t, η_t) as a multivariate Normal random variable. This gives the marginal distribution of ε_t the same mean and variance regardless of the correlation structure. In our specification, a greater correlation between ε_t and η_t implies a greater marginal variance of ε_t. We considered marginal moment restrictions on $f(\varepsilon_t)$, namely $E[\varepsilon_t] = 0$ and $Var[\varepsilon_t] = 1$. However, accommodating these restrictions is difficult when using any approximation other than a local polynomial; numerical integration is required to find the mean and variance of $\hat{g}(\eta_t)$. In addition, direct sampling from the full conditional posterior distribution of ρ would no longer be possible. Let $\Theta \equiv \{\beta = \{\beta_j\}, \sigma = \{\sigma_j\}, \delta = \{\delta_j\}, \rho\}$. The dependence between the degree of endogeneity and the marginal variance does not restrict the model overall because $Var[y_t|x_t, s_t, \Theta] = \sigma^2_{s_t} Var[\varepsilon_t]$. $g(\eta_t)$ determines the degree and nature of the endogeneity, while σ controls the variance of the error term. A possible way to weaken this relationship between endogeneity and unconditional variance is to use the alternative model

$$y_t = x_t' \beta_{s_t} + \varepsilon_t, \tag{11}$$

$$\varepsilon_t \sim N(g(\eta_t), \sigma^2_{s_t}), \;\; \eta_t \sim N(0, 1). \tag{12}$$

However, this model offers less flexibility with regard to within-regime variance. Under the alternative model, $Var[y_t|x_t, s_t, \Theta] = Var[g(\eta_t)] + \sigma^2_{s_t}$, as opposed to $\sigma^2_{s_t}(Var[g(\eta_t)] + 1)$ in the proposed model. Under the alternative model, we have different degrees of endogeneity depending on the regime. This last feature may be a desirable property, which is why we are currently researching the relative strengths of the alternative model.

3 Posterior Sampling

3.1 Sampling S_T and S_T^*

Let $Y_t \equiv (y_1, \ldots, y_t)'$, $S_t \equiv (s_1, \ldots, s_t)'$, $S_t^* \equiv (s_1^*, \ldots, s_t^*)'$. Regardless of the functional form of $g(\eta_t)$, the filtered regime probability can be calculated using

$$P(y_t, s_t|s_{t-1}) = \int_{\mathcal{B}_{s_t|s_{t-1}}} f(y_t|s_t, \eta_t) f(\eta_t) d\eta_t, \tag{13}$$

$$P(s_t|Y_t) \propto \sum_{s_{t-1}} P(y_t, s_t|s_{t-1}) P(s_{t-1}|Y_{t-1}). \tag{14}$$

$B_{s_t|s_{t-1}}$ is the region of integration where the values of η_t are consistent with s_t and s_{t-1}. The constant of proportionality can be obtained by summation over s_t. This enables a straightforward implementation of the algorithm of Chib (1996) for sampling the entire history of regimes as a single block. Numerical integration is performed using the trapezoid method and a fine grid of 500 points. Gauss Legendre quadrature would typically be a superior choice to the trapezoid method as it allows for exact integration of finite order polynomials and requires fewer function evaluations. The trapezoid method was chosen because function evaluations can be saved and reused in sampling S_T^*.

S_T and S_T^* are sampled as a single block by first sampling S_T marginally of S_T^* and then drawing from $\pi(S_T^*|S_T, \Theta, Y_T)$. Conditional on s_t, s_t^* can be drawn independently from the posterior $\pi(S_t^*|S_t, \Theta, Y_T)$ using a Metropolis Hastings step. We obtain near iid samples from the full conditional posterior using a Griddy Gibbs proposal density (Tierney, 1994). The posterior is first discretized by evaluating $f(y_t, S_t^*|s_t, \Theta)$ over an evenly-spaced grid. The discrete probability measure is calculated as

$$P(x_i) = \frac{f(y_t, x_i|s_t, \Theta)}{\sum_k f(y_t, x_k|s_t, \Theta)} \tag{15}$$

.

A candidate $s_t^{*\prime}$ is obtained by drawing x_i from the discrete distribution and then adding a continuous random variable:

$$s_t^{*\prime} = x_i + u, \quad u \sim N(0, \sigma_u^2). \tag{16}$$

The proposal density for $s_t^{*\prime}$ is then obtained by summation of $(s_t^{*\prime}, x_i)$ over the discrete component:

$$q(s_t^{*\prime}) = \sum_i P(x_i) f_N(s_t^{*\prime} - x_i, 0, \sigma_u^2). \tag{17}$$

Proposed draws are then accepted with the usual MH acceptance probability.

We follow the common practice in Bayesian MSMs of rejecting samples where s_t is constant over all periods. Accepting such draws can cause the sampler to get stuck in a particular region of the parameter space and mix very slowly. Chib (1996) pointed out that this restriction is not necessary if all priors are proper. Another quirk of Bayesian MSMs of which we must be mindful is label switching (Fruhwirth-Schnatter, 2001). This problem arises in parameter simulation because an unconstrained model with N regimes produces a likelihood with $N!$ modes. Failing to account for label switching can lead to nonsensical parameter estimates if one simply uses the sample mean. One solution is to use identifying restrictions, such as order restrictions on the intercepts or variances.

As η_1 depends on s_0, it must either be sampled or the dependence of η_1 on s_0 must be integrated out. We elect to sample s_0. Since there is no corresponding y_0 for s_0, it can be sampled analytically from its full conditional distribution. Let $\eta(s_0) \equiv s_1^* - z_1' \delta_{s_0}$. The full conditional distribution of s_0 can then be written as

$$P(s_0 | Y_T, S_{-0}) \propto f(y_1 | s_1, \eta(s_0)) f(s_1^* | s_0) \pi(s_0) \tag{18}$$

The constant of proportionality is obtained by summing over all values of s_0. $\pi(s_0)$, the unconditional probability of s_0, can be estimated in several ways. One common approach is to use the stationary distribution of the Markov chain (Chib, 1993; Chib, 1996). This becomes more complicated when transition probabilities are non-constant. If the variables in z_t are stationary, stationary transition probabilities can approximated by plugging the sample mean of z_t into the equation for s_t^* (Hwu et al., 2019). However, the approximation is invalid when nonstationary variables like time trends are included. Another solution is to let $\pi(s_0 = 1)$ be a parameter with prior distribution $\pi(s_0 = 1) \sim \mathcal{B}(p_1, p_2)$. One can then sample from the full conditional distribution

$$\pi(s_0 = 1)|Y_T, S_T \sim \mathcal{B}(p_1 + 1 - s_0, p_2 + s_0). \tag{19}$$

We use this specification in all estimations that follow.

