Time-Varying Variable Selection for Macroeconomic Forecasting: A Bayesian Regression Tree Approach*

Hyun Jae Stephen Chu†
Jaeho Kim‡
Kyu Ho Kang§

April 2026

Abstract

The growing availability of large datasets in macroeconomic forecasting has renewed interest in variable selection methods. However, most existing approaches assume stable relationships among variables, failing to account for the time-varying relevance of predictors due to nonlinear dynamics and structural changes in the economy. To address this issue, we propose a Bayesian Dirac Classification and Regression Tree (B-DART) model that jointly captures time-varying predictor relevance and performs regime-specific stochastic variable selection within a nonlinear tree-based framework. Empirical results show that the proposed B-DART model delivers superior forecast accuracy relative to Bayesian CART and conventional stochastic variable selection models. These gains are especially pronounced during the post-COVID-19 period, which is characterized by heightened volatility and structural instability.

JEL classification: C11, C22, C52, C53

Keywords: variable selection, Dirac spike-and-slab, Bayesian CART, nonlinear regression, forecasting

*We thank Yunjong Eo and Chirok Han for valuable comments. We also thank Munechika Katayama and participants of the 3rd Tokyo-Taipei-Seoul Macroeconomics Network Workshop for helpful comments. Kyu Ho Kang acknowledges the support from the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2025S1A5A2A01007094).

†(i) Department of Economics, University of Notre Dame, Notre Dame, USA, (ii) Department of Economics, Korea University, Seoul, Korea, Email: hchu@nd.edu.

‡Department of Economics, Sogang University, Seoul, Korea, Email: jaehoecon@sogang.ac.kr.

§*Corresponding author*, Professor, Department of Economics, Korea University, Seoul, Korea, Email: kyuho@korea.ac.kr.

1 Introduction

Recently, variable selection methods have received extensive attention in macroeconomic forecasting, alongside the increasing availability of big datasets.[1] Incorporating a large number of predictor variables into forecasting models can mitigate the loss of forecast accuracy caused by omitting important predictors and also guard against forecast instability.[2] Despite these advantages, expanding the dimensionality of the predictor space can exacerbate parameter uncertainty, potentially leading to inaccurate forecasts. As a result, variable selection methods have attracted considerable attention in the literature (see Giannone, Lenza, and Primiceri (2021)).

However, an often overlooked aspect of variable selection methods in the literature is that the set of important predictors may change over time. This is because relationships among economic variables are unstable and can shift due to external shocks, such as economic policy changes, fluctuating market conditions, or technological advances. Indeed, numerous studies have documented evidence of time-varying relationships across a wide range of macroeconomic variables, including inflation, GDP growth, the equity premium, stock market predictability, exchange rates, and business cycles (Stock and Watson, 1996, 2003; Welch and Goyal, 2008; Pesaran and Timmermann, 1995; Rapach and Wohar, 2006; Rossi, 2013b; Ng and Wright, 2013). In particular, recent studies such as Jordà, Singh, and Taylor (2022) and Bernanke and Blanchard (2023) empirically demonstrate that the COVID-19 pandemic fundamentally altered macroeconomic dynamics.

Beyond accounting for time-varying relations, achieving accurate forecasts requires models that can capture inherent nonlinearities in the economy. Economic activity does not always respond symmetrically to shocks. For instance, the effects of fiscal or monetary policy can differ substantially between recessions and expansions. The seminal work by Hamilton (1989) introduced Markov-switching models to capture such distinct eco-

[1] Varian (2014) outlines various methods using big data in econometric analysis, while Bok, Caratelli, Giannone, Sbordone, and Tambalotti (2018) provides examples of applying big data to track economic conditions and forecast macroeconomic variables.

[2] Rossi (2021) defines forecast instability as instability in the loss function. Under a quadratic loss function, forecast instability corresponds to time-varying forecast errors.

nomic regimes, like shifts between high and low growth periods. Similarly, Smooth Transition Autoregressive (STAR) models, detailed in the work of Granger and Teräsvirta (1993), allow for a more gradual evolution between states. More recently, Auerbach and Gorodnichenko (2012) demonstrated the power of this approach by showing that government spending multipliers are significantly larger during recessions. These findings underscore that linear models may provide a misleading picture of the economy's dynamics. Neglecting nonlinearity can result in substantial forecast errors, particularly around critical turning points in the business cycle.

The purpose of this study is to enhance forecast accuracy by allowing for time-varying variable selection, parameter instability, and nonlinearity in prediction. We achieve this through a novel framework that (1) partitions the data into distinct regimes, each governed by its own set of parameters, and (2) performs variable selection specific to each identified regime. While these modeling tools serve independent purposes, they must be estimated jointly, as regime changes and variable relevance interact dynamically. In a high-dimensional setting, we empirically demonstrate that the proposed integrated model produces superior out-of-sample forecasting performance than models that apply either regime-switching or variable selection in isolation.

To jointly identify latent regimes and select relevant predictors, we propose a Bayesian Dirac Classification and Regression Tree model, which we refer to as B-DART. Our approach integrates the Bayesian Classification and Regression Tree (B-CART) model (Chipman, George, and McCulloch (1998)) with a Dirac-type spike-and-slab stochastic variable selection strategy (Mitchell and Beauchamp (1988)). This novel approach recursively partitions the predictor space, effectively capturing complex nonlinear relationships and regime changes.[3] Notably, the number of regimes is not predetermined in our model but is instead inferred from the data as an unknown parameter.

Estimating the B-DART model is challenging due to the presence of multiple hyperparameters in the spike-and-slab prior. To address this estimation challenge, we employ a spike component with a Dirac mass at zero and adopt Zellner's g-prior for the slab

[3]As highlighted by Medeiros, Vasconcelos, Veiga, and Zilberman (2021) and Goulet Coulombe, Leroux, Stevanovic, and Surprenant (2022), such nonlinear model structures can markedly improve macroeconomic forecast performance.

(Zellner, 1986). The g-prior, governed by a single hyperparameter, also yields an analytically tractable conditional marginal likelihood. This approach helps circumvent the inherent difficulties associated with specifying reasonable priors for Bayesian variable selection.

Our study is closely related to a large body of literature on nonlinearity and instability in forecasting. Structural break tests, such as Bai and Perron (1998), identify abrupt changes in model parameters but are fundamentally in-sample diagnostics that cannot predict future breaks and require the number of breaks to be specified a priori.[4] Rossi and Soupre (2017) address some of these limitations by evaluating time-varying predictive accuracy under model instability in an out-of-sample setting. Alternative approaches, including Markov-switching, change-point, and smooth transition autoregressive (STAR) models, explicitly accommodate parameter instability and nonlinearity.[5] However, in high-dimensional predictor spaces, the accuracy of estimated breakpoints, regime classifications, and associated model parameters deteriorate rapidly due to the curse of dimensionality. The proposed B-DART model effectively addresses this issue.

Complementary to this literature, a growing body of work adopts tree-based methods as a flexible and data-driven approach to modeling nonlinear relationships in macroeconomic data. Bayesian additive regression trees (BART), proposed by Chipman, George, and McCulloch (2012), approximate unknown functions using ensembles of regression trees and have demonstrated strong predictive performance across a wide range of applications. In macroeconomic forecasting, recent contributions extend BART to multivariate time series settings. For example, Huber and Rossini (2022) develop a Bayesian additive vector autoregressive tree model, while Clark, Huber, Koop, Marcellino, and Pfarrhofer (2023) propose multivariate BART-based specifications that improve density and tail forecasts. Yet despite their flexibility in capturing regime-specific dynamics, none of these tree-based approaches incorporate variable selection that evolves with the identified regimes. The predictor set remains fixed across all nodes of the tree, leaving time-varying predictor relevance entirely unaddressed.

[4]For detailed examples, see Section 2.5 in Rossi (2021).

[5]See Kim and Nelson (1999) for Markov-switching and change-point models, and Teräsvirta (2006) for STAR models.

For empirical illustration, we compare the predictive performance of four competing models, including B-DART, using the dataset of McCracken and Ng (2020), which comprises 221 quarterly macro-financial variables. We generate one- to eight-quarter-ahead forecasts for eight macroeconomic variables: the personal consumption expenditures (PCE) price index, consumer price index (CPI), producer price index (PPI), crude oil spot price, federal funds effective rate, 10-year Treasury yield, real GDP, and the unemployment rate.

Our empirical evaluation yields two key findings. First, by jointly estimating time-varying variable inclusion and partition structure, the B-DART model outperforms both standalone Bayesian CART and stochastic variable selection (SVS) models in out-of-sample forecasting, achieving statistically significant gains during the volatile post-COVID-19 period. This result highlights a synergistic interaction between variable selection and tree-structure learning. Second, an analysis of the predictor sets selected over the forecasting periods confirms that the B-DART model dynamically adapts to new information by updating relevant predictors, whereas alternative models exhibit limited flexibility. These findings underscore the prevalence of model uncertainty and demonstrate that neglecting the temporal evolution of variable relevance can materially impair forecast performance.

The remainder of this paper is organized as follows. In Section 2, we present the formal specification of the B-DART model. Section 3 details the Bayesian estimation algorithm. In Section 4, we evaluate the out-of-sample forecasting performance of B-DART against alternative models using a large macro-finance dataset. Finally, Section 5 concludes.

2 Model Specification

This section introduces the proposed B-DART model. We first specify the underlying Bayesian CART structure, which recursively partitions the covariate space using binary decision rules. To enable flexible and localized variable selection, we then impose a hierarchical spike-and-slab prior within each terminal node of the tree.

2.1 Tree Structure and Splitting Rules

The Bayesian CART model is a widely used nonparametric method for approximating an unknown functional relationship between a response variable and a set of covariates. Consider the following data-generating process for the response variable y_t:

$$y_t = \mathcal{L}(\mathbf{x}_t; \theta) + \varepsilon_t, \tag{1}$$

where $\mathbf{x}_t = \{x_{1,t}, x_{2,t}, \ldots, x_{K,t}\}$ denotes the vector of covariates, K is the number of covariates, θ represents model parameters, and ε_t is the error term such that $E[\varepsilon_t|\mathbf{x}_t] = 0$. The CART model estimates the unknown function $\mathcal{L}(\cdot)$ by recursively partitioning the covariate space using a selected variable and its associated threshold at each internal node of a binary decision tree. Each split partitions the data into two child nodes, each representing a subset of observations with more homogeneous response behavior. This scheme enables the model to capture distinct regimes and nonlinear relationships in the outcome variable.

Our study extends the standard Bayesian CART framework by adopting a linear regression model as the basis function within each terminal node, allowing for local linear approximations of the unknown function $\mathcal{L}(\cdot)$. Specifically, the response variable y_t is modeled as

$$y_t = \left[\sum_{r=1}^{R} \mathbb{1}(d_t = r) \mathbf{x}_t' \boldsymbol{\beta}_r \right] + \varepsilon_t, \tag{2}$$

where $\boldsymbol{\beta}_r = [\beta_{r,1}, \beta_{r,2}, \ldots, \beta_{r,K}]'$ and R is the number of terminal nodes (i.e., regimes).[6] The latent indicator $d_t \in \{1,\ 2, \ldots,\ R\}$ denotes the terminal node to which observation t is assigned, and $\mathbb{1}(\cdot)$ is an indicator function that equals one if the given condition is satisfied and zero otherwise. To account for heteroskedasticity, we assume ε_t follows a normal distribution with regime-specific variance, $\varepsilon_t \sim \mathcal{N}(0, \sum_{r=1}^{R} \mathbb{1}(d_t = r)\sigma_r^2)$. The regime configuration is determined by a tree structure, $\mathcal{T}$, and the set of its internal-node decision rules, $\mathcal{I} = \{i^x, i^q\}$. We let $i^x \in \{1, 2, \ldots, K\}$ index the covariate used

[6] All regressions in the terminal node include an intercept term. The equivalence of using original data with an intercept term and the demeaned data without an intercept for Bayesian estimation is shown in *Appendix A*.

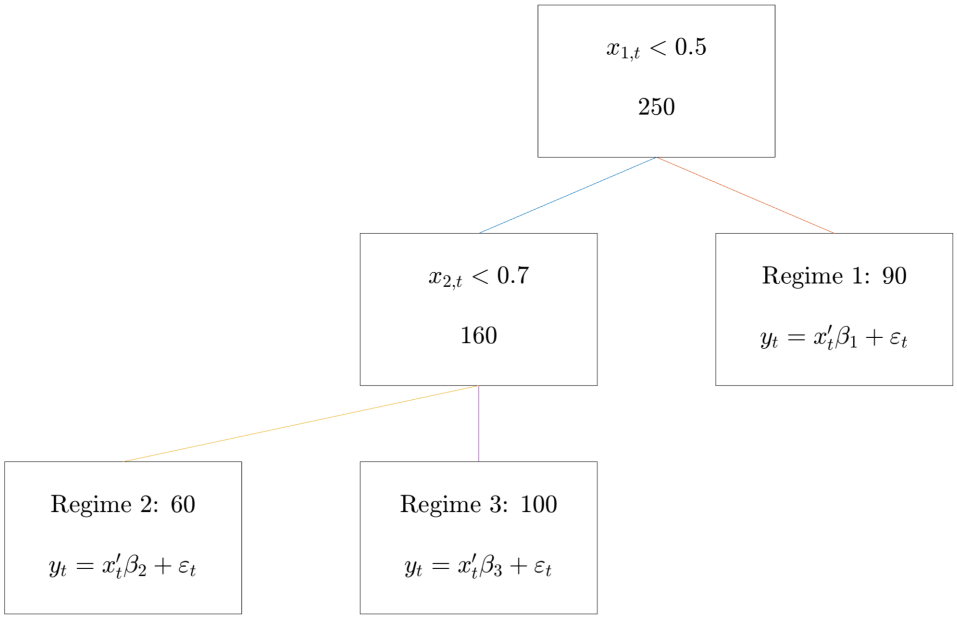


Figure 1: **Example of a Regression Tree**

Notes: This figure illustrates an example of a simple regression tree with two internal nodes and three terminal nodes (regimes), where splits occur at $x_{1,t} < 0.5$ and $x_{2,t} < 0.7$. The error terms of each terminal node have different variances σ_r^2 for $r = 1, 2, 3$.

for splitting, and $i^q \in \mathbb{R}$ specifies the corresponding threshold. Each decision rule at an internal node consists of a pair $\{i^x, i^q\}$. Together, $\mathcal{T}$ and $\mathcal{I}$ recursively partition the full sample into distinct regimes, within which a separate linear model is fit.

Figure 1 illustrates a simple regression tree with two internal nodes and three terminal nodes (regimes). The first split is based on whether $x_{1,t} < 0.5$, dividing the data into two branches. All time periods satisfying this condition are assigned to the left node, and the remaining observations are assigned to the right node. For this first split rule, $i^x = 1$ and $i^q = 0.5$, and the depth of the root node at which this split occurs is 0. The left branch further splits on $x_{2,t} < 0.7$. For this second split rule, $i^x = 2$ and $i^q = 0.7$, and the depth of this internal node is 1. The right branch does not split further and thus becomes a terminal node. This structure results in three distinct regimes, each associated with its own linear regression model characterized by parameters $\{\boldsymbol{\beta}_1, \sigma_1^2\}$, $\{\boldsymbol{\beta}_2, \sigma_2^2\}$, and $\{\boldsymbol{\beta}_3, \sigma_3^2\}$, respectively.

To specify a prior over the tree structure $\mathcal{T}$, which determines the regime configuration, we follow the approach of Chipman et al. (1998). This prior favors simpler trees by making the probability of a node splitting decrease with its depth. The probability that a node η at depth d_η (where the root node has depth $d_\eta = 0$) becomes an internal node (i.e., splits) is defined as

$$p_{\text{split}}(\eta) = \kappa(1 + d_\eta)^{-\rho}, \tag{3}$$

where hyperparameters $\kappa \in (0, 1)$ and $\rho \geq 0$ govern the complexity of the tree. The prior hyperparameter κ controls the baseline probability of splitting. A lower value of κ favors smaller trees overall, while ρ controls the penalty for depth. A higher value of ρ makes the splitting probability decrease more rapidly, thus discouraging complex, deep trees.

The full prior probability for a specific tree structure $\mathcal{T}$ is the product of the probabilities of splitting for all internal nodes and the probabilities of *not* splitting for all terminal (leaf) nodes. Let I be the set of internal nodes and L be the set of terminal (leaf) nodes for tree $\mathcal{T}$. The tree prior[7] is given by:

$$p(\mathcal{T}) = \left[\prod_{\eta \in I} \kappa(1 + d_\eta)^{-\rho}\right] \left[\prod_{\eta \in L} (1 - \kappa(1 + d_\eta)^{-\rho})\right]. \tag{4}$$

A splitting rule at any internal node is defined by selecting both a predictor variable for regime changes and a corresponding threshold value. We assign a uniform prior over the set of possible splitting rules. A predictor variable is first chosen uniformly from the K available predictors. This is equivalent to sampling from a Categorical distribution with equal probabilities:

$$i^x \sim \text{Categorical}\left(\frac{1}{K}, \ldots, \frac{1}{K}\right) \quad \text{for } i^x \in \{1, \ldots, K\}. \tag{5}$$

Next, conditional on the selected predictor $x_{i^x,\cdot}$, a threshold value i^q is drawn uniformly

[7]Refer to Figure 3 in Chipman et al. (1998) for a visualization of how the tree structure evolves under different combinations of κ and ρ.

from the observed range of that variable:

$$i^q | i^x \sim \text{Uniform}(\min(x_{i^x,\cdot}), \max(x_{i^x,\cdot})). \tag{6}$$

This prior specification reflects an initial belief that all predictors and their potential thresholds are equally informative.[8] We denote the joint prior for all splitting rules $\mathcal{I}$, conditional on the tree structure $\mathcal{T}$, as $p(\mathcal{I} \mid \mathcal{T})$, which is given by the product of the prior densities defined in Equations (5) and (6).