3.2 Sampling β and ρ

Once we condition on S_T, S_T^*, and δ, the model for Y_T becomes linear. We assume the conjugate priors

$$\beta \sim N(b_0, B_0), \tag{20}$$

$$\rho \sim N(r_0, R_0). \tag{21}$$

We use the hierarchical prior

$$R_0 = \tau_\rho^2 diag(\nu_1, \ldots, \nu_p), \tag{22}$$

$$\tau_\rho^2 \sim IG(\alpha_\rho/2, \gamma_\rho/2). \tag{23}$$

τ_ρ thus acts as a global smoothness parameter. It can be sampled from the full conditional posterior

$$\tau_\rho^2 \sim IG\left(\frac{\alpha_\rho + p}{2}, \frac{\gamma_\rho + \rho'(diag(\nu_1, \ldots, \nu_p))^{-1}\rho}{2}\right). \tag{24}$$

We set $\nu_1, \nu_2, \nu_p = 1$. For other entries, we set $\nu_i = k_i - k_{i-1}$. k_i is the i^{th} knot. Knots are set such that they are evenly spaced across standard Normal quantiles for linear splines and multiples of quantiles for higher order splines.

An equivalent way of writing (1) is

$$y_t = \begin{bmatrix} x_t' & s_t x_t' & \sigma_{s_t} b_t' \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 - \beta_0 \\ \rho \end{bmatrix} + \sigma_{s_t}\varepsilon_t^\dagger = x_{s_t}'\beta^* + \sigma_{s_t}\varepsilon_t^\dagger, \tag{25}$$

$$\varepsilon_t^\dagger \sim N(0,1). \tag{26}$$

Let $X_{S_T} \equiv (x_{s_1}, \ldots, x_{s_T})'$ and $\Sigma_{S_T} \equiv \text{diag}(\sigma_{s_1}^2, \ldots, \sigma_{s_T}^2)$. The likelihood can then be written as

$$f(Y_T|\Theta, S_T, S_T^*) = f_N(Y_T|X_{S_T}\beta^*, \Sigma_{S_T}). \tag{27}$$

We then arrive at the full conditional posterior for a classical linear model with heteroskedasticity.

$$\beta^*|Y_T, \Theta_{-\beta^*}, S_T, S_T^* \sim N(\hat{b}^*, \hat{B}^*), \tag{28}$$

$$\hat{B}^* = (B_0^{*-1} + X_{S_T}'\Sigma_{S_T}^{-1}X_{S_T})^{-1}, \tag{29}$$

$$\hat{b}^* = \hat{B}^*(B_0^{*-1}b_0^* + X_{S_T}'\Sigma_{S_T}^{-1}y), \tag{30}$$

$$b_0^* \equiv (b_0', r_0')', \tag{31}$$

$$B_0^* \equiv \begin{bmatrix} B_0 & 0_{2k\times p} \\ 0_{p\times 2k} & R_0 \end{bmatrix}. \tag{32}$$

When β is unrestricted, the entire vector β^* can be sampled at once. When an identifying restriction is placed on the intercepts, we will use the normalization $\beta_{11} > \beta_{01}$, where β_{j1} is the intercept for $s_t = j$. This leads to the full conditional posterior

$$\beta^*|Y_T, \Theta_{-\beta^*}, S_T, S_T^* \sim TN_{\beta^*_{k+1>0}}(\hat{b}^*, \hat{B}^*). \tag{33}$$

Since only one dimension of β^* is truncated, the marginal distribution of β^*_{k+1} is

$$\beta^*_{k+1}|Y_T, \Theta_{-\beta^*}, S_T, S_T^* \sim TN_{(0,\infty)}(\beta^*_{k+1}|\hat{b}^*_{k+1}, \hat{B}^*_{k+1,k+1}). \tag{34}$$

A well known result is that is the conditional distributions from a multivariate truncated Normal distribution are also truncated Normal distributions. This fact, combined with the lack of truncation for β^*_{-k+1}, tells us that $f(\beta^*_{-k+1}|\beta^*_{k+1}, Y_T, \Theta_{-\beta^*}, S_T, S_T^*)$ is a multivariate Normal density. We can then sample from $f(\beta^*|Y_T, \Theta_{-\beta^*}, S_T, S_T^*)$ by first sampling from $f(\beta^*_{k+1}|Y_T, \Theta_{-\beta^*}, S_T, S_T^*)$ and then from $f(\beta^*_{-k+1}|\beta^*_{k+1}, Y_T, \Theta_{-\beta^*}, S_T, S_T^*)$.

3.3 Sampling σ

Sampling σ_j is complicated by the nonstandard manner in which it enters the likelihood. The full conditional distribution takes the form

$$f(\sigma_j|Y_T, \Theta_{-\sigma_j}, S_T, S_T^*) \propto \pi(\sigma_j) \prod_{s_t=j} f_N(y_t|x_t'\beta_j + \sigma_j\rho' b_t, \sigma_j^2). \tag{35}$$

If just σ_j entered the conditional mean parameter of the likelihood, we could use a Normal prior for σ_j and sample it from a Normal full conditional posterior distribution. If just σ_j^2 entered the conditional variance parameter of the likelihood, we could use an Inverse-Gamma prior for σ_j^2 and sample it from an Inverse-Gamma full conditional posterior distribution. However, the appearance of both σ_j in the conditional mean and σ_j^2 in the conditional variance makes iid sampling from the full conditional posterior infeasible. Luckily, MH sampling with a tailored proposal density is a simple task. The mode of the full conditional distribution of σ_j has an analytical solution for several choices of prior distribution, including Gamma, Inverse-Gamma, and Generalized inverse Gaussian distributions. One first samples from $q(\sigma_j') = f_{t_\nu}(\sigma_j'|\hat{\sigma}_j, c\hat{V}_j)$. $\hat{\sigma}_j$ is the mode of the full conditional posterior. $\hat{V}_j$ is the negative inverse of the second derivative of the log posterior distribution evaluated at $\hat{\sigma}_j$. ν and c are positive tuning parameters. σ_j' is then accepted with probability

$$\alpha = min\left\{1, \frac{q(\sigma_j)\pi(\sigma_j')\prod_{s_t=j} f_N(y_t|x_t'\beta_j + \sigma_j'\rho' b_t, \sigma_j'^2)}{q(\sigma_j')\pi(\sigma_j)\prod_{s_t=j} f_N(y_t|x_t'\beta_j + \sigma_j\rho' b_t, \sigma_j^2)}\right\}. \tag{36}$$

3.4 Sampling δ

δ_j enters the likelihood in a highly nonlinear fashion via the vector of basis functions. This removes the option of analytical sampling that is present in exogenous and parametric endogenous models. As well, there is no general closed form solution for the mode of the full conditional posterior density. This leaves one with MH sampling and either a tailored proposal distribution that is found numerically or a random walk proposal distribution. A random walk proposal distribution is used in all estimations that follow. Let the prior distribution for $\pi(\delta_j) = f_N(d_{0j}, D_{0j})$. A candidate δ_j' is draw from $q(\delta_j'|\delta_j) = f_N(\delta_j'|\delta_j, \tau_{\delta_j}^2)$. Let $\tilde{\eta}_t \equiv s_t^* - z_t'\delta_j'$. δ_j' is then accepted with probability

$$\alpha = min\left\{1, \frac{\pi(\delta_j')\prod_{s_{t-1}=j} f(y_t|s_t, \tilde{\eta}_t)f(\tilde{\eta}_t)}{\pi(\delta_j)\prod_{s_{t-1}=j} f(y_t|s_t, \eta_t)f(\eta_t)}\right\}. \tag{37}$$

Bayesian MSMs sometimes require a strong prior for the transition probabilities for the model to be well-identified. In a model with 2 regimes and fixed transition probabilities, we can select priors to match our expectations about the average duration of a regime (Chib, 1996; Filardo and Gordon, 1998).

4 Simulation Study

This section presents the results of a simulation study. We generated 500 datasets with $T = 500$ observations and $k = 3$ variables: an intercept and two variables each drawn from $N(0_T, I_T)$ distributions. A sample of 13,000 draws from the posterior distribution of parameters was obtained for each dataset. Given the earlier discussion of mixing, this may seem like an insufficiently small sample size. However, we observed much faster mixing of the posterior distribution for the simulated datasets than with the output growth dataset used later. We used the identifying restriction $\beta_{01} < \beta_{11}$. In each instance, the first 3,000 draws were discarded as burn-in. For comparison, we also estimated models with parametric endogenous switching and exogenous switching. All datasets were simulated using the function $g(\eta_t) = \eta_t^2$. Table 1 reports the parameter estimation errors for all three models. The estimation error of a parameter is taken to be the parameter estimate minus the true parameter value. The results demonstrate that the existing models can produce biased estimates when $g(\eta_t)$ is nonlinear. A surprising result is that the exogenous model outperforms the parametric endogenous model. The large estimation errors of the parametric endogenous model are partly due to bimodality in the empirical error distribution. A natural cubic spline with 9 knots was used to approximate $g(\eta_t)$. Figure 2 shows that $g(\eta_t)$ is well-approximated by posterior estimates.

5 Model Comparison

The different models considered in this paper are compared using Bayes factors (Kass and Raftery, 1995). Since the sampler uses a mix of Gibbs and MH steps, marginal likelihood calculation is done using methods form Chib (1995), Chib (1998), and Chib and Jeliazkov (2001). Bayes factors have also been employed to select the number of regimes in both classical MSMs (Koop and Potter, 1999) and in models with endogenous switching (Kang, 2014). As in Chib (1995), the formula for the marginal likelihood is obtained from a simple application of Bayes' Formula:

$$f(Y_T|\mathcal{M}_i) = \frac{f(Y_T|\Theta^*, \mathcal{M}_i)\pi(\Theta^*|\mathcal{M}_i)}{f(\Theta^*|Y_T, \mathcal{M}_i)}. \tag{38}$$

Θ^* is taken to be the posterior mean of Θ. The likelihood $f(Y_T|\Theta^*, \mathcal{M}_i)$ is calculated using a modified version of the forward filtering algorithm of Hamilton (1989). $\pi(s_0 = 1)$ can be integrated out of the likelihood by replacing it with its prior mean. $f(\Theta^*|Y_T, \mathcal{M}_i)$ is rewritten as $f(\beta^*|\sigma^*, \delta^*, Y_T, \mathcal{M}_i)f(\sigma^*|\delta^*, Y_T, \mathcal{M}_i)f(\delta^*|Y_T, \mathcal{M}_i)$. All ordinates are estimated via simulation as in Chib and Jeliazkov (2001).