2.2 Stochastic Variable Selection

Conditional on the regime configuration determined by $\mathcal{T}$ and $\mathcal{I}$, we perform SVS within each regime. Let $\boldsymbol{\delta}_r = \{\delta_{r,1}, \ldots, \delta_{r,K}\}$ denote the latent inclusion indicators, where $\delta_{r,k} = 1$ (0) indicates that predictor k is active (inactive) in regime r. That is, if $\delta_{r,k} = 0$, the corresponding parameter $\beta_{r,k}$ is set to zero, effectively excluding it from the local linear regression for regime r. Conditional on $\boldsymbol{\delta}_r$, the prior for $\boldsymbol{\beta}_r$ is specified as:

$$p(\boldsymbol{\beta}_r | \sigma_r^2, \boldsymbol{\delta}_r) = \pi_{slab}(\boldsymbol{\beta}_{r,\delta}) \cdot \Big\{ \prod_{k:\delta_{r,k}=0} \pi_{spike}(\beta_{r,k}) \Big\}$$

where $\boldsymbol{\beta}_{r,\delta}$ denotes the collection of coefficients $\beta_{r,k}$ for which $\delta_{r,k} = 1$, and the spike prior for $\beta_{r,k}$ is a Dirac measure that assigns unit mass at 0.

We use the g-prior of Zellner (1986) as the slab distribution for $\boldsymbol{\beta}_{r,\delta}$. As noted by Malsiner-Walli and Wagner (2011), this prior choice is particularly advantageous because it yields a simple expression for the marginal likelihood and substantially reduces the computational burden by avoiding the need to compute matrix determinants. The g-prior is given by

$$\boldsymbol{\beta}_{r,\delta} \sim \mathcal{N}\left(0, \sigma_r^2 \cdot g \cdot (\mathbf{X}_{r,\delta}' \mathbf{X}_{r,\delta})^{-1}\right). \tag{7}$$

Here, $\mathbf{X}_r$ is the $T_r \times K$ matrix containing the time-series observations of all K covariates for regime r, where T_r denotes the number of observations in that regime. The matrix $\mathbf{X}_{r,\delta}$, used in Equation (7), is the $T_r \times K_r$ submatrix of $\mathbf{X}_r$ formed by including only the

[8] Alternatively, one may consider a non-uniform specification that encourages sparsity, such as the one proposed by Linero (2018) and Creal and Kim (2021).

columns for which the selection indicator $\delta_{r,k}$ equals 1, where K_r denotes the number of such columns. The hyperparameter g controls the prior variance, governing the degree of shrinkage toward the prior mean of zero.

We model the latent inclusion indicators $\delta_{r,k}$ using a hierarchical Beta–Bernoulli prior to allow for regime-specific sparsity. Within each regime r, the indicators are treated as independent draws from a Bernoulli distribution governed by a regime-specific inclusion probability p_r:

$$\delta_{r,k} \sim \text{Bernoulli}(p_r).$$

This structure allows the model to learn the propensity for variables to be included in each regime. The regime-specific probability p_r is itself assigned a conjugate Beta hyperprior,

$$p_r \sim \text{Beta}(a_0, c_0).$$

Finally, the error variance for each regime, σ_r^2, is assigned an Inverse-Gamma prior,

$$\sigma_r^2 \sim \text{IG}\left(\frac{\nu}{2}, \frac{\nu\lambda}{2}\right).$$

3 Bayesian Estimation

We employ a Markov Chain Monte Carlo (MCMC) algorithm to jointly estimate all parameters of the proposed B-DART model. The algorithm alternates between updating the model's two core components: the tree structure that partitions the data into regimes and the set of regime-specific parameters under SVS. In this section, we describe the update steps for each component in detail.

3.1 Metropolis-Hastings Algorithm for Tree Structure

To obtain the posterior samples of tree structure $\mathcal{T}$ and split rules $\mathcal{I}$, we derive the Metropolis-Hastings (MH) algorithm whose target posterior density is given by

$$p(\mathcal{T}, \mathcal{I}|\mathbf{X}, \mathbf{Y}) \propto p(\mathbf{Y}|\mathbf{X}, \mathcal{T}, \mathcal{I}) \cdot p(\mathcal{I}|\mathcal{T}) \cdot p(\mathcal{T}). \tag{8}$$

Following Chipman et al. (1998), we sample a candidate tree structure $\mathcal{T}^*$ from one of the following proposal moves, conditional on the tree structure $\mathcal{T}^{(j-1)}$ accepted at MCMC iteration $j-1$:

- **Grow**: Randomly select a terminal node and split it into two child nodes using a new splitting rule drawn from the prior distribution defined in Equations (5) and (6).
- **Prune**: Randomly select a parent node with two terminal children and collapse it into a single terminal node.
- **Change**: Randomly select an internal node and reassign a new splitting rule drawn from the prior distribution defined in Equations (5) and (6).
- **Swap**: Randomly select a parent–child pair that are both internal nodes and swap their splitting rules.

Each candidate move is assigned an equal selection probability of $p_{\text{move}} = 1/4$. Once the type of move is selected, the splitting rules affected by the chosen move are generated according to the prior distributions in Equations (5) and (6), conditional on the candidate tree $\mathcal{T}^*$. The resulting set of splitting rules after applying the candidate move to the previous set of splitting rules $\mathcal{I}^{(j-1)}$ is denoted by $\mathcal{I}^*$.

Once a candidate tree $\mathcal{T}^*$ and the corresponding set of splitting rules $\mathcal{I}^*$ are generated, the MH acceptance probability is computed as follows:

$$
\begin{aligned}
&\alpha(\mathcal{T}^*, \mathcal{I}^* | \mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)}) \qquad (9)\\
&= \min\left(1, \frac{p(\mathcal{T}^*, \mathcal{I}^* | \mathbf{X}, \mathbf{Y}) \cdot q(\mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)} | \mathcal{T}^*, \mathcal{I}^*)}{p(\mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)} | \mathbf{X}, \mathbf{Y}) \cdot q(\mathcal{T}^*, \mathcal{I}^* | \mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)})}\right)\\
&= \min\left(1, \frac{p(\mathbf{Y}|\mathbf{X}, \mathcal{T}^*, \mathcal{I}^*) p(\mathcal{T}^*, \mathcal{I}^*)}{p(\mathbf{Y}|\mathbf{X}, \mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)}) p(\mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)})} \cdot \frac{q(\mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)} | \mathcal{T}^*, \mathcal{I}^*)}{q(\mathcal{T}^*, \mathcal{I}^* | \mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)})}\right)\\
&= \min\left(1, \underbrace{\frac{p(\mathbf{Y}|\mathbf{X}, \mathcal{T}^*, \mathcal{I}^*)}{p(\mathbf{Y}|\mathbf{X}, \mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)})}}_{\text{(a)}} \cdot \underbrace{\frac{p(\mathcal{I}^*|\mathcal{T}^*) p(\mathcal{T}^*)}{p(\mathcal{I}^{(j-1)}|\mathcal{T}^{(j-1)}) p(\mathcal{T}^{(j-1)})}}_{\text{(b)}} \cdot \underbrace{\frac{q(\mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)} | \mathcal{T}^*, \mathcal{I}^*)}{q(\mathcal{T}^*, \mathcal{I}^* | \mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)})}}_{\text{(c)}}\right),
\end{aligned}
$$

where $q(\cdot)$ is the proposal density. Each of the three components in Equation (9) corresponds to the following: (a) the ratio of the likelihood functions of the previous and

candidate trees; (b) the ratio of the prior densities of the previous and candidate trees; and (c) the ratio of the probability of moving to the candidate tree given the previous tree to the probability of moving to the previous tree given the candidate tree. Additional constraints on $(\mathcal{T}, \mathcal{I})$ can be imposed during estimation. For example, to enforce the requirement that each regime contains at least 30 observations, we set the MH acceptance probability to zero for any candidate tree that violates this constraint. Details regarding the derivation of the acceptance probability for each candidate move are provided in *Appendix C*.

3.2 Approximate Conditional Marginal Likelihood

This section derives the marginal likelihood of the baseline linear regression model at each terminal node of the tree and introduces a computationally efficient algorithm for selecting relevant variables within the identified regimes. Let $\mathbf{y}_r$ denote the $T_r \times 1$ data vector of the dependent variable corresponding to regime r.[9] Ideally, for the estimation of the regression tree, we need to integrate out all regime-dependent parameters, such as $\boldsymbol{\delta}_r$ and p_r, from the following target density:

$$p(\mathbf{y}_r|\mathbf{X}_r) = \int\int p(\mathbf{y}_r|\mathbf{X}_r, \boldsymbol{\delta}_r)p(\boldsymbol{\delta}_r|p_r)p(p_r)d\boldsymbol{\delta}_r dp_r \tag{10}$$

where

$$p(\mathbf{y}_r|\mathbf{X}_r, \boldsymbol{\delta}_r) = \int\int p(\mathbf{y}_r|\mathbf{X}_r, \boldsymbol{\beta}_r, \sigma_r^2, \boldsymbol{\delta}_r)p(\boldsymbol{\beta}_r|\sigma_r^2, \boldsymbol{\delta}_r)p(\sigma_r^2)d\boldsymbol{\beta}_r d\sigma_r^2 \tag{11}$$

and use this quantity to draw posterior samples for the tree configuration. The marginal likelihood, conditional on $\boldsymbol{\delta}_r$, is analytically obtained as

$$p(\mathbf{y}_r|\mathbf{X}_r, \boldsymbol{\delta}_r) = (\pi)^{-\frac{(T_r-1)}{2}} \cdot T_r^{-\frac{1}{2}} \cdot (\nu\lambda)^{\frac{\nu}{2}} \cdot (g+1)^{-\frac{K_r}{2}} \cdot \frac{\Gamma(\frac{(T_r-1)+\nu}{2})}{\Gamma(\frac{\nu}{2})} \cdot (A_{r,\delta}+\nu\lambda)^{-\frac{(T_r-1)+\nu}{2}}, \tag{12}$$

where

$$A_{r,\delta} = \mathbf{y}_r'\mathbf{y}_r - \mathbf{y}_r'\mathbf{X}_{r,\delta}B_{r,\delta}\mathbf{X}_{r,\delta}'\mathbf{y}_r$$
$$\text{and } B_{r,\delta} = \frac{g}{g+1}(\mathbf{X}_{r,\delta}'\mathbf{X}_{r,\delta})^{-1}.$$

[9]For analytical convenience, variables here are assumed to be centered (demeaned), although the result is equivalent to the decentered case as shown in *Appendix A*.

Details on the derivation of the above marginal likelihood are provided in *Appendix A*.

Evaluating the marginal likelihood in Equation (10) is computationally infeasible because there are 2^K possible configurations of $\boldsymbol{\delta}_r$. This issue becomes particularly severe in high-dimensional settings. To resolve this computational bottleneck, we adopt the following numerical approximation to the target density:

$$p(\mathbf{y}_r|\mathbf{X}_r) \approx \int\int p(\mathbf{y}_r|\mathbf{X}_r, \boldsymbol{\delta}_r)\mathbb{1}(\boldsymbol{\delta}_r = \hat{\boldsymbol{\delta}}_r|p_r)\mathbb{1}(p_r = \hat{p}_r)d\boldsymbol{\delta}_r dp_r = p(\mathbf{y}_r|\mathbf{X}_r, \hat{\boldsymbol{\delta}}_r) \tag{13}$$

where $\mathbb{1}(\cdot)$ is a point-mass function that equals one when the condition inside the parentheses is satisfied and zero otherwise. The parameters $\hat{\boldsymbol{\delta}}_r$ and $\hat{p}_r$ in Equation (13) denote the estimated inclusion indicator vector and the estimated inclusion probability, respectively. More precisely, we approximate the integral in Equation (12) by evaluating the conditional marginal likelihood at a single point estimate of $\boldsymbol{\delta}_r$ and p_r.

The approximate marginal likelihood in Equation (12) requires the estimates $\hat{\boldsymbol{\delta}}_r = \{\hat{\delta}_{r,1}, \hat{\delta}_{r,2}, \ldots, \hat{\delta}_{r,K}\}$ and $\hat{p}_r$. To obtain the parameter estimates, we use a Gibbs sampling algorithm to draw posterior samples of $\boldsymbol{\Theta}_r = \{\boldsymbol{\beta}_r, \sigma_r^2, p_r, \boldsymbol{\delta}_r\}$ conditional on $\{\mathcal{T}, \mathcal{I}\}$ for $r = 1, 2, \ldots, R$. The target posterior density for this step is given by

$$p(\boldsymbol{\beta}_r, \sigma_r^2, p_r, \boldsymbol{\delta}_r|\mathbf{y}_r, \mathbf{X}_r). \tag{14}$$

In Equation (14) above, $\{\mathcal{T}, \mathcal{I}\}$ is omitted from the conditioning set for notational simplicity.

The Gibbs sampling procedure is outlined in *Algorithm 1* and a detailed derivation of the posterior distributions for each parameter, along with the SVS steps, is provided in *Appendix B*. Note that $p(\boldsymbol{\delta}_r|\mathbf{y}_r, p_r)$ and $p(\sigma_r^2|\mathbf{y}_r, \boldsymbol{\delta}_r)$ are obtained by integrating out nuisance parameters and thus are not the standard full conditional distributions. Because analytical derivations are available, this leads to a more efficient sampling scheme by reducing the dimensionality of the parameter space.

After running *Algorithm 1*, we set the k-th inclusion indicator $\hat{\delta}_{r,k} = 1$ if the posterior mean of $\delta_{r,k}$ exceeds the estimated regime-specific inclusion probability, that is, if $\overline{\delta}_{r,k} = \frac{1}{J}\sum_{j=1}^{J}\mathbb{I}(\delta_{r,k}^{(j)} = 1) \geq \hat{p}_r$, and $\hat{\delta}_{r,k} = 0$ otherwise. Here, $\overline{\delta}_{r,k}$ represents the posterior probability that covariate k is included in the baseline regression for regime r, $\delta_{r,k}^{(j)}$ denotes

Algorithm 1: Gibbs Sampling Algorithm for the SVS Block

Given initial values of $\boldsymbol{\delta}_r, p_r$, repeat the following steps J times, conditional on $\mathcal{T}, \mathcal{I}$, for $r = 1, 2, \ldots, R$:

1: For each $k = 1, 2, \ldots, K$, draw $\delta_{r,k} \in \{0, 1\}$ from its conditional posterior distribution:

$$p(\delta_{r,k}|\boldsymbol{\delta}_{r,-k}, \mathbf{y}_r, \mathbf{X}_r) = \frac{p(\mathbf{y}_r|\mathbf{X}_r, \delta_{r,k}, \boldsymbol{\delta}_{r,-k}) \cdot p_r}{p(\mathbf{y}_r|\mathbf{X}_r, \delta_{r,k} = 1, \boldsymbol{\delta}_{r,-k}) \cdot p_r + p(\mathbf{y}_r|\mathbf{X}_r, \delta_{r,k} = 0, \boldsymbol{\delta}_{r,-k}) \cdot (1 - p_r)}$$

where $\boldsymbol{\delta}_{r,-k}$ denotes $\boldsymbol{\delta}_r$ excluding $\delta_{r,k}$ and $p(\mathbf{y}_r|\mathbf{X}_r, \delta_{r,k}, \boldsymbol{\delta}_{r,-k})$ is the conditional marginal likelihood in Equation (12).

2: Draw σ_r^2 from $\mathcal{IG}(\frac{\overline{\nu}_r}{2}, \frac{\overline{\lambda}_r}{2})$ where

$$\overline{\nu}_r = \nu + T_r \text{ and } \overline{\lambda}_r = \nu \cdot \lambda + A_{r,\delta}.$$

3: Draw $\boldsymbol{\beta}_{r,\delta}$ from $\mathcal{N}(\overline{\mu}_{r,\delta}, \overline{V}_{r,\delta})$ where

$$\overline{V}_{r,\delta} = \sigma_r^2 B_{r,\delta} \text{ and } \overline{\mu}_{r,\delta} = B_{r,\delta} \mathbf{X}'_{r,\delta} \mathbf{y}_r,$$

and set $\boldsymbol{\beta}_{r,-\delta} = \mathbf{0}$ where $\boldsymbol{\beta}_{r,-\delta}$ is the vector of regression coefficients corresponding to the excluded covariates.

4: Draw p_r from

$$p_r|\boldsymbol{\delta}_r \sim Beta(a_0 + K_r, c_0 + (K - K_r))$$

where K_r is the number of included predictors.

the j-th posterior draw, and J is the total number of iterations of the conditional Gibbs sampler.