6 Application to GDP Data

We applied the model of (1) - (4) to data on quarterly real GDP growth. The dataset runs from 1947:Q2 to 2019:Q4. The most recent recession was omitted because the magnitudes of the shifts are much greater than in the rest of the sample. Estimations that included this period did not perform well at identifying previous recessions. They tended to classify every period prior to 2020:Q1 as an expansion. Let the dependent variable be defined as $y_t \equiv ln(GDP_t) - ln(GDP_{t-1})$. We estimated a 2-state switching means model with constant scaling factor σ:

$$y_t = \beta_{s_t} + \sigma\varepsilon_t. \tag{39}$$

We experimented with different autoregressive specifications, allowing for up to 4 lags of y_t and switching scaling factors. However, the simple switching means model with constant scaling factor performed the best in identifying latent states that correspond to business cycles. $g(\eta_t)$ is approximated using a natural cubic spline with 9 knots. Identification is achieved through the restriction $\beta_0 < \beta_1$. We also estimated a parametric endogenous model and an exogenous model. We used the prior $\pi(\sigma) = f_{TN_{(0,\infty)}}(\sigma|0,1)$ and the hyperpriors $b_0 = (-.1815, .4196, 0_p')'$, $B_0 = .25I_2$, $\alpha_\rho = 1$, $\gamma_\rho = .1$, $d_0 = (-.6, 1.66)'$, $D_0 = I_2$, $p_1, p_2 = 0$. We set $p = 1$ and $p = 0$ for the parametric endogenous and exogenous models, respectively. b_0 was chosen to match the average growth rates during recessions and expansions as classified by the National Bureau of Economic Research (NBER). d_0 was chosen to match the average durations of recessions and expansions. Each sample of parameters consisted of 300,000 draws after burn-in samples were discarded.

As can be seen in Figure 4, the posterior estimate of $\hat{g}(\eta_t)$ is rather nonlinear. To better understand how nonlinearity in the conditional mean of ε_t affects its marginal distribution, we estimated the densities $f(\varepsilon_\tau|Y_T)$ and $\{f(\varepsilon_\tau|Y_T, s_\tau, s_{\tau-1})\}$. The subscript τ is used instead of t to stress that these distributions are not conditioned on any time period in the sample. We would ideally remove the dependence of ε_τ on ρ and δ through direct integration:

$$f(\varepsilon_\tau|Y_T) = \int f(\varepsilon_\tau|\eta_\tau, Y_T, \rho) f(\eta_\tau) f(\rho|Y_T) d\eta_\tau d\rho, \tag{40}$$

$$f(\varepsilon_\tau|Y_T, s_\tau, s_{\tau-1}) = \int f(\varepsilon_\tau|\eta_\tau, Y_T, \rho) f(\eta_\tau|Y_T, s_\tau, s_{\tau-1}, \delta) f(\rho, \delta|Y_T) d\eta_\tau d\rho d\delta. \tag{41}$$

The intractability of these integrals forces us to instead use a mixture of numerical and monte carlo integration. At each iteration m of the MCMC sampler, we evaluate the integrals $\int f(\varepsilon_\tau|\eta_\tau, Y_T, \rho^{(m)}) f(\eta_\tau) d\eta_\tau$ and $\int f(\varepsilon_\tau|\eta_\tau, Y_T, \rho^{(m)}) f(\eta_\tau|Y_T, s_\tau, s_{\tau-1}, \delta^{(m)}) d\eta_\tau$ using numerical methods. The reader should note that $f(\eta_\tau) = f_N(\eta_\tau|0,1)$ and $f(\eta_\tau|Y_T, s_\tau, s_{\tau-1}, \delta^{(m)})$ is a truncated standard Normal density with region of truncation $\mathcal{B}^{(m)}_{s_t|s_{t-1}}$. The rest of the integration is done by averaging over MCMC draws. We use the approximations

$$f(\varepsilon_\tau|Y_T) \approx M^{-1}\sum_{m=1}^{M}\int f(\varepsilon_\tau|\eta_\tau, Y_T, \rho^{(m)})f(\eta_\tau)d\eta_\tau, \tag{42}$$

$$f(\varepsilon_\tau|Y_T, s_\tau, s_{\tau-1}) \approx M^{-1}\sum_{m=1}^{M}\int f(\varepsilon_\tau|\eta_\tau, Y_T, \rho^{(m)})f(\eta_\tau|Y_T, s_\tau, s_{\tau-1}, \delta^{(m)})d\eta_\tau. \tag{43}$$

M is the number of remaining MCMC draws after burn-in samples are discarded. Approximations $\hat{f}(\varepsilon_\tau|Y_T)$ and $\hat{f}(\varepsilon_\tau|Y_T, s_\tau, s_{\tau-1})$ are plotted in Figures 5 and 6, respectively. $\hat{f}(\varepsilon_\tau|Y_T)$ is skewed to the right, making extreme positive values more likely than in a Gaussian distribution. We observe interesting deviations from $\hat{f}(\varepsilon_\tau|Y_T)$ when we condition on past and current regimes. $\hat{f}(\varepsilon_\tau|Y_T, s_\tau = 0, s_{\tau-1} = 0)$ is postively skewed and centered around a positive number, meaning we are more likely to see positive deviations from the average growth rate during a recession. The distribution of errors in $\hat{f}(\varepsilon_\tau|Y_T, s_\tau = 0, s_{\tau-1} = 1)$ is more Gaussian, but there is more mass in the positive region of ε_τ. This implies that average growth is higher in the first period of a recession. We can interpret this as a transitional period between high and low growth. We also see a large amount of probability mass in the positive region of ε_τ for $\hat{f}(\varepsilon_\tau|Y_T, s_\tau = 0, s_{\tau-1} = 1)$. This corresponds to a high growth recovery in which average growth is higher in the first quarter following a recession. $\hat{f}(\varepsilon_\tau|Y_T, s_\tau = 1, s_{\tau-1} = 1)$ is the density that most closely resembles a Gaussian distribution centered at 0. This results from the quasilinear shape of $\hat{E}[\hat{g}(\eta_t)|Y_T]$ in the region [-1,.5].

Parameter estimates are displayed in Table 2. Estimates for β are lower for the nonparamteric model than the other two. This is likely caused by the positive skew in $f(\varepsilon_\tau|Y_T)$. The posterior estimate for σ is also lowest in the nonparamteric model, indicating that there is less residual variation in the data when we allow ε_t to have a nonlinear conditional mean function. The estimate for δ_0 is highest in the nonparametric model, corresponding to less persistent recessions. There does not appear to be a large variation in the persistence of expansions predicted by the three models.

Figure 3 shows smoothed recession probabilities from the nonparametric model along with NBER recession dates. We observes a spike in recession probabilities during every recession. The one false positive occurs in the first quarter of the sample. This is a reasonable error for the model to produce because real output growth was negative in this period.

The natural logarithm of the Bayes factors for choosing the Nonparametric model over the exogenous model and the parametric endogenous model are 5.64 and 4.03, respectively. Using an uninformative uniform prior for model probabilities, this makes the posterior probability of the nonparametric model 281.46 times that of the exogenous model and 56.26 times that of the parametric endogenous model.

7 Conclusion

We developed a MSM with nonparametric endogenous switching. The nonparametric model offers substantial flexibility with regard to the marginal distribution of observation errors. A simulation study demonstrated that exisiting parametric models can produce biased results when the true data generating process entails nonlinear endogenous switching. The model was applied to data on US real GDP growth. The estimated model had a significantly higher Bayes factor than estimates for parametric endogenous and exogenous models. The nonparametric model is also able to identify all recessions as calculated by the NBER. Estimated marginal error distributions indicated the innovations in the observation equation are non-Gaussian and generally skewed to the right.

There are many areas in which this paper could be extended. One obvious application is to financial data. A flexible error distribution is called for in a landscape where fat tails and skewness are expected. Other directions for further research are extensions to multivariate data and more than 2 regimes.

References

Albert, J. H. and S. Chib. 1993. Bayes Inference via Gibbs Sampling of Autoregressive Time Series Subject to Markov Mean and Variance Shifts, *Journal of Business & Economic Statistics* 11, 1-15.

Auestad, B. Tjøstheim, D. 1990. Identification of nonlinear time series: First order characterization and order estimation. *Biometrika* 77, 669-687.

Baum, L. E. and T. Petrie. 1966. Statistical Inference for Probabilistic Functions of Finite State Markov Chains. *Annals of Mathematical Statistics* 37, 1554-1563.

Beaudry, P. and Koop, G. 1993. Do Recessions Permanently Change Output? *Journal of Monetary Economics* 31, 149–163.

Boldin, M. D. 1996. A check on the robustness of Hamilton's Markov-switching model approach to the economic analysis of the business cycle. *Studies in Nonlinear Dynamics and Econometrics* 1, 35–46.

Chauvet, M. 1998. An Econometric Characterization of Business Cycle Dynamics with Factor Structure and Regime Switching. *International Economic Review* 39, 969–996.

Chib, S. 1995. Marginal Likelihood from the Gibbs Output. *Journal of the American Statistical Association* 90, 1313–1321.

Chib, S. 1996. Calculating the posterior distributions and modal estimates in Markov mixture models, *Journal of Econometrics* 75, 79-97.

Chib, S. 1998. Estimation and comparison of multiple change-point models, *Journal of Econometrics* 86, 221-241.

Chib, S. and M. Dueker. 2004. Non-Markovian regime switching with endogenous states and time-varying state strengths. Working Paper 2008-31, Federal Reserve Bank of St Louis.

Chib, S. and I. Jeliazkov. 2001. Marginal likelihood from the Metropolis–Hastings output. *Journal of the American Statistical Association* 96, 270–81.

Delatola, E.I. and J.E. Griffin. 2011. Bayesian nonparametric modelling of the return distribution with stochastic volatility. *Bayesian Analysis* 6, 901–926.

Diebold, F. X., J.-H. Lee and G. C. Weinbach. 1994. Regime switching with time-varying transition probabilities. In: C. Hargreaves (ed.), *Nonstationary Time Series Analysis and Cointegration*, pp. 283–302. Oxford, New York: Oxford University Press.

Dueker, M. 1997. Markov Switching in GARCH Processes and Mean-Reverting Stock Market Volatility. *Journal of Business and Economic Statistics* 15, 26-34

Filardo, A. J. 1994. Business cycle phases and their transitional dynamics. *Journal of Business and Economic Statistics* 12, 299–308.

Filardo, A. J. and S. F. Gordon. 1998. Business cycle durations. *Journal of Econometrics* 85, 99–123.

Fruhwirth-Schnatter, S. 2001. Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models. *Journal of the American Statistical Association* 96, 194-209.

Ghysels, E., R. E. Mcculloch, R. S. Tsay. 1998. Bayesian inference for periodic regime-switching models, *Journal of Applied Econometrics* 13, 129-143.

Goldfeld, S. M. and R. E. Quandt. 1973. A Markov model for switching regressions. *Journal of Econometrics* 1, 3–16.

Guidolin, M. and A. Timmermann. 2005. Economic implications of bull and bear regimes in UK stock and bond returns. *Economic Journal* 115, 111–143.

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

Hamilton, J. D. 2001. A Parametric Approach to Flexible Nonlinear Inference. *Econometrica* 69, 537– 573.

Härdle, W., Tsybakov, A., and Yang, L. 1998. Nonparametric Vector Autoregression. *Journal of Statistical Planning and Inference* 68, 221–245.

Huber F, and L. Rossini. 2021. Inference in Bayesian Additive Vector Autoregressive Tree models. *arXiv* 2006.16333.

Haas, M., S. Mittnik and M. S. Paolella, 2004, A New Approach to Markov-Switching GARCH Models, *Journal of Financial Econometrics* 2, 493–530.

Hwu, S. T. 2018. "Essays on Nonparametric and Dynamic Time-Series Econometrics." PhD dissertation. University of Washington.

Hwu, S. T., C. J. Kim, and J. Piger. 2019. An n-state endogenous Markov-switching model with applications in macroeconomics and finance. *Macroeconomic Dynamics*. Cambridge University Press, 1–29.

Jeliazkov, I. , 2013. Nonparametric Vector Autoregressions: Specification, Estimation, and Inference. In *VAR Models in Macroeconomics–New Developments and Applications: Essays in Honor of Christopher A. Sims*. Emerald Group Publishing Limited.

Jensen, M.J. and J.M. Maheu. 2010. Bayesian semiparametric stochastic volatility modeling. *Journal of Econometrics* 157:2, 306–316.

Kang, K. H. 2014. Estimation of state-space models with endogenous Markov regime-switching parameters. *Econometrics Journal* 17, 56–82.

Kim, M.Y. and K.H. Kang. 2020. Bayesian Inference of Multivariate Regression Models with Endogenous Markov Regime-Switching Parameters. *Journal of Financial Econometrics*, 1-46.