Note that the hyperparameter p_r represents the prior inclusion probability and determines how many predictors are expected to play an important role among the full set of candidate regressors within regime r on average. Since p_r is estimated from the data, the resulting $\hat{p}_r$ summarizes the overall sparsity level of regime r implied by the model. The criterion $\overline{\delta}_{r,k} \geq \hat{p}_r$ therefore compares the posterior importance of each predictor with this baseline prior sparsity level. If $\overline{\delta}_{r,k}$ exceeds $\hat{p}_r$, this implies that the data provide stronger evidence for including the predictor than implied by the prior benchmark, and the variable is retained when evaluating the conditional marginal likelihood. This rule is useful because it is difficult to determine *a priori* how many predictors are

Algorithm 2: MCMC Algorithm

Given the previously accepted draw $\{\mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)}\}$, perform the following steps:

1: Randomly select a candidate move among *Grow, Prune, Change, Swap* with equal probability $p_{move} = \frac{1}{4}$.
2: Generate a candidate tree $\{\mathcal{T}^*, \mathcal{I}^*\}$ from the proposal distribution $q(\mathcal{T}^*, \mathcal{I}^* | \mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)})$ based on the selected move in Step 1.
3: For each terminal node (regime), run *Algorithm* 1 to obtain $\hat{\boldsymbol{\delta}}_r$ and set $\hat{\delta}_{r,k} = 1$ if $\overline{\delta}_{r,k} \geq \hat{p}_r$, and 0 otherwise.
4: Compute the acceptance probability $\alpha(\mathcal{T}^*, \mathcal{I}^* | \mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)})$ in Equation (9) using the approximate conditional marginal likelihood in Equation (13).
5: Accept or reject $\{\mathcal{T}^*, \mathcal{I}^*\}$ according to the acceptance probability α. If accepted, set $\{\mathcal{T}^{(j)}, \mathcal{I}^{(j)}\} = \{\mathcal{T}^*, \mathcal{I}^*\}$, otherwise set $\{\mathcal{T}^{(j)}, \mathcal{I}^{(j)}\} = \{\mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)}\}$.
6: Conditional on $\{\mathcal{T}^{(j)}, \mathcal{I}^{(j)}\}$, run *Algorithm* 1 with $J = 1$ to obtain posterior draws of $\boldsymbol{\Theta}_r^{(j)} = \{\boldsymbol{\beta}_r^{(j)}, \sigma_r^{2(j)}, p_r^{(j)}, \boldsymbol{\delta}_r^{(j)}\}$ for each terminal node r.

relevant within each regime, and the threshold $\hat{p}_r$ is used only as a data-driven measure of sparsity rather than a restriction on the number of predictors selected. *Algorithm* 2 summarizes the main B-DART estimation procedure.

4 Empirical Application

Appendix F validates the in-sample performance of the B-DART model through a simulation exercise. In this section, we empirically examine whether the proposed B-DART model delivers superior forecast accuracy relative to competing models. For a comprehensive comparison, we consider a static variable selection model employing a Dirac spike-and-slab prior, a Bayesian CART model without variable selection, and a benchmark AR(1) model. We evaluate forecast performance based on out-of-sample predictions for a set of macroeconomic variables.

4.1 Data Description

We utilize the FRED-QD dataset compiled by McCracken and Ng (2020), which comprises 221 quarterly macroeconomic and financial variables spanning the period from

1967Q1 to 2024Q2.[10] Using this dataset, we generate forecasts at horizons ranging from one to eight quarters ahead for eight key macroeconomic variables: Personal Consumption Expenditures (PCE), the Consumer Price Index (CPI), the Producer Price Index (PPI), crude oil prices, the Federal Funds Effective Rate (FFE), the 10-year Treasury yield, real Gross Domestic Product (GDP), and the unemployment rate.

A significant challenge in handling high-dimensional datasets arises from multicollinearity, as macroeconomic indicators often exhibit strong interdependencies. A widely adopted solution to mitigate collinearity is the use of factor-based dimension-reduction techniques, such as those proposed by Stock and Watson (2006), Bai and Ng (2008), and Kim and Swanson (2018). Following the conventional approach, we apply principal component analysis (PCA) separately within each of the 14 variable categories provided by McCracken and Ng (2020) to extract representative principal components.[11] The representativeness of each principal component is evaluated based on its eigenvalue ratio, which measures the proportion of the total variance explained.

Table 1 reports 14 categories of variables, including the number of variables in each category, the number of selected principal components based on cumulative eigenvalue ratios, and the eigenvalue ratios for the top three principal components. The results in the table are obtained using the full in-sample dataset. We consider several thresholds for the cumulative eigenvalue ratios. Specifically, we consider thresholds of 0.4, 0.5, 0.7, and 0.8. Each represents the minimum proportion of variance that must be explained for selecting representative principal components. For instance, a threshold of 0.4 implies that the chosen principal components must collectively explain at least 40% of the total variation within their respective category groups. Applying these thresholds results in the selection of 21, 26, 49, and 68 principal components, respectively.

In the prediction evaluation, we retain only those principal components that cumulatively explain at least 40% of the variance within each category group. This criterion is chosen since stricter thresholds (e.g., 50% or higher) often result in nearly singular pre-

[10] Although the original FRED-QD dataset contains 246 variables and covers the period from 1959Q1 to 2024Q2, we restrict our analysis to a balanced subsample beginning in 1967Q1.

[11] Because PCA is highly sensitive to extreme observations, we remove 84 outlier observations from the original 221 predictor variables before performing PCA. Outliers are identified as observations that deviate from the median by more than ten times the interquartile range (IQR) for each variable.

Table 1: Result of PCA for Each Category Groups

Category	Abbrev.	Variables	0.4	0.5	0.7	0.8	1st Eigenvalue	2nd Eigenvalue	3rd Eigenvalue
NIPA	NIPA	22	2	2	5	7	0.3584	0.1461	0.0923
Industrial Production	IP	16	1	2	3	4	0.4665	0.1482	0.136
Employment	EMP	49	2	2	5	9	0.3334	0.1756	0.0803
Housing	HOUSE	11	1	1	3	4	0.597	0.0939	0.0909
Inventories, Orders, and Sales	INV	6	1	2	2	3	0.4723	0.3102	0.104
Prices	PRICE	46	2	3	8	11	0.3545	0.1038	0.0653
Earnings and Productivity	EAPROD	11	2	2	3	4	0.3283	0.2919	0.1572
Interest Rates	INT	15	1	2	3	4	0.4429	0.2242	0.1203
Money and Credit	M&C	13	3	3	5	7	0.2264	0.1487	0.1299
Household Balance Sheets	HBAL	9	1	2	3	3	0.4572	0.2372	0.1232
Exchange Rates	EXCH	4	1	1	2	3	0.5249	0.2524	0.1403
Stock Market	STOCK	5	1	1	2	3	0.5848	0.2108	0.1096
Non-household Balance Sheets	NHBAL	13	2	2	4	5	0.3279	0.215	0.1064
Others (Sentiment)	SENTI	1	1 (Use original variable without PCA)						
Sum		221	21	26	49	68			

Notes: This table shows the result of the Principal Component Analysis (PCA) for 14 category groups provided by McCracken and Ng (2020). The numbers 0.4, 0.5, 0.7, and 0.8 are the total variations of each category explained by the principal components (PC). Also, the last category group, denoted as Others, consists only of the Consumer Sentiment survey from the University of Michigan. Therefore, we use the original variable itself without applying PCA. Sum indicates one plus the total number of PCs for which the cumulative explained variance exceeds each threshold.

dictor matrices in the subsequent regression analysis. Although principal components are orthogonal within each category by construction, the components extracted from different categories may still be correlated with each other. Including too many components across categories therefore introduces multicollinearity in the predictor variable matrix, causing near singularity and numerical instability. An exception is made for the "Other" category group, which contains only a single variable (the consumer sentiment index). Consequently, our final predictor set comprises 20 principal components alongside this single original variable.

Finally, while Table 1 reports the principal components obtained using the full in-sample data, the components used in the out-of-sample forecasting are recursively re-estimated using the available subsample at each out-of-sample forecast period. As a result, the number of principal components used for forecasting may vary across out-of-sample periods. Additional details on this recursive PCA procedure are provided in *Appendix E.*

4.2 Prediction Models

We consider four distinct models for comparison of prediction performance. The first specification is the proposed B-DART model, which simultaneously identifies distinct regimes and conducts SVS via a spike-and-slab prior. The second model assumes a single regime and implements only SVS. We denote this second model as the SVS model. The third model allows for flexible regime determination but retains all covariates without employing variable selection. This model is referred to as the Bayesian CART (B-CART) model. Finally, as a benchmark, we consider a simple AR(1) model with an intercept term.

The first three specifications adopt an autoregressive distributed lag (ADL) structure, incorporating four lagged values of the dependent variable and one lag of each of the selected principal components.[12] These model specifications share the following unified framework for h-period-ahead predictions,

$$y_{t+h} = \sum_{r=1}^{R} \mathbb{1}(d_t = r)\mathbf{x}_t'\boldsymbol{\beta}_r + \varepsilon_{t+h},$$

where $\mathbf{x}_t = [1, y_{t-1}, \ldots, y_{t-4}, \mathbf{f}_{t-1}']'$, $\mathbf{f}_{t-1}$ is the vector of lagged selected principal components, $\varepsilon_{t+h} \sim \mathcal{N}(0, \sum_{r=1}^{R} \mathbb{1}(d_t = r)\sigma_r^2)$, d_t is the regime indicator variable, and $r = 1, 2, \ldots, R$. We conduct forecasts for horizons $h = 1, 2, \ldots, 8$. Here, R denotes the total number of identified regimes, and the regimes are identified by the regression tree model outlined in detail in Sections 2.1 and 3.1.

In the empirical analysis, we carefully choose prior hyperparameters. First, for the tree prior $p(\mathcal{T})$, we set $\kappa = 0.5$ and $\rho = 0.5$, ensuring that the maximum depth of the tree is approximately 10. The prior hyperparameter g of the g-slab prior is set to $g = T_r^2$, where T_r denotes the sample size of regime r.[13] For the inclusion-probability

[12] While incorporating additional lags could potentially capture more past information, this also increases the risk of near-singular predictor matrices in a high-dimensional setting, rendering linear regression computationally infeasible. Thus, our choice of predictors balances informational richness and numerical stability.

[13] Although Fernandez, Ley, and Steel (2001) suggest that $g = T_r$ is adequate, we instead set $g = T_r^2$. Since the tree structure partitions the sample into regimes, some regimes may contain relatively few observations where the likelihood provides limited information and parameter estimates can become unstable. Thus, choosing $g = T_r^2$ corresponds to assuming a more diffuse prior, which helps stabi-

prior in the Dirac variable selection, the hyperparameters are set to $a_0 = 5$ and $c_0 = 5$. When computing the conditional marginal likelihood, we assign $\hat{\delta}_{r,k} = 0$ if $\overline{\delta}_{r,k} < \hat{p}_r$ and $\hat{\delta}_{r,k} = 1$ otherwise, where $\hat{p}_r$ is the estimated regime-specific inclusion probability and $\overline{\delta}_{r,k}$ is the posterior mean of $\delta_{r,k}$ obtained from *Algorithm* 2 given $\{\mathcal{T}, \mathcal{I}\}$. Additionally, we impose a stopping rule requiring each group to contain at least 30 observations to prevent overfitting in excessively small subsamples.

We also tune the hyperparameters related to σ_r^2, as the structure of the generated trees is highly sensitive to this choice. Following Chipman et al. (2012), we calibrate σ_r^2 using the sample-based estimate $\hat{\sigma}_r^2$, setting $\nu = 5$ and adjusting λ proportionally to the standard deviation of Y. Details of the tuning procedure and the final selected hyperparameters are reported in *Appendix G*.

4.3 Predictive Accuracy Comparison

We now present the out-of-sample forecasting accuracy of the four competing models. The evaluation period for out-of-sample forecasting spans 40 quarters, from Q2:2014 to Q1:2024. We use recursive forecasting to obtain the root-mean-squared error (RMSE), which is the criterion for evaluating the forecasting accuracy of each model:

$$RMSE = \left(\frac{1}{T'} \sum_{t=\tau+1}^{\tau+T'} (y_t - \hat{y}_t)^2 \right)^{\frac{1}{2}}, \tag{15}$$

where τ denotes the terminal time period for the in-sample estimation, T' denotes the size of the out-of-sample period, y_t is the realized value of the response variable at time t, and $\hat{y}_t$ is the predicted value of y_t at time t.

We choose a recursive forecasting approach to ensure sufficient sample sizes for identifying distinctive regimes.[14] While a rolling-window approach is a common alternative, it is not suitable in our framework because it discards earlier observations and relies on

lize estimation in such small subsamples and mitigates spurious regime splits driven by small-sample instability.

[14] All computations were performed on a computer equipped with an Intel i7-13700K (16-core) CPU and 64 GB of DDR5 RAM, using parallel computing that utilized 14 of the 16 cores. The entire process took an average of 45.86 hours per macroeconomic variable across eight forecast horizons, including five rounds of hyperparameter tuning and 40 out-of-sample forecasting periods.

Table 2: Relative RMSEs over the Full Out-of-Sample Period

	PCE			CPI			PPI			Crude Oil		
H	B-DART	SVS	B-CART	B-DART	SVS	B-CART	B-DART	SVS	B-CART	B-DART	SVS	B-CART
1	**0.9157**	0.9689	1.0995	1.0291	1.0474	1.1693	**0.9376**	0.9433	1.0656	1.0256	**0.9802**	1.2600
2	**0.9954**	1.0248	1.0328	**0.9707**	0.9819	1.1203	1.0188	1.0017	1.0358	**0.9264**	0.9758	1.1112
3	1.0551	1.0043	1.2695	1.0381	1.0098	1.3438	1.0702	**0.9977**	1.1281	1.0167	1.0027	1.1570
4	1.0134	1.0042	1.4407	1.0820	1.0095	1.4110	1.0480	**0.9931**	1.1719	0.9922	**0.9876**	1.1384
5	1.0083	**0.9927**	1.2733	**0.9830**	0.9998	1.1127	**0.9744**	0.9918	1.0451	0.9763	**0.9762**	1.0324
6	**0.9535**	0.9617	0.9974	0.9956	**0.9769**	1.0640	**0.9930**	1.0009	1.0517	1.0086	**0.9970**	1.0209
7	1.0219	**0.9903**	0.9908	1.0379	**0.9979**	1.2556	**0.9318**	0.9948	1.0310	1.0038	**0.9905**	1.0432
8	1.1140	1.0214	1.2222	1.0159	1.0264	1.4145	1.0005	**0.9839**	1.3730	1.0008	1.0149	1.1542
	B-DART = AR(1) ≻ SVS ≻ B-CART			AR(1) ≻ B-DART = SVS ≻ B-CART			B-DART ≻ SVS ≻ AR(1) ≻ B-CART			SVS ≻ B-DART = AR(1) ≻ B-CART		
	FFE			Long INT			Real GDP			Unemp		
H	B-DART	SVS	B-CART	B-DART	SVS	B-CART	B-DART	SVS	B-CART	B-DART	SVS	B-CART
1	1.2633	1.2493	4.7426	1.0758	1.0297	1.5714	**0.9795**	1.4545	1.5894	**0.7047**	0.9131	0.7690
2	1.0071	1.0685	1.8026	1.0700	**0.9929**	1.3005	1.0741	1.0047	1.3316	**0.9133**	0.9207	1.0330
3	1.0005	1.1391	1.6068	1.2848	1.0046	1.0004	1.0431	**0.9919**	1.0396	1.0504	1.0560	1.0821
4	1.3479	1.0411	2.3198	1.0991	1.0010	1.6222	**0.8683**	0.9962	1.0927	**0.9451**	0.9874	2.0104
5	1.1737	1.0465	2.0343	1.1815	**0.9614**	1.1618	1.0251	**0.9879**	1.1566	**0.9276**	0.9944	1.0610
6	**0.9588**	1.0620	1.2073	**0.9878**	0.9883	1.4678	**0.9610**	1.0043	1.2753	**0.9787**	0.9968	1.2612
7	**0.9633**	1.0285	1.1908	1.0515	1.0058	1.1631	1.0431	1.0160	1.1384	1.0033	**0.9874**	1.2306
8	**0.8539**	0.9302	1.1468	1.5468	1.0173	1.2954	**0.9778**	0.9838	1.1457	1.0401	**0.9819**	1.0580
	AR(1) ≻ B-DART ≻ SVS = B-CART			AR(1) ≻ SVS ≻ B-DART ≻ B-CART			B-DART ≻ SVS = AR(1) ≻ B-CART			B-DART ≻ SVS ≻ AR(1) ≻ B-CART		
	Total # of B-DART 1st = 23, 35.94%						Total # of B-CART 1st = 0, 0.00%					
	Total # of SVS 1st = 18, 28.13%						Total # of AR(1) 1st = 23, 35.94%					

Notes: This table reports the relative root mean squared errors (RMSEs) for out-of-sample forecasts across competing models at horizons $h = 1, \ldots, 8$ for eight macroeconomic variables. All RMSEs are normalized by the AR(1) benchmark, whose value is set to 1. Values below 1 indicate an improvement over the benchmark. Boldface entries denote the best-performing model for each variable and forecast horizon, while rows with no bold values indicate that the AR(1) performs the best. The final row in each panel summarizes the overall ranking based on the number of times each model achieves the lowest RMSE. The evaluation is based on 40 recursive out-of-sample forecasts from 2014Q2 to 2024Q1.

a smaller sample when constructing regression trees, potentially leading to less reliable regime estimates.

Table 2 compares the out-of-sample forecasting performance of the competing models based on relative RMSEs, benchmarked against the AR(1) model.[15] Boldface values in the table indicate that a corresponding model outperforms the other competing models for a given variable and forecast horizon. For each variable, the model ordering shown in the last row of each panel is determined by counting the number of bold values across

[15]The RMSE of the AR(1) benchmark model is normalized to 1.

all forecast horizons. This ordering provides a clear summary of each model's overall performance. The final row of the table reports, for each model, the total number of cases in which it achieves the lowest RMSE, along with the corresponding proportion out of the 64 total cases (8 variables $\times$ 8 forecast horizons). All results are obtained using the optimal hyperparameters for σ_r^2, with details on the selected values provided in *Appendix G*.

Overall, the B-DART model outperforms its counterparts, highlighting the importance of accounting for time-varying relationships in variable selection to enhance forecasting accuracy. Specifically, out of 64 cases (8 variables $\times$ 8 horizons) in total, the B-DART and AR(1) models each achieve the best forecast accuracy in 23 cases (35.94%), followed by the SVS model in 19 cases (28.13%). The B-CART model performs worse than the other models in all cases. Although the B-DART model demonstrates overall superior performance, the best-performing model differs depending on the macroeconomic variable being forecast. This finding reinforces the need to assess forecast accuracy at the individual variable level.

Even at the individual variable level, the B-DART model demonstrates superior performance for indicators of economic activity (real GDP and the unemployment rate) and for certain price-related variables (the PCE and PPI price indices). For economic activity indicators, the B-DART model shows superior performance in forecasting real GDP at the one-, four-, six-, and eight-quarter horizons, and consistently outperforms other models in forecasting the unemployment rate over short- and medium-term horizons (up to six quarters ahead). For price indicators, the B-DART model delivers better performance at longer horizons (five or more quarters ahead) for the PPI price index, and at shorter horizons (up to four quarters ahead) for the PCE price index.

To better understand how different forecasting models respond to structural shifts of this scale, we examine the pre- and post-COVID-19 periods separately. To do this, we divide the full out-of-sample period (Q2 2014 to Q1 2024) into pre- and post-COVID-19 subperiods. This allows us to assess how forecast accuracy changes across more stable and more volatile economic regimes. We choose Q2 2020 as the dividing point, since it marks the onset of the COVID-19 pandemic. Among the 40 out-of-sample periods, this

Table 3: Relative RMSEs over the Pre-COVID-19 Period

	PCE			CPI			PPI			CrudeOil		
H	B-DART	SVS	B-CART	B-DART	SVS	B-CART	B-DART	SVS	B-CART	B-DART	SVS	B-CART
1	1.0460	**0.9322**	0.9617	1.0493	**0.9244**	0.9727	0.8987	**0.8618**	0.8702	1.0682	**0.9868**	1.1066
2	1.0365	1.0186	1.2483	**0.9348**	1.0024	1.2090	1.1407	1.0007	1.1857	**0.9039**	0.9712	1.0687
3	1.0662	**0.9932**	1.3471	1.0489	**0.9705**	1.3126	**0.9651**	0.9975	1.2206	**0.9957**	1.0006	1.1265
4	1.0162	**0.9774**	1.2409	**0.9281**	0.9739	1.1913	1.0785	1.0234	1.1321	**0.9397**	0.9781	1.1153
5	1.0134	1.0148	1.1889	**0.9217**	1.0047	1.0286	0.9848	0.9914	**0.9666**	1.0028	**0.9842**	1.0377
6	1.0070	**0.9963**	1.0121	1.0257	**0.9917**	1.0648	**0.9824**	0.9878	1.0598	**0.9857**	0.9935	1.0219
7	**0.7976**	0.9858	0.9619	**0.9085**	0.9912	1.0070	1.0163	0.9925	**0.9662**	**0.9594**	0.9935	1.0096
8	1.1157	1.0338	1.1712	1.0483	1.0469	1.1916	1.0238	**0.9955**	1.0289	**0.9307**	1.0030	0.9907
	SVS ≻ AR(1) ≻ B-DART ≻ B-CART			B-DART ≻ SVS ≻ AR(1) ≻ B-CART			B-DART = SVS = B-CART = AR(1)			B-DART ≻ SVS ≻ AR(1) = B-CART		
	FFE			Long INT			Real GDP			Unemp		
H	B-DART	SVS	B-CART	B-DART	SVS	B-CART	B-DART	SVS	B-CART	B-DART	SVS	B-CART
1	1.1900	1.2533	1.7867	1.0478	1.0237	1.5542	1.2508	1.3814	1.3883	1.1454	**0.9565**	1.2815
2	1.0362	1.0104	2.6390	1.1267	**0.9907**	1.5984	1.1756	1.1082	1.2298	**0.9449**	1.1861	1.6044
3	1.5347	1.3735	2.5768	1.3480	**0.9753**	1.1095	1.3050	1.0448	1.2706	1.3693	1.2009	1.2766
4	1.5614	**0.9956**	1.4392	1.1802	**0.9943**	1.2451	1.0605	1.0035	1.3244	1.1180	1.1484	1.3354
5	1.9528	1.2224	2.0536	1.4297	**0.9903**	1.3265	**0.9698**	0.9844	1.3148	**0.9310**	0.9551	1.0361
6	1.1247	1.1417	1.5853	1.0160	1.0008	1.5515	1.0751	**0.9995**	1.1743	1.1641	0.9597	**0.9485**
7	1.1524	**0.8643**	1.0709	1.0407	**0.9897**	1.1842	1.3917	1.0004	1.4428	1.0325	**0.9035**	1.2786
8	**0.9836**	0.9872	1.1699	1.4702	1.0063	1.4008	**0.8338**	1.0223	1.4885	1.2889	**0.8637**	1.4168
	AR(1) ≻ SVS ≻ B-DART = B-CART			SVS ≻ AR(1) ≻ B-DART = B-CART			AR(1) ≻ B-DART ≻ SVS ≻ B-CART			SVS ≻ B-DART = AR(1) ≻ B-CART		
	Total # of B-DART 1st = 18, 28.13%						Total # of B-CART 1st = 3, 4.69%					
	Total # of SVS 1st = 22, 34.38%						Total # of AR(1) 1st = 21, 32.81%					

Notes: See Table 2 for details on the construction of the statistics. The results are computed over the pre-COVID-19 period from 2014Q2 to 2020Q1.

quarter represents the most recent major event likely to have altered the underlying data-generating process. Shocks from the pandemic, along with later geopolitical events such as the Russia–Ukraine war, triggered a period of high inflation. These events significantly altered the relationships among macroeconomic variables. As a result, aggregating the full out-of-sample results could obscure regime-specific dynamics.

The forecast accuracy of the candidate models during the pre-COVID-19 period is reported in Table 3. In terms of overall performance, the SVS model achieves the best forecast accuracy in 22 out of 64 cases (34.38%), followed by the AR(1) model in 21 cases (32.81%), the B-DART model in 18 cases (28.13%), and the B-CART model in only three cases (4.69%). These results suggest that models that do not account for structural changes, such as the SVS and AR(1), tend to perform better during the pre-COVID-19 period, when the relationships among variables are relatively stable.

Table 4: Relative RMSEs over the Post-COVID-19 Period

	PCE			CPI			PPI			CrudeOil		
H	B-DART	SVS	B-CART	B-DART	SVS	B-CART	B-DART	SVS	B-CART	B-DART	SVS	B-CART
1	**0.8183**	0.9924	1.1787	**0.9245**	1.1092	1.2641	**0.8997**	0.9793	1.1506	**0.8973**	0.9711	1.4001
2	**0.8574**	1.0317	0.9068	0.9880	**0.9716**	1.0719	**0.9171**	1.0008	0.9743	**0.9516**	0.9815	1.1537
3	1.0385	1.0068	1.2140	**0.9378**	1.0365	1.3665	1.0181	**0.9964**	1.0810	**0.9057**	1.0022	1.1842
4	1.0092	1.0199	1.5621	1.0074	1.0322	1.5270	**0.9741**	0.9829	1.1902	1.0308	**0.9960**	1.1653
5	**0.9467**	0.9817	1.3150	1.0171	**0.9971**	1.1577	**0.9722**	0.9937	1.0833	**0.9495**	0.9653	1.0223
6	**0.9018**	0.9452	0.9894	**0.9092**	0.9710	1.0620	1.0001	1.0047	1.0498	1.0171	1.0008	1.0198
7	1.0018	**0.9947**	1.0073	**0.9910**	1.0014	1.3741	**0.8545**	0.9956	1.0583	1.0090	**0.9894**	1.0730
8	**0.9793**	1.0158	1.2492	**0.9339**	1.0174	1.5190	0.9902	**0.9791**	1.5043	1.0375	1.0280	1.3097
	B-DART $\succ$ AR(1) $\succ$ SVS $\succ$ B-CART			B-DART $\succ$ SVS $\succ$ AR(1) $\succ$ B-CART			B-DART $\succ$ SVS $\succ$ AR(1) $\succ$ B-CART			B-DART $\succ$ SVS = AR(1) $\succ$ B-CART		
	FFE			Long INT			Real GDP			Unemp		
H	B-DART	SVS	B-CART	B-DART	SVS	B-CART	B-DART	SVS	B-CART	B-DART	SVS	B-CART
1	1.1376	1.2488	5.0033	1.0408	1.0352	1.5831	**0.9169**	1.4626	1.6108	**0.6926**	0.9123	0.7593
2	**0.9815**	1.0805	1.6246	1.0432	**0.9980**	1.1172	1.0038	**0.9921**	1.3418	**0.9061**	0.9135	1.0167
3	**0.8846**	1.0938	1.3804	1.1635	1.0207	**0.9321**	**0.9446**	0.9855	1.0107	1.0389	1.0518	1.0765
4	1.1172	1.0467	2.4313	**0.9988**	1.0053	1.7876	**0.8445**	0.9955	1.0643	**0.9105**	0.9812	2.0310
5	**0.9809**	1.0160	2.0322	1.0080	**0.9523**	1.0791	0.9954	**0.9880**	1.1377	**0.9276**	0.9960	1.0619
6	**0.8694**	1.0498	1.1344	**0.9401**	0.9811	1.4221	**0.9474**	1.0047	1.2861	**0.9588**	0.9986	1.2729
7	**0.9252**	1.0546	1.2070	**0.9861**	1.0177	1.1527	**0.9957**	1.0164	1.0971	1.0002	**0.9903**	1.2284
8	**0.8364**	0.9274	1.1507	**0.9070**	1.0227	1.2282	**0.9757**	0.9797	1.1025	0.9903	**0.9885**	1.0361
	B-DART $\succ$ AR(1) $\succ$ SVS = B-CART			B-DART $\succ$ SVS $\succ$ AR(1) = B-CART			B-DART $\succ$ SVS $\succ$ AR(1) = B-CART			B-DART $\succ$ SVS = AR(1) $\succ$ B-CART		
	Total # of B-DART 1st = 41, 64.06%						Total # of B-CART 1st = 1, 1.56%					
	Total # of SVS 1st = 12, 18.75%						Total # of AR(1) 1st = 10, 15.63%					

Notes: See Table 2 for details on the construction of the statistics. The results are computed over the post-COVID-19 period from 2020Q2 to 2024Q1.

An exception arises for price-related indicators such as the CPI and crude oil prices, for which the B-DART model outperforms the other models. In particular, it achieves significantly better accuracy across most forecast horizons for crude oil prices. This implies that commodity prices and price indices may exhibit nonlinear or regime-switching behavior, such as sudden jumps or collapses in oil prices. By accounting for such dynamics, the B-DART model is better equipped to capture these nonlinear relationships and mitigate forecast instability.

The forecast performance of the models for the post-COVID-19 period is reported in Table 4. During this volatile period, the B-DART model significantly outperforms the other models, delivering the best results in 41 out of 64 cases (64.06%). In comparison, the SVS, AR(1), and B-CART models achieve the best performance in only 12 (18.75%), 10 (15.63%), and 1 (1.56%) cases, respectively. These results suggest that the

proposed B-DART model provides more accurate forecasts in a post-pandemic environment marked by abrupt changes and nonlinear dynamics.

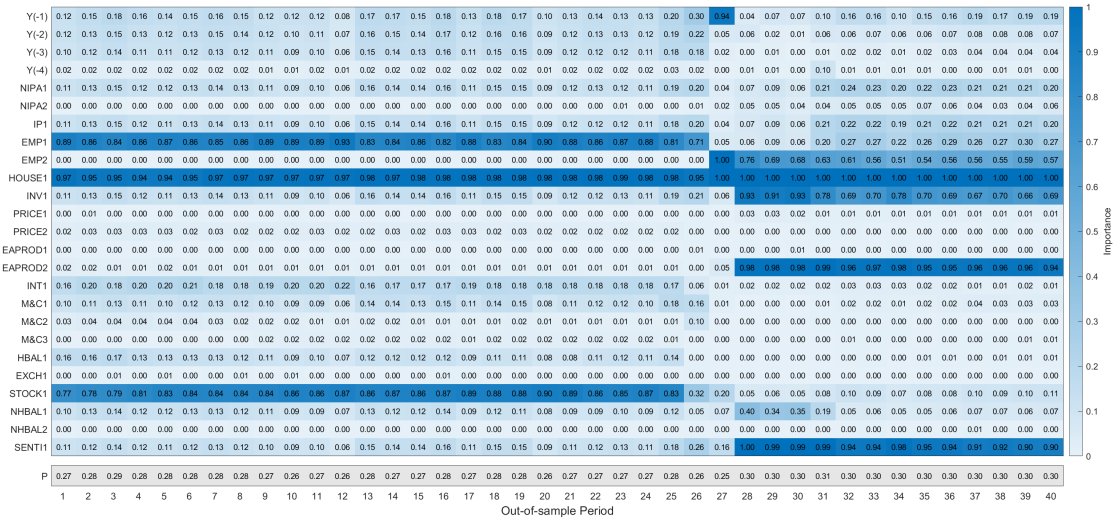
For monetary policy indicators, the B-DART model performs especially well at longer forecast horizons. Similarly, for economic activity indicators, it offers higher accuracy for short- and medium-term unemployment rate forecasts (up to six quarters ahead) and at selected horizons for real GDP. Finally, the B-DART model consistently outperforms all other models for price indicators, effectively capturing shifts in the relationships among macroeconomic variables during the recent high-inflation, post-pandemic period.

4.4 Selected Predictors and Their Importance

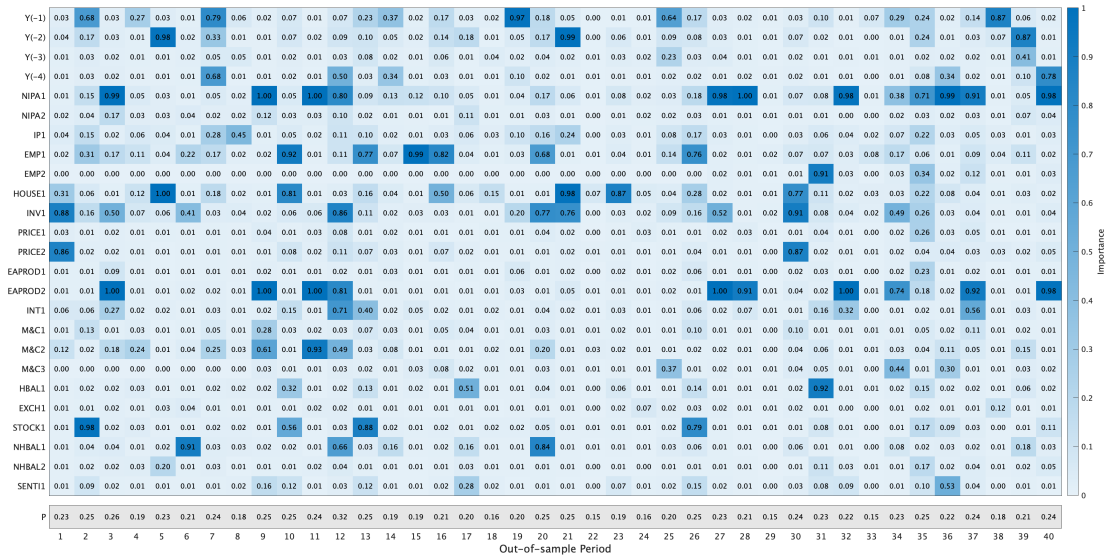
To investigate the underlying reasons for the observed differences in performance, we analyze the predictors selected by the SVS and B-DART models over the out-of-sample periods. Due to space constraints, we limit the discussion to the variable selection results for one- and eight-quarter-ahead forecasts of the unemployment rate, rather than presenting all 64 outcomes. These cases represent situations in which the B-DART model either clearly dominates or is dominated by the SVS model depending on the forecast horizon. Additional results for other macroeconomic variables and other forecast horizons are provided in *Appendix H*.

Figures 2 and 3 present heatmaps showing which predictors are selected for forecasting the unemployment rate. The horizontal axis represents the out-of-sample periods, while the vertical axis lists the predictor variables. The colors in the heatmaps reflect the predictive importance estimated by the posterior mean of $\hat{\delta}_{r,k}$ (i.e., $\bar{\delta}_{r,k}$), with values ranging from 0 to 1, where darker colors indicate greater importance.[16] The final row of the heatmaps, shaded in gray, displays the estimated inclusion probability $\hat{p}_r$ for each out-of-sample period, which serves as the threshold for selecting meaningful predictors for forecasting. However, in the forecasting stage we retain all predictors, even when their estimated importance falls below $\hat{p}_r$. This follows Giannone et al. (2021), who show

[16] As shown in Figure D.1 and Table D.1 in the Appendix, the number of principal components representing each category varies across out-of-sample periods. Specifically, the 9th period (Q2:2016) includes an additional principal component for the Money and Credit category, while the 27th period (Q4:2020) includes an additional principal component for the Employment and Unemployment category. In periods where these principal components are not included, their importance values are set to zero.