Kass, R. and A. Raftery, 1995. Bayes factors. *Journal of the American Statistical Association* 90, 773-795.

Litterman, R. 1986. Forecasting With Bayesian Vector Autoregressions – Five Years of Experience. *Journal of Business and Economic Statistics*, 4, 25–38.

Kalli, M. and J.E. Griffin. 2018. Bayesian nonparametric vector autoregressive model. *Journal of Econometrics* 203, 267–282.

Kim, C.J., J. Piger and R. Start. 2008. Estimation of Markov regime-switching regressions with endogenous switching. *Journal of Econometrics* 143, 263–273.

Kim, C.J. 1994. Dynamic linear models with Markov switching. *Journal of Econometrics*

60, 1–22.

Kim, C.J. and C. R. Nelson. 1999. *State-Space Models with Regime Switching*. Cambridge, MA: The MIT Press.

Koop, G. and D. Korobilis. 2013. Large time-varying parameter VARs. *Journal of Econometrics*, 177, 185-198.

Koop, G. and S. Potter. 1999. Bayes factors and nonlinearity: Evidence from economic time series. *Journal of Econometrics* 88:2, 251-281.

Primiceri, G. E. 2005. Time Varying Structural Vector Autoregressions and Monetary Policy. *Review of Economic Studies* 72, 821–852.

Shahbaba, B. 2009. Discovering hidden structures using mixture models: application to nonlinear time series processes. *Studies in Nonlinear Dynamics and Econometrics* 13:5.

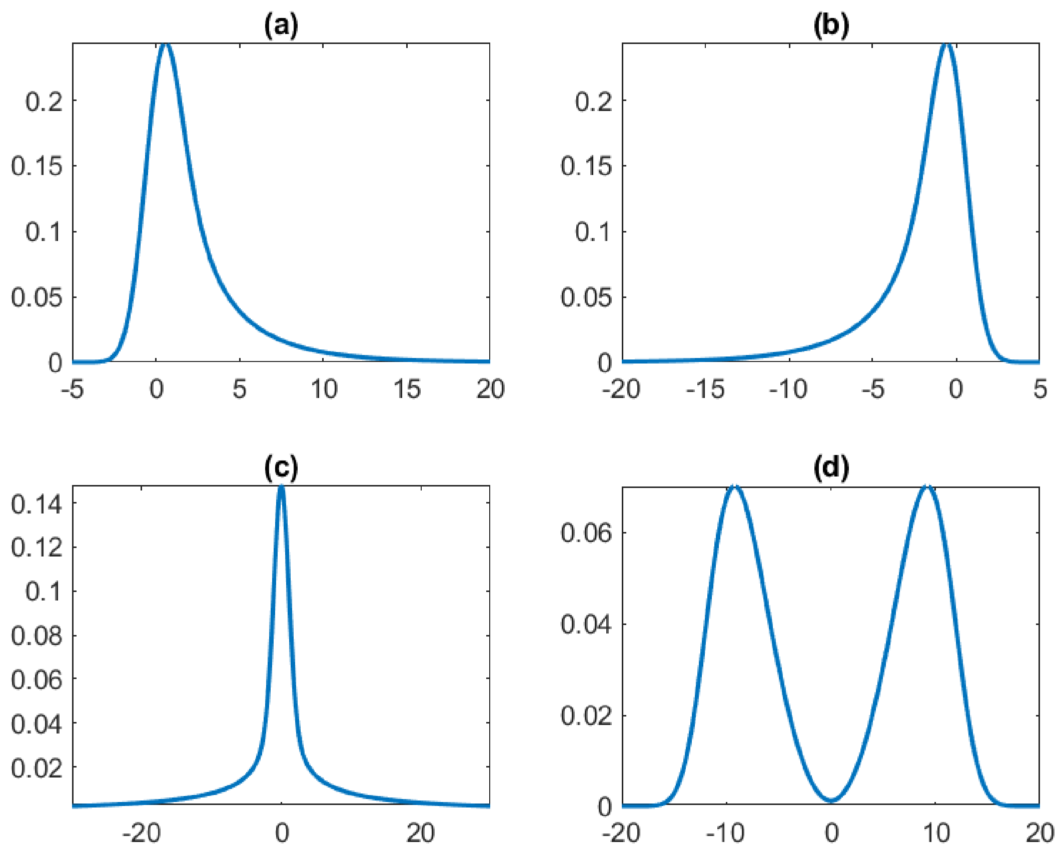
Sims, C. A. 1980. Macroeconomics and Reality. *Econometrica*, 48, 1–48.

Sims, C. A. and T. Zha. 2006. Were There Regime Switches in U.S. Monetary Policy? *American Economic Review*, 96, 54–81.

Uribe, P.W. and H.F. Lopes, 2020. Dynamic sparsity on dynamic regression models. *arXiv* 2009.14131.

Vigfusson, R. 1997. Switching between Chartists and Fundamentalists: A Markov Regime-Switching Approach. *International Journal of Finance and Economics* 2, 291–305.

Figure 1: Marginal Distributions of ε_t



(a) $g(\eta_t) = \eta_t^2$, (b) $g(\eta_t) = -\eta_t^2$, (c) $g(\eta_t) = 10\eta_t^3$, (d) $g(\eta_t) = 10\eta_t^{1/3}$ (the real root).

Table 1: Average Estimation Errors

	Nonparametric Endogenous Model	Parametric Endogenous Model	Exogenous Model
β_{01}	0.0316	-1.3478	0.2404
	(0.0512)	(1.4932)	(0.0267)
β_{11}	0.1071	2.0731	0.5028
	(0.3088)	(1.5123)	(0.0634)
β_{02}	0.0033	1.3325	-0.0035
	(0.0869)	(1.2015)	(0.0281)
β_{12}	-0.0035	-1.3373	-0.0185
	(0.0975)	(1.1851)	(0.0570)
β_{03}	-0.0076	-1.3320	-0.0007
	(0.0867)	(1.1978)	(0.0271)
β_{13}	0.0066	1.3262	0.0225
	(0.0969)	(1.1751)	(0.0552)
σ_0	0.0194	0.7062	0.3222
	(0.0373)	(0.5127)	(0.0673)
σ_1	0.0089	0.9562	0.1888
	(0.0359)	(0.5127)	(0.0397)
δ_{01}	0.0142	0.7404	0.0211
	(0.1075)	(0.6413)	(0.1239)
δ_{11}	-0.0094	-0.6912	-0.0128
	(0.1074)	(0.6643)	(0.1252)

The values presented are the estimation errors (parameter estimate - true value) for all parameters in θ. Standard deviations are listed below estimation errors in parentheses.