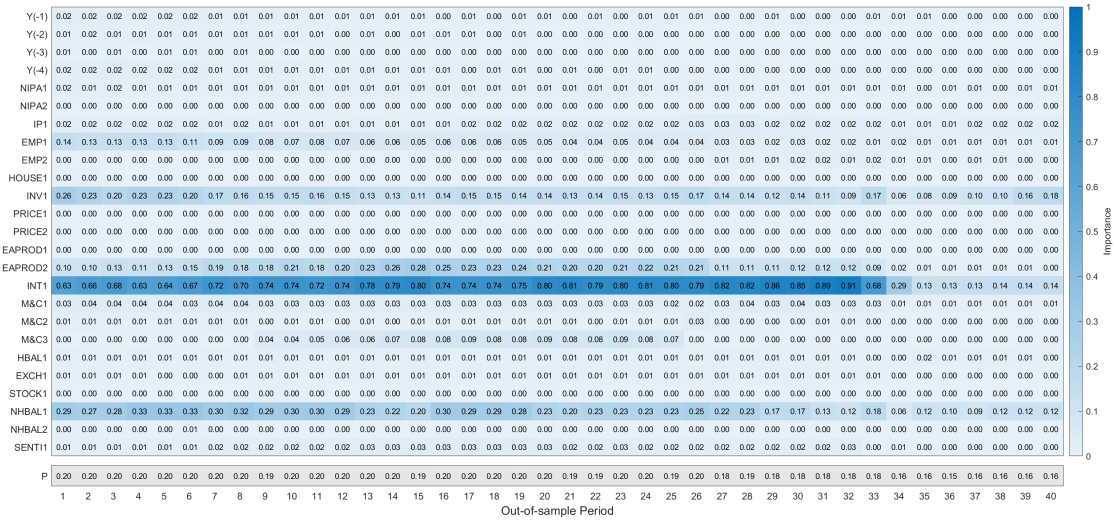
(a) SVS Model



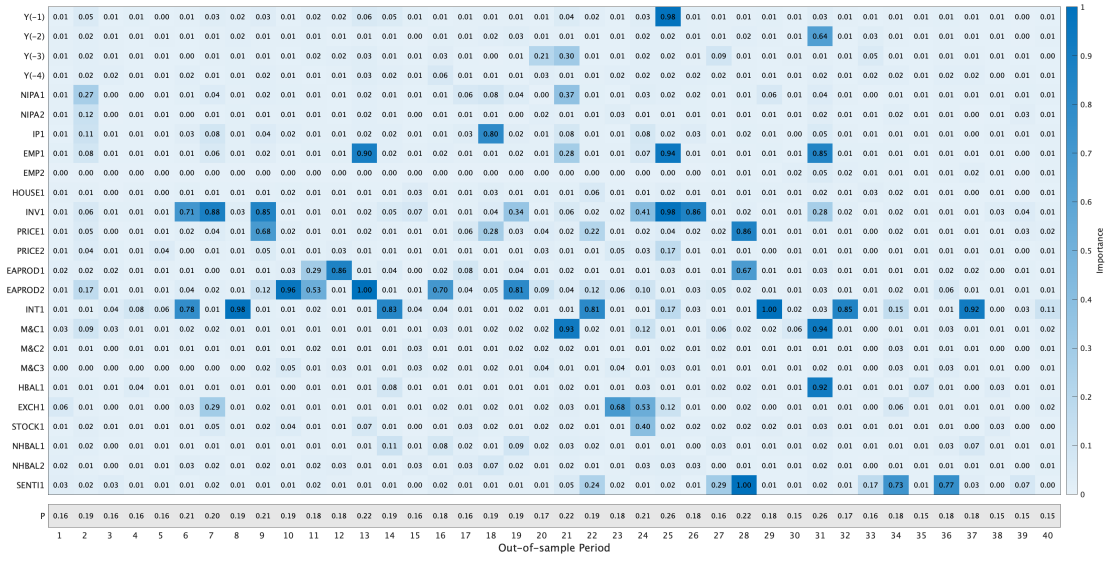
(b) B-DART Model

Figure 2: One-Quarter-Ahead Predictor Importance

Note: This figure reports the importance of each predictor variable for 40 out-of-sample periods when forecasting the one-quarter-ahead unemployment rate. The heatmap becomes darker as the estimated importance of the predictor variable approaches 1. The last row is the estimated inclusion probability $\hat{p}_r$ of the selected terminal node r used in forecasting.



(a) SVS Model



(b) B-DART Model

Figure 3: Eight-Quarter-Ahead Predictor Importance

Note: This figure reports the importance of each predictor variable for 40 out-of-sample periods when forecasting the eight-quarter-ahead unemployment rate. The heatmap becomes darker as the estimated importance of the predictor variable approaches 1. The last row is the estimated inclusion probability $\hat{p}_r$ of the selected terminal node r used in forecasting.

that the posterior inclusion probabilities do not guarantee the systematic importance of individual predictors. Instead, considerable uncertainty may remain regarding whether certain predictors should be included in the model, i.e., model uncertainty is pervasive, and excluding such predictors may therefore lead to an "illusion of sparsity".

Overall, both models reflect changes in variable importance as new information is incorporated and the sample size used for forecasting expands. However, the nature of these information updates differs substantially across the two models. In the SVS model, the importance of predictors rarely changes qualitatively before or after the COVID-19 pandemic, but it undergoes a structural shift during the pandemic. For example, in Figure 2, which reports the one-quarter-ahead unemployment rate forecasts, we observe that the SVS model exhibits a pronounced shift in variable importance around the 27th out-of-sample period (Q4:2020), coinciding with the onset of the COVID-19 pandemic. Even in the case of eight-quarter-ahead unemployment rate forecasts, reported in Figure 3, the importance of predictors remains relatively stable over time with no pronounced shifts comparable to those observed at shorter horizons. In particular, only the contribution of INT1 gradually diminishes, while the overall pattern of variable importance remains largely unchanged. This rigidity in updating new information in the SVS model may impair its forecast performance relative to the B-DART model.

In contrast, the B-DART model produces more abrupt and frequent adjustments in predictor importance. The implications of this flexibility depend on the underlying data environment. On the one hand, as shown in Table 4, the ability to rapidly update variable importance allows the B-DART model to adapt effectively in relatively volatile environments characterized by structural change. On the other hand, this same flexibility may lead to unnecessary adjustments when the underlying relationships are relatively stable, contributing to weaker performance as observed in Table 3. Overall, as shown in Table 2, the B-DART model achieves one of the best forecasting performances among the models in the full out-of-sample period. These results suggest that, while the B-DART model is not uniformly superior, its ability to flexibly select important predictors becomes particularly valuable when the data-generating process is subject to instability.

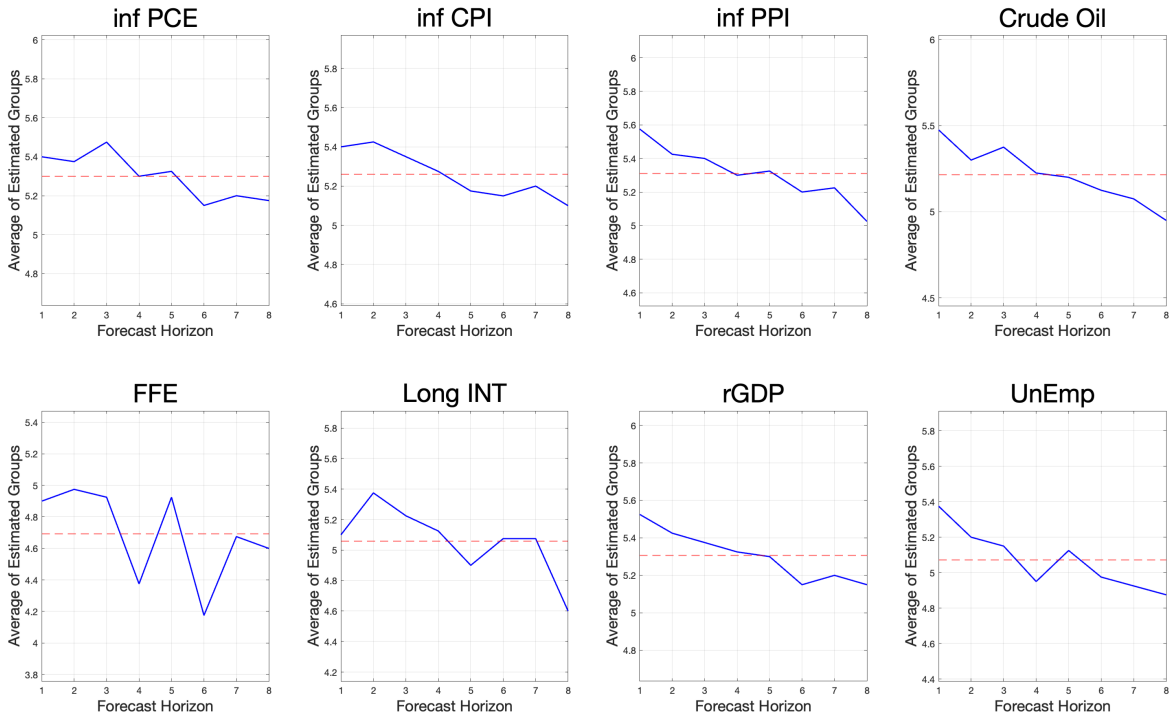


Figure 4: Posterior Mean of the Regime Count

Note: This figure depicts the estimated average number of groups for macroeconomic variables in the B-DART model. The blue line shows the average number of groups across each forecast horizon, while the red dotted line indicates the overall mean across all eight forecast horizons.

4.5 Regression Tree Structure

This section presents additional findings from applying the B-DART model to forecast eight macroeconomic variables. Detailed results on the number of regimes estimated by the B-DART model are shown in Figure 4. In the figure, the blue line represents the posterior mean of the regime count across forecast horizons for each target variable. The red dotted line indicates the average regime count across all eight horizons. On average, the B-DART model identifies five regimes when forecasting macroeconomic variables. An exception is the federal funds effective rate (FFE), where the number of regimes fluctuates between four and five. In general, the number of regimes tends to decrease as the forecast horizon increases.

In contrast, the B-CART model captures only a single group on average for most

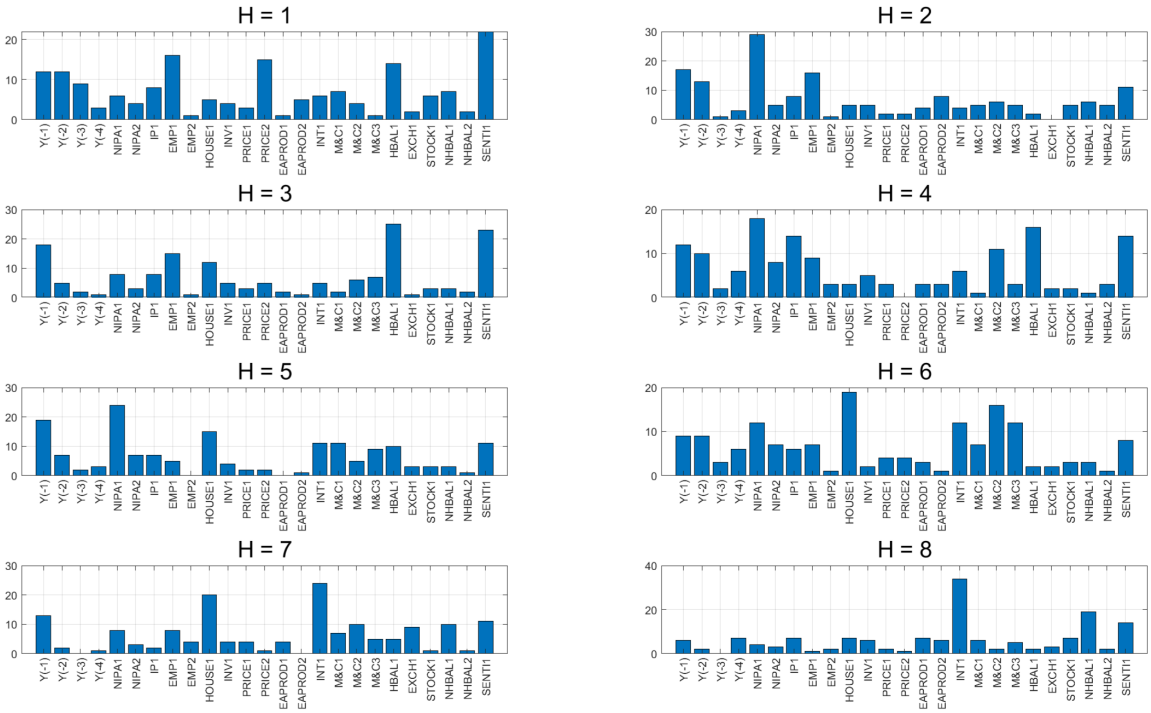


Figure 5: Predictor Usage Frequency in Splitting Rules for Unemployment Rate

Note: This figure depicts the frequency of predictor variables used as the splitting criteria of the B-DART model for forecasting the unemployment rate.

macroeconomic variables.[17] This result suggests that the B-CART model struggles to detect meaningful regime shifts in macroeconomic relationships. Its limited effectiveness is likely due to the combination of small within-regime subsamples and the absence of informative priors. When too many irrelevant parameters are introduced, the statistical significance of the tree structure deteriorates. These findings, along with the consistent outperformance of the B-DART model over the B-CART model in Tables 2, 3, and 4, indicate that integrating variable selection with regime configuration within a regression tree framework creates a synergistic effect.

The B-DART model enables the identification of predictors that are important for regime classification. Due to space limitations, we again focus on the case of the unemployment rate.[18] Figure 5 displays the frequency with which each predictor is used in the

[17] See Figure D.2 in *Appendix D* for detailed average group counts across forecast horizons for each macroeconomic variable.

[18] Additional results for other variables are available in *Appendix H*.

splitting rules across different forecast horizons. Interestingly, some predictors emerge more frequently as splitting variables than others. For example, autoregressive lags of the unemployment rate are consistently selected in the splitting rules across almost all forecast horizons, except for the eight-quarter-ahead case. Economic Sentiment is another predictor that is frequently selected, although its selection rate tends to decline at longer horizons.

For short-term forecasts (one to four quarters), the principal component representing Household Balance Sheets is a key splitting variable, but its importance fades as the forecast horizon lengthens. Conversely, the principal component representing Housing becomes more prominent as a splitting variable in the five- to seven-quarter-ahead forecasts. Finally, the principal component for Interest Rates shows a noticeable increase in selection frequency at horizons of six to eight quarters ahead.

5 Conclusion

This paper proposes a new framework for macroeconomic forecasting that jointly accounts for nonlinearity, parameter instability, and time-varying predictor relevance. Motivated by the growing evidence that both the set of relevant predictors and their importance evolve over time, especially during periods of structural change, we develop a novel Bayesian forecasting model that we refer to as B-DART. The proposed approach integrates a tree-based regime-partitioning structure with regime-specific SVS based on a Dirac spike-and-slab prior. This unified framework allows the model to simultaneously discover latent regimes, capture nonlinearities, and adapt the set of relevant predictors to changing economic environments.

Using a large quarterly macroeconomic dataset, we conduct an extensive out-of-sample forecasting exercise for eight key U.S. macroeconomic variables across multiple horizons. The empirical results show that the B-DART model consistently outperforms both the Bayesian CART model and conventional stochastic variable selection models, particularly during the post-COVID-19 period characterized by heightened volatility and structural instability. While simpler linear or stable-parameter models remain compet-

itive in periods of relative stability, our findings demonstrate that explicitly modeling nonlinear regime changes together with time-varying variable selection yields substantial gains in forecast accuracy when the underlying economic relationships shift abruptly.

The contribution of this study is threefold. First, from a methodological perspective, we introduce a Bayesian framework that unifies regime discovery and variable selection within a single model. In contrast to existing approaches that treat regime switching and variable selection as separate problems, the B-DART model estimates them jointly, thereby capturing their endogenous interaction. This feature is particularly important in high-dimensional environments, where both the timing of structural change and the set of relevant predictors are inherently uncertain.

Second, we develop a computationally feasible Bayesian estimation strategy by combining a Dirac spike-and-slab prior with a g-prior slab component, which yields an analytically tractable (approximate) marginal likelihood for tree updates. This substantially reduces the computational burden typically associated with high-dimensional Bayesian variable selection in nonlinear models, and makes the joint estimation of tree structures and regime-specific predictor sets practical.

Third, from an empirical perspective, we show that accounting for time-varying variable selection is not merely a modeling refinement. Rather, it can be central to macroeconomic forecasting in unstable environments. In particular, the strong post-COVID-19 performance of the B-DART model highlights the importance of allowing both the functional form and the predictor set to adjust to structural change. Our results suggest that a failure to accommodate either nonlinearity or time-varying predictor relevance can lead to systematically inferior forecasts.

Despite these contributions, several limitations of the current framework point to promising directions for future research. First, the present model treats regime changes as driven solely by observed predictors through the tree structure. An interesting extension would be to incorporate explicit stochastic regime dynamics, for example, by combining the B-DART framework with Markov-switching or change-point mechanisms. This could allow for a richer representation of persistent versus transitory structural shifts.

Second, while we focus on point forecasts and evaluate performance using RMSE, many policy applications require accurate density and tail-risk forecasts. Extending the B-DART framework to distributional forecasting and evaluating its performance using density-based or decision-theoretic criteria would be a natural and important next step. Finally, in this paper we rely on principal components to summarize large information sets before applying the B-DART model. Although this approach is standard and effective, it remains an open question as to how the model would perform when applied directly to very high-dimensional raw predictors using more aggressive shrinkage or sparsity-inducing priors. Exploring such high-dimensional implementations, possibly combined with more scalable computational strategies, is another promising avenue for future work.

References

Auerbach, A. J. and Gorodnichenko, Y. (2012), "Measuring the output responses to fiscal policy," *American Economic Journal: Economic Policy*, 4, 1–27.

Bai, J. and Ng, S. (2008), "Forecasting economic time series using targeted predictors," *Journal of Econometrics*, 146, 304–317.

Bai, J. and Perron, P. (1998), "Estimating and testing linear models with multiple structural changes," *Econometrica*, 47–78.

Bernanke, B. and Blanchard, O. (2023), "What caused the US pandemic-era inflation?" *Peterson Institute for International Economics Working Paper*.

Bok, B., Caratelli, D., Giannone, D., Sbordone, A. M., and Tambalotti, A. (2018), "Macroeconomic nowcasting and forecasting with big data," *Annual Review of Economics*, 10, 615–643.

Chipman, H. A., George, E. I., and McCulloch, R. E. (1998), "Bayesian CART model search," *Journal of the American Statistical Association*, 93, 935–948.

— (2012), "BART: Bayesian additive regression trees," *Annals of Applied Statistics*, 6, 266–298.

Clark, T. E., Huber, F., Koop, G., Marcellino, M., and Pfarrhofer, M. (2023), "Tail forecasting with multivariate Bayesian additive regression trees," *International Economic Review*, 64, 979–1022.

Creal, D. and Kim, J. (2021), "Empirical Asset Pricing with Bayesian Regression Trees," .

Fernandez, C., Ley, E., and Steel, M. F. (2001), "Benchmark priors for Bayesian model averaging," *Journal of Econometrics*, 100, 381–427.

Giannone, D., Lenza, M., and Primiceri, G. E. (2021), "Economic predictions with big data: The illusion of sparsity," *Econometrica*, 89, 2409–2437.

Goulet Coulombe, P., Leroux, M., Stevanovic, D., and Surprenant, S. (2022), "How is machine learning useful for macroeconomic forecasting?" *Journal of Applied Econometrics*, 37, 920–964.

Granger, C. W. and Teräsvirta, T. (1993), *Modelling nonlinear economic relationships*, oxford university Press.

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

Huber, F. and Rossini, L. (2022), "Inference in Bayesian additive vector autoregressive tree models," *The Annals of Applied Statistics*, 16, 104–123.

Jordà, Ò., Singh, S. R., and Taylor, A. M. (2022), "Longer-run economic consequences of pandemics," *Review of Economics and Statistics*, 104, 166–175.

Kim, C.-J. and Nelson, C. R. (1999), "State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications," *MIT Press Books*, 1.

Kim, H. H. and Swanson, N. R. (2018), "Mining big data using parsimonious factor, machine learning, variable selection and shrinkage methods," *International Journal of Forecasting*, 34, 339–354.

Linero, A. R. (2018), "Bayesian regression trees for high-dimensional prediction and variable selection," *Journal of the American Statistical Association*, 113, pp.626–636.

Malsiner-Walli, G. and Wagner, H. (2011), "Comparing Spike and Slab Priors for Bayesian Variable Selection," *Austrian Journal of Statistics*, 40, 241–264.

McCracken, M. W. and Ng, S. (2020), "FRED-QD: A Quarterly Database for Macroeconomic Research," Tech. rep., Federal Reserve Bank of St. Louis.

Medeiros, M. C., Vasconcelos, G. F., Veiga, Á., and Zilberman, E. (2021), "Forecasting inflation in a data-rich environment: the benefits of machine learning methods," *Journal of Business & Economic Statistics*, 39, 98–119.

Mitchell, T. J. and Beauchamp, J. J. (1988), "Bayesian variable selection in linear regression," *Journal of the American Statistical Association*, 83, 1023–1032.

Ng, S. and Wright, J. H. (2013), "Facts and challenges from the great recession for forecasting and macroeconomic modeling," *Journal of Economic Literature*, 51, 1120–1154.

Pesaran, M. H. and Timmermann, A. (1995), "Predictability of stock returns: Robustness and economic significance," *The Journal of Finance*, 50, 1201–1228.

Rapach, D. E. and Wohar, M. E. (2006), "Structural breaks and predictive regression models of aggregate US stock returns," *Journal of Financial Econometrics*, 4, 238–274.

Rossi, B. (2013a), "Advances in forecasting under instability," in *Handbook of economic forecasting*, Elsevier, vol. 2, pp. 1203–1324.

— (2013b), "Exchange rate predictability," *Journal of Economic Literature*, 51, 1063–1119.

— (2021), "Forecasting in the presence of instabilities: How we know whether models predict well and how to improve them," *Journal of Economic Literature*, 59, 1135–1190.

Rossi, B. and Soupre, M. (2017), "Implementing tests for forecast evaluation in the presence of instabilities," *The Stata Journal*, 17, 850–865.

Stock, J. H. and Watson, M. W. (1996), "Evidence on structural instability in macroeconomic time series relations," *Journal of Business & Economic Statistics*, 14, 11–30.

— (2003), "Forecasting output and inflation: The role of asset prices," *Journal of Economic Literature*, 41, 788–829.

— (2006), "Forecasting with many predictors," *Handbook of Economic Forecasting*, 1, 515–554.

Teräsvirta, T. (2006), "Forecasting economic variables with nonlinear models," *Handbook of Economic Forecasting*, 1, 413–457.

Varian, H. R. (2014), "Big data: New tricks for econometrics," *Journal of Economic Perspectives*, 28, 3–28.

Welch, I. and Goyal, A. (2008), "A comprehensive look at the empirical performance of equity premium prediction," *The Review of Financial Studies*, 21, 1455–1508.

Zellner, A. (1986), "On assessing prior distributions and Bayesian regression analysis with g-prior distributions," *Bayesian Inference and Decision Techniques*.

A Deriving the Conditional Marginal Likelihood

We present the details of the derivation of the marginal likelihood for each terminal node when the data generating process (DGP) explicitly considers both the intercept and the coefficients of predictor variables. Then, we prove that this marginal likelihood is equivalent to that obtained using demeaned data without the intercept term. Finally, we derive the conditional marginal likelihood where a Dirac spike-and-slab prior is applied to the coefficient terms.

A.1 Marginal Likelihood of Each Terminal Node (Decentered)

Suppose the spike-and-slab prior is not imposed on β_r and the data generating process (DGP) is specified as:

$$Y = Z \cdot \alpha + \varepsilon$$

where $Z = (\mathbf{1}_T, X)$ and $\alpha = (\mu, \beta)'$. Note that $\mathbf{1}_T$ denotes a vector of ones with $\dim(\mathbf{1}_T) = T \times 1$ and $\dim(Z) = T \times (K+1)$. The DGP of each terminal node (regime) can thus be expressed by explicitly considering both the intercept and the coefficients of predictor variables as follows:

- $Y_r|\mu_r, \beta_r, \sigma_r^2 \sim \mathcal{N}(\mu_r \cdot \mathbf{1}_{T_r} + X_r\beta_r, \sigma_r^2 \cdot I_{T_r})$
- $\beta_r|\sigma_r^2 \sim \mathcal{N}(\beta_{0,r}, \sigma_r^2 \cdot B_{0,r})$
- $\sigma_r^2 \sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu \cdot \lambda}{2})$
- $p(\mu_r) \propto 1$; Improper (flat) prior[19]

[19]This corresponds to assuming that the predictor variables are centered with the null vector as their mean, i.e. $X_r' \cdot \mathbf{1}_{T_r} = \mathbf{0}$. Precisely, let

$$Y_r = \mu_r \cdot \mathbf{1}_{T_r} + X_r \cdot \beta_r + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma_r^2 \cdot I_{T_r}).$$

where $\mathbf{y_{r,c}} = Y_r - \overline{y_r} \cdot \mathbf{1}_{T_r}$ and $\mathbf{X_{r,c}} = X_r - \frac{1}{T_r}\mathbf{1}_{T_r}\mathbf{1}_{T_r}'X_r$. If $X_r' \cdot \mathbf{1}_{T_r} = 0 \iff \mathbf{1}_{T_r}' \cdot X_r = 0$, then $\mathbf{X_{r,c}} = X_r$. Thus, the model simplifies to:

$$\mathbf{y_{r,c}} = X_r \cdot \beta_r + \varepsilon.$$

Let $\Theta = \{\boldsymbol{\beta}_r, \mu_r, \sigma_r^2\}_{r=1}^R$ denote the parameter set that contains the regime-specific regression coefficients and error variances. Then, the marginal likelihood can be simplified analytically as follows:

$$\begin{aligned} p(Y|X,\mathcal{T}) &= \iiint p(Y|X,\Theta,\mathcal{T}) \cdot p(\Theta|\mathcal{T})\, d\Theta \\ &= \iiint \prod_{r=1}^{R} \left(p(Y_r|\mu_r,\beta_r,\sigma_r^2) \cdot p(\beta_r|\sigma_r^2) \cdot p(\mu_r) \cdot p(\sigma_r^2) \right) d\{\beta_r\}\, d\{\mu_r\}\, d\{\sigma_r^2\}. \end{aligned}$$

Focusing on a single regime r for simplicity, we obtain

$$p(Y_r|X_r,\mathcal{T}) = \iiint p(Y_r|\mu_r,\beta_r,\sigma_r^2) \cdot p(\beta_r|\sigma_r^2) \cdot p(\mu_r) \cdot p(\sigma_r^2)\, d\beta_r\, d\mu_r\, d\sigma_r^2.$$

Note that the probability density functions for the likelihood and prior distributions are:

$$p(Y_r|\mu_r,\beta_r,\sigma_r^2) = (2\pi\sigma_r^2)^{-\frac{T_r}{2}} \cdot \exp\left(-\frac{1}{2\sigma_r^2}(Y_r - \mu_r \mathbf{1}_{T_r} - X_r\beta_r)'(Y_r - \mu_r \mathbf{1}_{T_r} - X_r\beta_r) \right),$$

$$p(\beta_r|\sigma_r^2) = (2\pi\sigma_r^2)^{-\frac{K}{2}} \cdot |B_{0,r}|^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2\sigma_r^2}(\beta_r - \beta_{0,r})' B_{0,r}^{-1}(\beta_r - \beta_{0,r}) \right),$$

$$p(\sigma_r^2) = \frac{(\frac{\nu\lambda}{2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} \cdot (\sigma_r^2)^{-\frac{\nu}{2}-1} \cdot \exp\left(-\frac{\nu\lambda}{2\sigma_r^2} \right).$$

To integrate out β_r, we first express the joint distribution of Y_r and β_r as follows:

$$\begin{aligned} &p(Y_r,\beta_r|\mu_r,\sigma_r^2) \\ &= p(Y_r|\mu_r,\beta_r,\sigma_r^2) \cdot p(\beta_r|\sigma_r^2) \\ &= (2\pi\sigma_r^2)^{-\frac{T_r}{2}} \exp\left(-\frac{1}{2\sigma_r^2}(Y_r - \mu_r \mathbf{1}_{T_r} - X_r\beta_r)'(Y_r - \mu_r \mathbf{1}_{T_r} - X_r\beta_r) \right) \\ &\quad \times (2\pi\sigma_r^2)^{-\frac{K}{2}} \cdot |B_{0,r}|^{-\frac{1}{2}} \cdot \exp\left(-\frac{1}{2\sigma_r^2}(\beta_r - \beta_{0,r})' B_{0,r}^{-1}(\beta_r - \beta_{0,r}) \right) \\ &= \underbrace{(2\pi\sigma_r^2)^{-\frac{T_r}{2}} \cdot (2\pi\sigma_r^2)^{-\frac{K}{2}} \cdot |B_{0,r}|^{-\frac{1}{2}}}_{(*)} \\ &\quad \times \exp\left(-\frac{1}{2\sigma_r^2} \underbrace{\{(Y_r - \mu_r \mathbf{1}_{T_r} - X_r\beta_r)'(Y_r - \mu_r \mathbf{1}_{T_r} - X_r\beta_r) + (\beta_r - \beta_{0,r})' B_{0,r}^{-1}(\beta_r - \beta_{0,r})\}}_{(**)} \right). \end{aligned} \tag{A.1}$$

Simplifying the terms in (*) yields

$$(2\pi\sigma_r^2)^{-\frac{T_r+K}{2}} \cdot |B_{0,r}|^{-\frac{1}{2}}. \tag{A.2}$$

For (**), the terms involving β_r can be grouped as follows:

$$\begin{aligned}
&(Y_r - \mu_r \mathbf{1}_{T_r} - X_r\beta_r)'(Y_r - \mu_r \mathbf{1}_{T_r} - X_r\beta_r) \\
&= (Y_r' - \mu_r \mathbf{1}_{T_r}' - \beta_r' X_r')(Y_r - \mu_r \mathbf{1}_{T_r} - X_r\beta_r) \\
&= Y_r'Y_r - 2\mu_r \mathbf{1}_{T_r}' Y_r - 2\beta_r' X_r'(Y_r - \mu_r \mathbf{1}_{T_r}) + \mu_r^2 \cdot T_r + \beta_r' X_r' X_r \beta_r,
\end{aligned}$$

and

$$(\beta_r - \beta_{0,r})' B_{0,r}^{-1}(\beta_r - \beta_{0,r}) = \beta_r' B_{0,r}^{-1}\beta_r - 2\beta_r' B_{0,r}^{-1}\beta_{0,r} + \beta_{0,r}' B_{0,r}^{-1}\beta_{0,r}.$$

which leads to

$$\begin{aligned}
&Y_r'Y_r - 2\mu_r \mathbf{1}_{T_r}' Y_r + \mu_r^2 T_r + \beta_{0,r}' B_{0,r}^{-1}\beta_{0,r} \\
&\quad + \beta_r'(X_r'X_r + B_{0,r}^{-1})\beta_r - 2\beta_r'(X_r'(Y_r - \mu_r \mathbf{1}_{T_r}) + B_{0,r}^{-1}\beta_{0,r}).
\end{aligned} \tag{A.3}$$

Then the terms involving β_r can be further expressed by completing the square for β_r as

$$\begin{aligned}
&\beta_r' B_{1,r}^{-1}\beta_r - 2\beta_r' B_{1,r}^{-1}\beta_{1,r} \\
&= (\beta_r - \beta_{1,r})' B_{1,r}^{-1}(\beta_r - \beta_{1,r}) - \beta_{1,r}' B_{1,r}^{-1}\beta_{1,r},
\end{aligned} \tag{A.4}$$

where $B_{1,r}^{-1} = X_r'X_r + B_{0,r}^{-1}$ and $B_{1,r}^{-1}\beta_{1,r} = X_r'(Y_r - \mu_r \mathbf{1}_{T_r}) + B_{0,r}^{-1}\beta_{0,r}$, leading to

$$B_{1,r} = (X_r'X_r + B_{0,r}^{-1})^{-1} \quad \text{and} \quad \beta_{1,r} = B_{1,r}(X_r'(Y_r - \mu_r \mathbf{1}_{T_r}) + B_{0,r}^{-1}\beta_{0,r}).$$

Therefore, by substituting equation (A.4) into (A.3), it follows that

$$\Rightarrow Y_r'Y_r - 2\mu_r \mathbf{1}_{T_r}' Y_r + \mu_r^2 T_r + \beta_{0,r}' B_{0,r}^{-1}\beta_{0,r} + (\beta_r - \beta_{1,r})' B_{1,r}^{-1}(\beta_r - \beta_{1,r}) - \beta_{1,r}' B_{1,r}^{-1}\beta_{1,r} \tag{A.5}$$

Now, by applying results (A.2) and (A.5) to (A.1),

$$p(Y_r, \beta_r | \mu_r, \sigma_r^2) = (2\pi\sigma_r^2)^{-\frac{T_r+K}{2}} \cdot |B_{0,r}|^{-\frac{1}{2}}$$

$$\times \exp\left(-\frac{1}{2\sigma_r^2}(Y_r'Y_r - 2\mu_r \mathbf{1}_{T_r}'Y_r + \mu_r^2 T_r + \beta_{0,r}'B_{0,r}^{-1}\beta_{0,r} - \beta_{1,r}'B_{1,r}^{-1}\beta_{1,r})\right)$$
$$\times \exp\left(-\frac{1}{2\sigma_r^2}(\beta_r - \beta_{1,r})'B_{1,r}^{-1}(\beta_r - \beta_{1,r})\right)$$

Finally, we can integrate out β_r as follows.

$$\begin{aligned}
\int p(Y_r, \beta_r|\mu_r, \sigma_r^2)\, d\beta_r &= (2\pi\sigma_r^2)^{-\frac{T_r+K}{2}} \cdot |B_{0,r}|^{-\frac{1}{2}} \\
&\quad \times \exp\left(-\frac{1}{2\sigma_r^2}(Y_r'Y_r - 2\mu_r \mathbf{1}_{T_r}'Y_r + \mu_r^2 T_r + \beta_{0,r}'B_{0,r}^{-1}\beta_{0,r} - \beta_{1,r}'B_{1,r}^{-1}\beta_{1,r})\right) \\
&\quad \times (2\pi\sigma_r^2)^{\frac{K}{2}}|B_{1,r}|^{\frac{1}{2}} \\
&\quad \times \underbrace{\int (2\pi\sigma_r^2)^{-\frac{K}{2}}|B_{1,r}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma_r^2}(\beta_r - \beta_{1,r})'B_{1,r}^{-1}(\beta_r - \beta_{1,r})\right) d\beta_r}_{=1 \text{ since } \mathcal{N}(\beta_{1,r}, \sigma_r^2 B_{1,r})} \\
&= (2\pi\sigma_r^2)^{-\frac{T_r}{2}} \left(\frac{|B_{1,r}|}{|B_{0,r}|}\right)^{\frac{1}{2}} \\
&\quad \times \exp\left(-\frac{1}{2\sigma_r^2}(Y_r'Y_r - 2\mu_r \mathbf{1}_{T_r}'Y_r + \mu_r^2 T_r + \beta_{0,r}'B_{0,r}^{-1}\beta_{0,r} - \beta_{1,r}'B_{1,r}^{-1}\beta_{1,r})\right) \\
&\equiv p(Y_r|\mu_r, \sigma_r^2)
\end{aligned}$$