Figure 2: The Distribution of $\hat{g}(\eta_t)$, Simulation Study

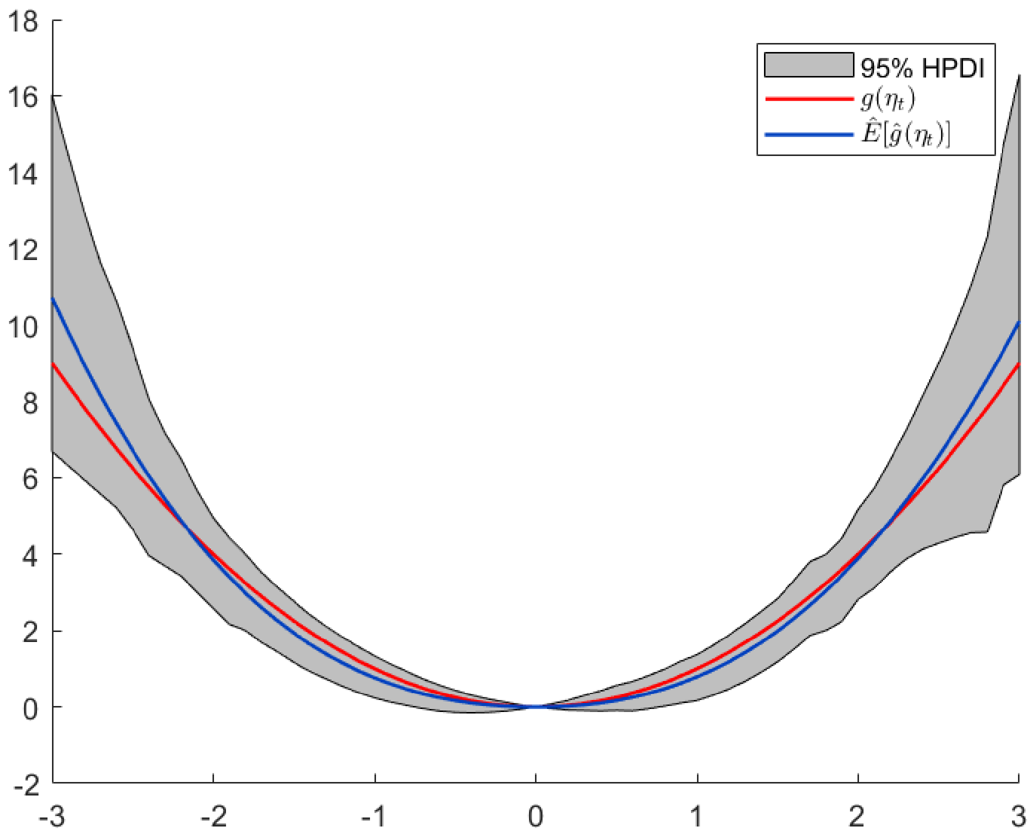


Table 2: Posterior Estimates for Output data

	Nonparametric Endogenous Model	Parametric Endogenous Model	Exogenous Model
β_0	-0.4089	-0.2368	-0.0049
	(0.1298)	(0.2593)	(0.2080)
β_1	0.3702	0.4500	0.5357
	(0.0642)	(0.1317)	(0.1624)
σ	0.2430	0.3475	0.3497
	(0.0362)	(0.0186)	(0.0178)
δ_0	-0.3880	-0.5434	-0.90901
	(0.2429)	(0.5533)	(0.5449)
δ_1	1.3523	1.6845	1.2814
	(0.2625)	(0.4062)	(0.4811)
$ln(f(Y_T\|\Theta^*, \mathcal{M}_i))$	-123.4621	-137.8244	-140.4360
$ln(f(Y_T\|\mathcal{M}_i))$	-139.3520	-143.3834	-144.9914

Standard deviations are listed below parameter estimates in parentheses.