Next, we can integrate out μ_r. The joint distribution of Y_r and μ_r is as follows:

$$\begin{aligned}
p(Y_r, \mu_r|\sigma_r^2) &= p(Y_r|\mu_r, \sigma_r^2) \cdot p(\mu_r) \\
&= (2\pi\sigma_r^2)^{-\frac{T_r}{2}} \cdot \left(\frac{|B_{1,r}|}{|B_{0,r}|}\right)^{\frac{1}{2}} \\
&\quad \times \exp\left(-\frac{1}{2\sigma_r^2}\Big\{ \underbrace{Y_r'Y_r - 2\mu_r \mathbf{1}_{T_r}'Y_r + \mu_r^2 T_r}_{\text{terms involving } \mu_r} + \underbrace{\beta_{0,r}'B_{0,r}^{-1}\beta_{0,r} - \beta_{1,r}'B_{1,r}^{-1}\beta_{1,r}}_{\text{terms not involving } \mu_r} \Big\}\right)
\end{aligned} \tag{A.6}$$

Then the terms involving μ_r can be further expressed by completing the square for μ_r as follows:

$$\Rightarrow Y_r'Y_r - 2\mu_r \mathbf{1}_{T_r}'Y_r + \mu_r^2 T_r = Y_r'Y_r + T_r(\mu_r - \overline{y_r})^2 - T_r\overline{y_r}^2 \tag{A.7}$$

where

$$\overline{y_r} = \frac{1}{T_r}\mathbf{1}'_{T_r} Y_r.$$

Substituting equation (A.7) into (A.6), we obtain

$$p(Y_r, \mu_r|\sigma_r^2) = (2\pi\sigma_r^2)^{-\frac{T_r}{2}} \cdot \left(\frac{|B_{1,r}|}{|B_{0,r}|}\right)^{\frac{1}{2}} \times \exp\left(-\frac{1}{2\sigma_r^2}\Big\{ Y_r'Y_r + T_r(\mu_r - \overline{y_r})^2 - T_r\overline{y_r}^2 + \beta_{0,r}'B_{0,r}^{-1}\beta_{0,r} - \beta_{1,r}'B_{1,r}^{-1}\beta_{1,r}\Big\}\right)$$

Finally, we integrate out μ_r as:

$$\begin{aligned}
\int p(Y_r, \mu_r|\sigma_r^2)\, d\mu_r &= (2\pi\sigma_r^2)^{-\frac{T_r}{2}} \cdot \left(\frac{|B_{1,r}|}{|B_{0,r}|}\right)^{\frac{1}{2}} \\
&\quad \times \exp\left(-\frac{1}{2\sigma_r^2}(Y_r'Y_r - T_r\overline{y_r}^2 + \beta_{0,r}'B_{0,r}^{-1}\beta_{0,r} - \beta_{1,r}'B_{1,r}^{-1}\beta_{1,r})\right) \\
&\quad \times (2\pi\sigma_r^2)^{\frac{1}{2}} \cdot T_r^{-\frac{1}{2}} \cdot \underbrace{\int (2\pi\sigma_r^2)^{-\frac{1}{2}} T_r^{\frac{1}{2}} \exp\left(-\frac{T_r}{2\sigma_r^2}(\mu_r - \overline{y_r})^2\right) d\mu_r}_{=1 \text{ since } \mathcal{N}(\overline{y_r}, \frac{1}{T_r}\sigma_r^2)} \\
&= (2\pi\sigma_r^2)^{-\frac{(T_r-1)}{2}} \cdot T_r^{-\frac{1}{2}} \cdot \left(\frac{|B_{1,r}|}{|B_{0,r}|}\right)^{\frac{1}{2}} \cdot \exp\left(-\frac{1}{2\sigma_r^2}A\right) \\
&\equiv p(Y_r|\sigma_r^2)
\end{aligned}$$

where

$$A \equiv Y_r'Y_r + \beta_{0,r}'B_{0,r}^{-1}\beta_{0,r} - \beta_{1,r}'B_{1,r}^{-1}\beta_{1,r} - T_r \cdot \overline{y_r}^2.$$

Next, we can integrate out σ_r^2. The joint distribution of Y_r and σ_r^2 is as follows:

$$\begin{aligned}
p(Y_r, \sigma_r^2) &= p(Y_r|\sigma_r^2) \cdot p(\sigma_r^2) \\
&= (2\pi)^{-\frac{(T_r-1)}{2}} (\sigma_r^2)^{-\frac{(T_r-1)}{2}} \cdot T_r^{-\frac{1}{2}} \cdot \left(\frac{|B_{1,r}|}{|B_{0,r}|}\right)^{\frac{1}{2}} \\
&\quad \times \exp\left(-\frac{1}{2\sigma_r^2}A\right) \cdot \frac{\left(\frac{\nu\lambda}{2}\right)^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} (\sigma_r^2)^{-\frac{\nu}{2}-1} \exp\left(-\frac{1}{2\sigma_r^2}\nu\lambda\right) \\
&= (2\pi)^{-\frac{(T_r-1)}{2}} \cdot T_r^{-\frac{1}{2}} \cdot \left(\frac{|B_{1,r}|}{|B_{0,r}|}\right)^{\frac{1}{2}} \frac{\left(\frac{\nu\lambda}{2}\right)^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})}
\end{aligned}$$

$$\times (\sigma_r^2)^{-\frac{(T_r-1)}{2}-\frac{\nu}{2}-1} \exp\left(-\frac{1}{2\sigma_r^2}(A+\nu\lambda)\right)$$

Integrating out σ_r^2,

$$\begin{aligned}
&\int p(Y_r, \sigma_r^2)\, d\sigma_r^2 \\
&= (2\pi)^{-\frac{(T_r-1)}{2}} \cdot T_r^{-\frac{1}{2}} \cdot \left(\frac{|B_{1,r}|}{|B_{0,r}|}\right)^{\frac{1}{2}} \frac{\left(\frac{\nu\lambda}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \\
&\quad \times \int (\sigma_r^2)^{-\frac{(T_r-1)}{2}-\frac{\nu}{2}-1} \exp\left(-\frac{1}{2\sigma_r^2}(A+\nu\lambda)\right) d\sigma_r^2 \\
&= (2\pi)^{-\frac{(T_r-1)}{2}} \cdot T_r^{-\frac{1}{2}} \cdot \left(\frac{|B_{1,r}|}{|B_{0,r}|}\right)^{\frac{1}{2}} \frac{\left(\frac{\nu\lambda}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \frac{\Gamma\left(\frac{(T_r-1)+\nu}{2}\right)}{\left(\frac{A+\nu\lambda}{2}\right)^{\frac{(T_r-1)+\nu}{2}}} \\
&\quad \times \underbrace{\int \frac{\left(\frac{A+\nu\lambda}{2}\right)^{\frac{(T_r-1)+\nu}{2}}}{\Gamma\left(\frac{(T_r-1)+\nu}{2}\right)} (\sigma_r^2)^{-\frac{(T_r-1)}{2}-\frac{\nu}{2}-1} \exp\left(-\frac{1}{2\sigma_r^2}(A+\nu\lambda)\right) d\sigma_r^2}_{=1 \text{ since } \mathcal{IG}\left(\frac{(T_r-1)+\nu}{2}, \frac{A+\nu\lambda}{2}\right)} \\
&= (2\pi)^{-\frac{(T_r-1)}{2}} \left(\frac{|B_{1,r}|}{|B_{0,r}|}\right)^{\frac{1}{2}} \left(\frac{\nu\lambda}{2}\right)^{\frac{\nu}{2}} \frac{\Gamma\left(\frac{(T_r-1)+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \cdot \left(\frac{A+\nu\lambda}{2}\right)^{-\frac{(T_r-1)+\nu}{2}} \\
&= (\pi)^{-\frac{(T_r-1)}{2}} (\nu\lambda)^{\frac{\nu}{2}} \left(\frac{|B_{1,r}|}{|B_{0,r}|}\right)^{\frac{1}{2}} \frac{\Gamma\left(\frac{(T_r-1)+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \cdot (A+\nu\lambda)^{-\frac{(T_r-1)+\nu}{2}}
\end{aligned}$$

Consequently, the marginal likelihood can be analytically simplified as:

$$p(Y_r|X_r, \mathcal{T}) = \prod_{r=1}^{G} \left\{ (\pi)^{-\frac{(T_r-1)}{2}} (\nu\lambda)^{\frac{\nu}{2}} \left(\frac{|B_{1,r}|}{|B_{0,r}|}\right)^{\frac{1}{2}} \frac{\Gamma\left(\frac{(T_r-1)+\nu}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \cdot (A+\nu\lambda)^{-\frac{(T_r-1)+\nu}{2}} \right\} \quad \text{(A.8)}$$

where

$$A = Y_r'Y_r + \beta_{0,r}'B_{0,r}^{-1}\beta_{0,r} - \beta_{1,r}'B_{1,r}^{-1}\beta_{1,r} - T_r \cdot \overline{y_r}^2 \quad \text{(A.9)}$$

$$B_{1,r} = (X_r'X_r + B_{0,r}^{-1})^{-1} \quad \text{(A.10)}$$

$$\beta_{1,r} = B_{1,r}\left(X_r'(Y_r - \mu_r \mathbf{1}_{T_r}) + B_{0,r}^{-1}\beta_{0,r}\right) \quad \text{(A.11)}$$

A.2 Equivalence of Marginal Likelihood under Centered Data

In the initial DGP of the previous section, there are three parameters to be estimated: μ_r, β_r, and σ_r^2. However, by assuming that the data are centered, i.e., both Y_r and X_r are demeaned, one can avoid estimating μ_r and instead consider only β_r and σ_r^2. That is, suppose the following DGP:

- $\mathbf{y_{r,c}}|\beta_r, \sigma_r^2 \sim \mathcal{N}(X_r \cdot \beta_r, \sigma_r^2 \cdot I_{T_r})$
- $\beta_r|\sigma_r^2 \sim \mathcal{N}(\beta_{0,r}, \sigma_r^2 \cdot B_{0,r})$
- $\sigma_r^2 \sim \mathcal{IG}(\frac{\nu}{2}, \frac{\nu\lambda}{2})$

where $\mathbf{y_{r,c}} = Y_r - \overline{y_r} \cdot \mathbf{1}_{T_r} = Y_r - \frac{1}{T_r}\mathbf{1}_{T_r}\mathbf{1}'_{T_r}Y_r$,since $\overline{y_r} = \frac{1}{T_r}\mathbf{1}'_{T_r}Y_r$, and $X_r = \mathbf{X_{r,c}}$. [20] Then the marginal likelihood of each terminal node is analytically given by:

$$p(\mathbf{y_{r,c}}|X_r) = \iint p(\mathbf{y_{r,c}}|\beta_r, \sigma_r^2) \cdot p(\beta_r|\sigma_r^2) \cdot p(\sigma_r^2)\, d\beta_r\, d\sigma_r^2.$$

Rather than repeating the detailed process of integrating out β_r and σ_r^2 as above, we confirm the equivalence by showing that the key components A, $B_{1,r}$, and $\beta_{1,r}$ are unchanged compared to the decentered data case.

First, we can easily check that $B_{1,r}$ is the same as in equation (A.10) since $X_r = \mathbf{X_{r,c}}$, i.e.,

$$B_{1,r} = (X_r'X_r + B_{0,r}^{-1})^{-1}.$$

Next, we show that $\beta_{1,r}$ is the same as in equation (A.11).

Proposition 1. *Suppose that the predictor variables are centered with the null vector as their mean, i.e., $X_r' \cdot \mathbf{1}_{T_r} = 0$. Then*

$$X_r'(Y_r - \mu_r \cdot \mathbf{1}_{T_r}) = X_r' \cdot \mathbf{y_{r,c}}.$$

Proof. By definition, $\mathbf{y_{r,c}} = Y_r - \overline{y_r} \cdot \mathbf{1}_{T_r}$, where $\overline{y_r} = \frac{1}{T_r}\mathbf{1}'_{T_r}Y_r$. Then

$$X_r' \cdot \mathbf{y_{r,c}} = X_r'(Y_r - \overline{y_r} \cdot \mathbf{1}_{T_r})$$

[20] This is due to the fact that $p(\mu_r) \propto 1$. For details, refer to footnote 19.

$$
\begin{aligned}
&= X_r'\Big(Y_r - \frac{1}{T_r}\mathbf{1}_{T_r} \cdot \mathbf{1}_{T_r}' Y_r\Big), \quad \text{since } \dim(\mathbf{1}_{T_r}' Y_r) = 1, \\
&= X_r' Y_r - \frac{1}{T_r} X_r' \mathbf{1}_{T_r} \cdot \mathbf{1}_{T_r}' Y_r \\
&= X_r' Y_r,
\end{aligned}
$$

where the last equality is due to $X_r' \cdot \mathbf{1}_{T_r} = 0$. Thus,

$$
\begin{aligned}
X_r'(Y_r - \mu_r \cdot \mathbf{1}_{T_r}) &= X_r' Y_r - \mu_r \cdot X_r' \mathbf{1}_{T_r} \\
&= X_r' Y_r, \quad \text{since } X_r' \cdot \mathbf{1}_{T_r} = 0, \\
&= X_r' \cdot \mathbf{y_{r,c}}.
\end{aligned}
$$

□

Therefore, Proposition 1 implies that $\beta_{1,r}$ is the same as in the decentered data case.

Finally, we show that A is the same as in equation (A.9).

Proposition 2. *The following equality holds:*

$$
\mathbf{y_{r,c}}' \cdot \mathbf{y_{r,c}} = Y_r' Y_r - T_r \cdot \overline{y_r}^2.
$$

Proof. By definition, $\mathbf{y_{r,c}} = Y_r - \overline{y_r} \cdot \mathbf{1}_{T_r}$, where $\overline{y_r} = \frac{1}{T_r}\mathbf{1}_{T_r}' Y_r$. Then

$$
\begin{aligned}
\mathbf{y_{r,c}}' \cdot \mathbf{y_{r,c}} &= (Y_r - \overline{y_r} \cdot \mathbf{1}_{T_r})'(Y_r - \overline{y_r} \cdot \mathbf{1}_{T_r}) \\
&= (Y_r' - \overline{y_r} \cdot \mathbf{1}_{T_r}')(Y_r - \overline{y_r} \cdot \mathbf{1}_{T_r}) \\
&= Y_r' Y_r - \overline{y_r} \cdot Y_r' \mathbf{1}_{T_r} - \overline{y_r} \cdot \mathbf{1}_{T_r}' Y_r + \overline{y_r}^2 \cdot \mathbf{1}_{T_r}' \mathbf{1}_{T_r} \\
&= Y_r' Y_r - \overline{y_r} \cdot T_r \cdot \overline{y_r} - \overline{y_r} \cdot T_r \cdot \overline{y_r} + \overline{y_r}^2 \cdot T_r, \\
&\quad \text{since } \mathbf{1}_{T_r}' Y_r = Y_r' \mathbf{1}_{T_r} = T_r \cdot \overline{y_r}, \\
&= Y_r' Y_r - T_r \cdot \overline{y_r}^2.
\end{aligned}
$$

□

By Proposition 2, we observe that A is the same as that of the decentered data case.

Since A, $B_{1,r}$, and $\beta_{1,r}$ are identical to equations (A.9) to (A.11), we can conclude that the marginal likelihood under the centered and decentered data is the same. Therefore,

for simplicity of estimation, we use the centered (demeaned) data and estimate only β_r and σ_r^2 for each terminal node.

A.3 Marginal Likelihood under Spike-and-Slab Prior

We consider the case where the spike-and-slab prior is applied to β_r. For this case, the data generating process (DGP) of each terminal node under centered data is given as follows:

- $\mathbf{y_{r,c}} \mid \beta_r, \sigma_r^2 \sim \mathcal{N}(X_r\beta_r, \sigma_r^2 I_{T_r})$
- $\beta_r \mid \sigma_r^2, \boldsymbol{\delta}_r \sim p(\beta_r \mid \sigma_r^2, \boldsymbol{\delta}_r) = \pi_{\text{slab}}(\beta_{r,\delta}) \cdot \prod_{k:\delta_k=0} \pi_{\text{spike}}(\beta_{r,k})$

  where
  $$\begin{cases} \beta_{r,\delta} \sim \mathcal{N}(\beta_{0,\delta}, \sigma_r^2 \cdot B_{0,r,\delta}) \\ \beta_{r,-\delta} \sim \text{density function with all mass at zero} \end{cases}$$
- $\delta_{r,k} \sim \text{Bernoulli}(p_r)$
- $p_r \sim \text{Beta}(a_0, c_0)$
- $\sigma_r^2 \sim \text{InverseGamma}(\frac{\nu}{2}, \frac{\nu\lambda}{2})$

Note that $\beta_{r,\delta}$ denotes all coefficients $\beta_{r,k}$ for which $\delta_{r,k} = 1$, while $\beta_{r,-\delta}$ denotes those coefficients for which $\delta_{r,k} = 0$.