Figure 3: Smoothed Recession Probabilities

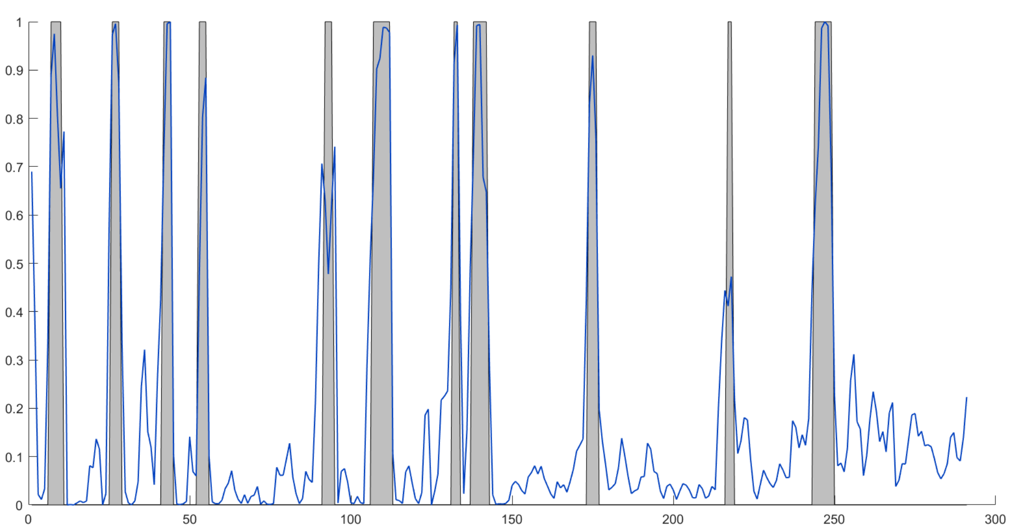


Figure 4: The Distribution of $\hat{g}(\eta_t)$, GDP Model

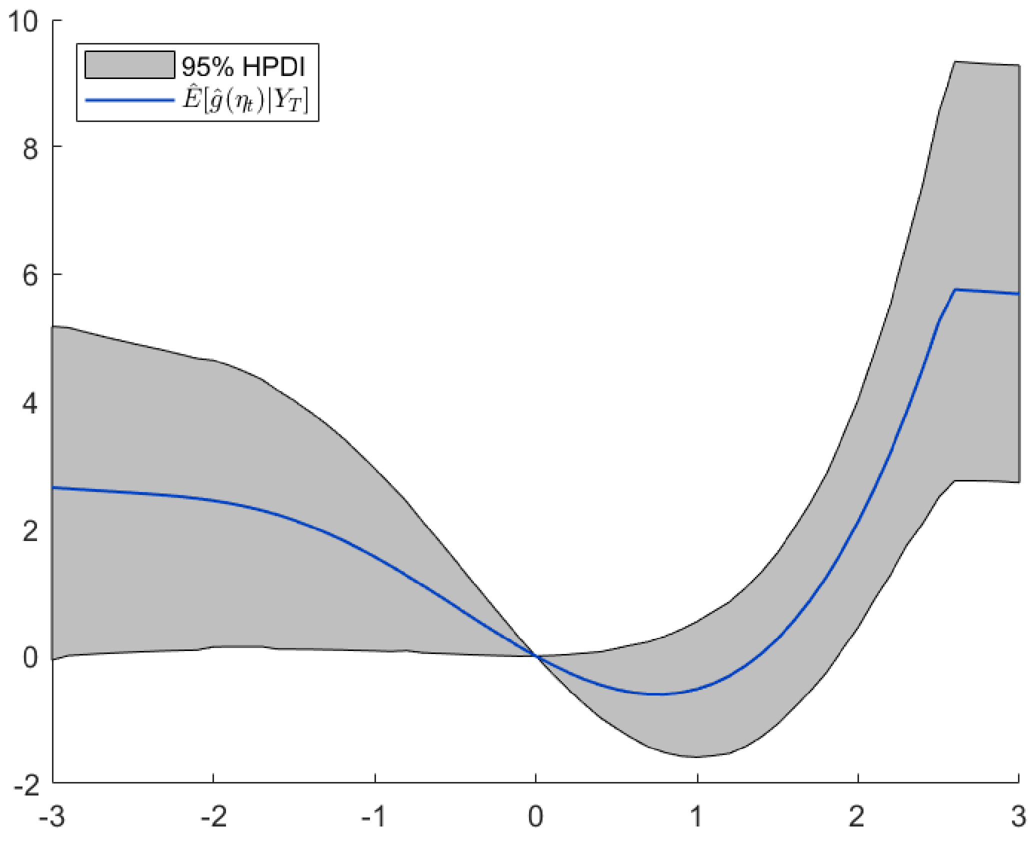


Figure 5: $\hat{f}(\varepsilon_\tau|Y_T)$

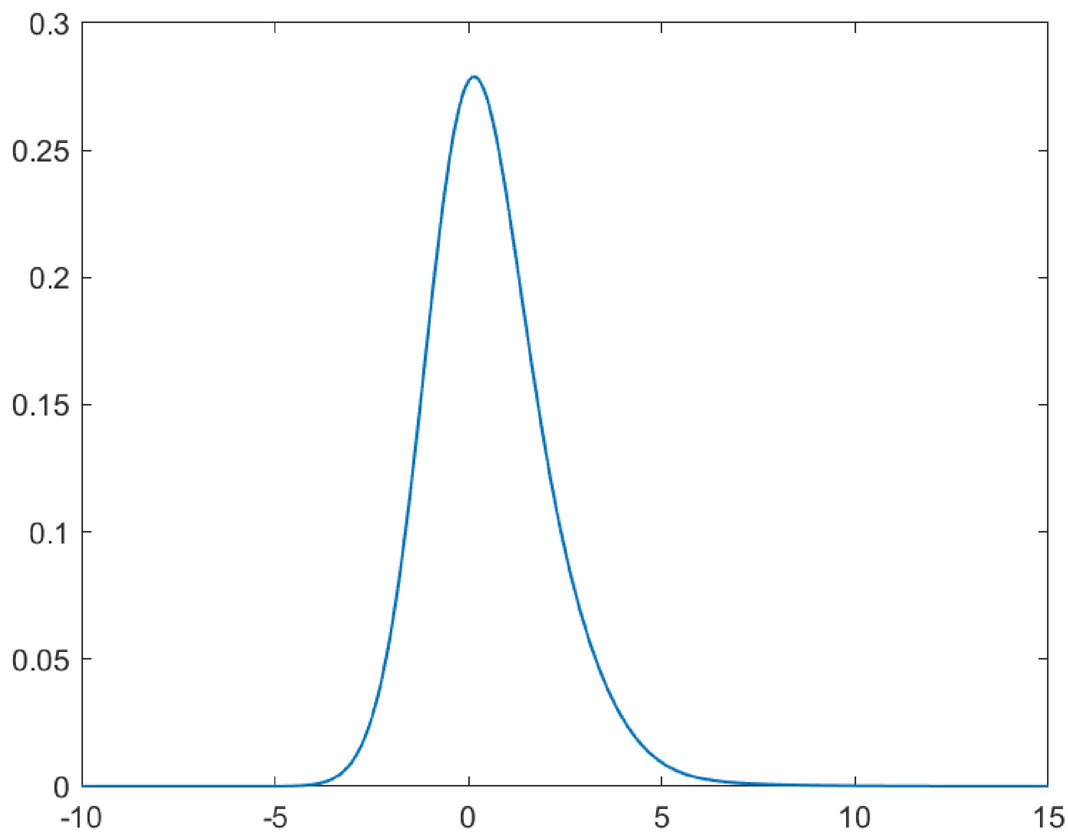


Figure 6: $\hat{f}(\varepsilon_\tau | Y_T, s_\tau, s_{\tau-1})$