Let the slab prior have mean and variance given by a g-slab as follows:

$$\beta_{0,\delta} = 0 \text{ and } B_{0,r,\delta} = g \cdot (X'_{r,\delta} X_{r,\delta})^{-1} \tag{A.12}$$

where $X_{r,\delta}$ is the design matrix consisting of the predictor variables and $\beta_{r,k}$ with $\delta_{r,k} = 1$. Then given $\boldsymbol{\delta}_r$, the marginal likelihood in equation (A.8) becomes

$$f(\mathbf{y_{r,c}}|X_r, \boldsymbol{\delta}_r) = (\pi)^{-\frac{(T_r-1)}{2}} \cdot T_r^{-\frac{1}{2}} \cdot (\nu\lambda)^{\frac{\nu}{2}} \cdot \left(\frac{|B_{1,r,\delta}|}{|B_{0,r,\delta}|}\right)^{\frac{1}{2}} \cdot \frac{\Gamma(\frac{(T_r-1)+\nu}{2})}{\Gamma(\frac{\nu}{2})} \cdot (A_{r,\delta} + \nu\lambda)^{-\frac{(T_r-1)+\nu}{2}} \tag{A.13}$$

where

$$A_{r,\delta} = \mathbf{y_{r,c}}'\mathbf{y_{r,c}} + \beta'_{0,\delta} B^{-1}_{0,r,\delta} \beta_{0,\delta} - \beta'_{1,r,\delta} B^{-1}_{1,r,\delta} \beta_{1,r,\delta}$$

$$B_{1,r,\delta} = (X'_{r,\delta}X_{r,\delta} + B^{-1}_{0,r,\delta})^{-1}$$
$$\beta_{1,r,\delta} = B_{1,r,\delta}(X'_{r,\delta}\mathbf{y_{r,c}} + B^{-1}_{0,r,\delta}\beta_{0,\delta})$$

Now using $\beta_{0,\delta}$ and $B_{0,r,\delta}$ of the g-slab, we can further simplify equation (A.13) as follows. First,

$$B^{-1}_{0,r,\delta} = \frac{1}{g}(X'_{r,\delta}X_{r,\delta})$$

leading to

$$B_{1,r,\delta} = (X'_{r,\delta}X_{r,\delta} + \frac{1}{g}(X'_{r,\delta}X_{r,\delta}))^{-1} = (\frac{g+1}{g}X'_{r,\delta}X_{r,\delta})^{-1} = \frac{g}{g+1}(X'_{r,\delta}X_{r,\delta})^{-1} \quad \text{(A.14)}$$

and

$$\beta_{1,r,\delta} = B_{1,r,\delta}(X'_{r,\delta}\mathbf{y_{r,c}} + \frac{1}{g}(X'_{r,\delta}X_{r,\delta})\beta_{0,\delta}) = B_{1,r,\delta}X'_{r,\delta}\mathbf{y_{r,c}} \quad \text{(A.15)}$$

Further, $A_{r,\delta}$ can be expressed as follows:

$$\begin{aligned} A_{r,\delta} &= \mathbf{y_{r,c}}'\mathbf{y_{r,c}} + \beta'_{0,\delta}B^{-1}_{0,r,\delta}\beta_{0,\delta} - \beta'_{1,r,\delta}B^{-1}_{1,r,\delta}\beta_{1,r,\delta} \\ &= \mathbf{y_{r,c}}'\mathbf{y_{r,c}} - (B_{1,r,\delta}X'_{r,\delta}\mathbf{y_{r,c}})'B^{-1}_{1,r,\delta}(B_{1,r,\delta}X'_{r,\delta}\mathbf{y_{r,c}}) \\ &= \mathbf{y_{r,c}}'\mathbf{y_{r,c}} - \mathbf{y_{r,c}}'X_{r,\delta}B'_{1,r,\delta}B^{-1}_{1,r,\delta}B_{1,r,\delta}X'_{r,\delta}\mathbf{y_{r,c}} \\ &= \mathbf{y_{r,c}}'\mathbf{y_{r,c}} - \mathbf{y_{r,c}}'X_{r,\delta}B_{1,r,\delta}X'_{r,\delta}\mathbf{y_{r,c}} \end{aligned} \quad \text{(A.16)}$$

Finally, using the fact that $|a \cdot A| = a^K \cdot |A|$ where $\dim(A) = K$, the ratio of determinants of the prior and posterior variance of β_r can be simplified as:

$$\frac{|B_{1,r,\delta}|^{\frac{1}{2}}}{|B_{0,r,\delta}|^{\frac{1}{2}}} = \frac{|\frac{g}{g+1}(X'_{r,\delta}X_{r,\delta})^{-1}|^{\frac{1}{2}}}{|g(X'_{r,\delta}X_{r,\delta})^{-1}|^{\frac{1}{2}}} = \frac{(\frac{g}{g+1})^{\frac{K_r}{2}}}{g^{\frac{K_r}{2}}} = (g+1)^{-\frac{K_r}{2}} \quad \text{(A.17)}$$

This is because $B_{1,r,\delta}$ is a scalar multiple of $B_{0,r,\delta}$ as shown in equation (A.14) above. Therefore, the simplified conditional marginal likelihood given $\boldsymbol{\delta}_r$ is

$$f(\mathbf{y_{r,c}}|X_r, \boldsymbol{\delta}_r) = (\pi)^{-\frac{(T_r-1)}{2}}\, T_r^{-\frac{1}{2}}\, (\nu\lambda)^{\frac{\nu}{2}} \cdot (g+1)^{-\frac{K_r}{2}} \cdot \frac{\Gamma(\frac{(T_r-1)+\nu}{2})}{\Gamma(\frac{\nu}{2})} \cdot (A_{r,\delta}+\nu\lambda)^{-\frac{(T_r-1)+\nu}{2}} \quad \text{(A.18)}$$

where

$$A_{r,\delta} = \mathbf{y_{r,c}}'\mathbf{y_{r,c}} - \mathbf{y_{r,c}}'X_{r,\delta}B_{1,r,\delta}X'_{r,\delta}\mathbf{y_{r,c}} \qquad B_{1,r,\delta} = \frac{g}{g+1}(X'_{r,\delta}X_{r,\delta})^{-1}$$

B MCMC Sampling for Stochastic Variable Selection

In this section, we describe the MCMC sampling procedure for SVS in each terminal node r, which means the regression tree $\mathcal{T}$ and split rules $\mathcal{I}$ are given. We omit $\{\mathcal{T}, \mathcal{I}\}$ in the conditional information set for notational simplicity.

The target posterior distribution for SVS can be decomposed as

$$p(\boldsymbol{\beta}_r, \sigma_r^2, \boldsymbol{\delta}_r | \mathbf{y_{r,c}}, p_r) = p(\boldsymbol{\beta}_r | \mathbf{y_{r,c}}, p_r, \sigma_r^2, \boldsymbol{\delta}_r) \cdot p(\sigma_r^2 | \mathbf{y_{r,c}}, p_r, \boldsymbol{\delta}_r) \cdot p(\boldsymbol{\delta}_r | \mathbf{y_{r,c}}, p_r). \tag{B.1}$$

B.1 Posterior Sampling of $\boldsymbol{\delta}_r$

We first illustrate how to sample $\boldsymbol{\delta}_r$ based on the relation:

$$p(\boldsymbol{\delta}_r | \mathbf{y_{r,c}}, p_r) \propto p(\mathbf{y_{r,c}} | \boldsymbol{\delta}_r, p_r) \cdot p(\boldsymbol{\delta}_r | p_r)$$

The posterior distribution of $\delta_{r,k}$ is proportional to the product of the likelihood and the prior distributions as follows:

$$\begin{aligned}
&p(\delta_{r,k} | \mathbf{y_{r,c}}, \boldsymbol{\beta}_r, \sigma_r^2, \delta_{r,-k}, p_r) \\
&\propto p(\mathbf{y_{r,c}} | X_r \boldsymbol{\beta}_r, \sigma_r^2) \cdot p(\boldsymbol{\beta}_r | \sigma_r^2, \delta_{r,k}, \delta_{r,-k}) \cdot p(\sigma_r^2) \cdot p(\delta_{r,k} | p_r) \cdot p(\delta_{r,-k} | p_r) \cdot p(p_r) \\
&\propto p(\mathbf{y_{r,c}} | \delta_{r,k}, \delta_{r,-k}) \cdot p(\delta_{r,k} | p_r)
\end{aligned}$$

where $\delta_{r,-k}$ denotes all $\delta_{r,\cdot}$ values except for $\delta_{r,k}$. Using the property that $\delta_{r,k}$ is a discrete variable,

$$\begin{aligned}
&p(\delta_{r,k} = 1 | \delta_{r,-k}, \mathbf{y_{r,c}}) \\
&= \frac{p(\delta_{r,k} = 1 | p_r) \cdot p(\mathbf{y_{r,c}} | \delta_{r,k} = 1, \delta_{r,-k})}{p(\delta_{r,k} = 1 | p_r) \cdot p(\mathbf{y_{r,c}} | \delta_{r,k} = 1, \delta_{r,-k}) + p(\delta_{r,k} = 0 | p_r) \cdot p(\mathbf{y_{r,c}} | \delta_{r,k} = 0, \delta_{r,-k})}.
\end{aligned} \tag{B.2}$$

Now using the conditional marginal likelihood given $\boldsymbol{\delta}_r$ in equation (A.18), equation (B.2) can be further simplified as:

$$\begin{aligned}
&p(\delta_{r,k} = 1 | \delta_{r,-k}, \mathbf{y_{r,c}}) \\
&= \frac{p_r (g+1)^{-\frac{K_r}{2}} (A_{r,\delta_k} + \nu\lambda)^{-\frac{(T_r - 1) + \nu}{2}}}{p_r (g+1)^{-\frac{K_r}{2}} (A_{r,\delta_k} + \nu\lambda)^{-\frac{(T_r - 1) + \nu}{2}} + (1 - p_r)(g+1)^{-\frac{K_{-r}}{2}} (A_{r,\delta_{-k}} + \nu\lambda)^{-\frac{(T_r - 1) + \nu}{2}}}
\end{aligned}$$

$$= \frac{p_r}{p_r + (1-p_r)(g+1)^{\frac{1}{2}}\left(\frac{A_{r,\delta_{-k}}+\nu\lambda}{A_{r,\delta_k}+\nu\lambda}\right)^{-\frac{(T_r-1)+\nu}{2}}} \tag{B.3}$$

where K_{-r} is the number of predictor variables with $\delta_{r,k} = 0$, $K_r = K_{-r} + 1$, $A_{r,\delta_{-k}}$ and A_{r,δ_k} are each $A_{r,\delta}$ derived for the two cases of $\delta_{r,k} = 0$ and $\delta_{r,k} = 1$.

B.2 Posterior Sampling of σ_r^2

Next, σ_r^2 is sampled using the relation:

$$p(\sigma_r^2|\mathbf{y_{r,c}}, p_r, \boldsymbol{\delta}_r) \propto p(\mathbf{y_{r,c}}|\sigma_r^2, \boldsymbol{\delta}_r) \cdot p(\sigma_r^2) \tag{B.4}$$

Note that equation (B.4) can be derived from integrating out σ_r^2 in *Appendix* A as follows:

$$p(\sigma_r^2|\mathbf{y_{r,c}}, \boldsymbol{\delta}_r) \propto \mathcal{IG}(\frac{\alpha_{r,1}}{2}, \frac{\gamma_{r,1}}{2}) \tag{B.5}$$

where $\alpha_{r,1} = \nu + T_r$ and $\gamma_{r,1} = A_{r,\delta} + \nu\lambda$ with $A_{r,\delta} = \mathbf{y_{r,c}}'\mathbf{y_{r,c}} - \mathbf{y_{r,c}}'X_{r,\delta}B_{1,r,\delta}X_{r,\delta}'\mathbf{y_{r,c}}$ in equation (A.18).

B.3 Posterior Sampling of $\boldsymbol{\beta}_r$

We sample $\boldsymbol{\beta}_r$ using the relation:

$$p(\boldsymbol{\beta}_r|\mathbf{y_{r,c}}, p_r, \sigma_r^2, \boldsymbol{\delta}_r) \propto p(\mathbf{y_{r,c}}|\boldsymbol{\beta}_r, \sigma_r^2, \boldsymbol{\delta}_r) \cdot p(\boldsymbol{\beta}_r|\sigma_r^2, \boldsymbol{\delta}_r) \tag{B.6}$$

Note that $\beta_{r,-\delta}$ has a point mass at zero, that is, $\beta_{r,k} = 0$ for all k such that $\delta_{r,k} = 0$. The full conditional distribution for the non-zero elements in $\boldsymbol{\beta}_r$ from equation (B.6) can be expressed as shown in *Appendix* A:

$$p(\beta_{r,\delta}|\mathbf{y_{r,c}}, \sigma_r^2) = \mathcal{N}(\beta_{1,r,\delta}, \sigma_r^2 \cdot B_{1,r,\delta}) \tag{B.7}$$

where $\beta_{1,r,\delta} = B_{1,r,\delta}(X_{r,\delta}'\mathbf{y_{r,c}} + B_{0,r,\delta}^{-1}\beta_{0,\delta})$ and $B_{1,r,\delta} = (X_{r,\delta}'X_{r,\delta} + B_{0,r,\delta}^{-1})^{-1}$. Under the g-slab, $\beta_{1,r,\delta}$ and $B_{1,r,\delta}$ are given as follows:

$$\beta_{1,r,\delta} = B_{1,r,\delta}X_{r,\delta}'\mathbf{y_{r,c}}$$

and

$$B_{1,r,\delta} = \frac{g}{g+1}(X'_{r,\delta}X_{r,\delta})^{-1}$$

B.4 Posterior Sampling of p_r

Finally, the full conditional distribution of p_r for each group (or terminal node) is proportional to the likelihood and the product of prior distribution as follows:

$$\begin{aligned} p(p_r|\boldsymbol{\beta}_r, \sigma_r^2, \boldsymbol{\delta}_r, \mathbf{y_{r,c}}) &\propto p(\boldsymbol{\delta}_r|p_r) \cdot p(p_r) \\ &= p_r^{K_r} \cdot (1-p_r)^{K_{-r}} \cdot \frac{\Gamma(a_0+c_0)}{\Gamma(a_0)\Gamma(c_0)} \cdot p_r^{a_0-1} \cdot (1-p_r)^{c_0-1} \\ &= \text{Beta}(a_0+K_r, c_0+K_{-r}) \end{aligned}$$

where K_{-r} is the number of predictor variables with $\delta_{r,k} = 0$ and K_r is the number of predictor variables with $\delta_{r,k} = 1$.

C Acceptance Probabilities for the Metropolis-Hastings Algorithm

Here we derive the acceptance probabilities for the four candidate moves of the MH algorithm. Recall that the MH acceptance probability for the candidate tree $\mathcal{T}^*$ and splitting rules $\mathcal{I}^*$, conditional on the current state $(\mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)})$, was given in Equation (9).

C.1 Acceptance Probability of the Grow Move

If the grow move is chosen, a randomly selected terminal node $\eta \in \mathcal{T}^{(j-1)}$ is split into two child nodes according to a new splitting rule. The likelihood ratio (a) and the prior ratio (b) of the acceptance probability in Equation (9) are easy to evaluate. The proposal ratio (c) requires deriving the transition probabilities. The probability of proposing the candidate tree $\mathcal{T}^*$ given $\mathcal{T}^{(j-1)}$ for the grow move is:

$$q(\mathcal{T}^*, \mathcal{I}^*|\mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)}) = p_{\text{move}} \times \frac{1}{M(\mathcal{T}^{(j-1)})} \times p(\text{rule})$$

where $p_{\text{move}} = 1/4$ is the probability of choosing the grow move, $M(\mathcal{T}^{(j-1)})$ is the number of terminal nodes in the current tree, and $p(\text{rule})$ is the probability of selecting the specific splitting rule for the new internal node. Conversely, the probability of returning to $\mathcal{T}^{(j-1)}$ given $\mathcal{T}^*$ by the prune move is:

$$q(\mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)} | \mathcal{T}^*, \mathcal{I}^*) = p_{\text{move}} \times \frac{1}{W(\mathcal{T}^*)}$$

where $W(\mathcal{T}^*)$ denotes the number of internal nodes in $\mathcal{T}^*$ that have two terminal child nodes (i.e., nodes available for pruning).

C.2 Acceptance Probability of the Prune Move

If the prune move is chosen, a randomly selected internal node with two terminal children is collapsed into a single terminal node. Here, $\mathcal{T}^{(j-1)}$ is the previous tree and $\mathcal{T}^*$ is the pruned candidate. Once again, evaluating (a) and (b) of Equation (9) is straightforward. The proposal ratio (c) is the inverse of that of the grow move. The probability of proposing the pruned tree $\mathcal{T}^*$ given $\mathcal{T}^{(j-1)}$ is:

$$q(\mathcal{T}^*, \mathcal{I}^* | \mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)}) = p_{\text{move}} \times \frac{1}{W(\mathcal{T}^{(j-1)})}$$

The probability of proposing the larger tree $\mathcal{T}^{(j-1)}$ given $\mathcal{T}^*$ via the prune move is:

$$q(\mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)} | \mathcal{T}^*, \mathcal{I}^*) = p_{\text{move}} \times \frac{1}{M(\mathcal{T}^*)} \times p(\text{rule})$$

where $M(\mathcal{T}^*)$ is the number of terminal nodes in the pruned tree. Thus, the grow and prune moves are reversible counterparts.

C.3 Acceptance Probability of the Change Move

If the change move is chosen, a randomly selected internal node is reassigned a new splitting rule. Since the tree structure topology remains constant (only $\mathcal{I}$ changes), the number of internal nodes is identical for both $\mathcal{T}^{(j-1)}$ and $\mathcal{T}^*$. The probability of proposing a local change to $\mathcal{I}^*$ given $\mathcal{I}^{(j-1)}$ is:

$$q(\mathcal{T}^*, \mathcal{I}^* | \mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)}) = p_{\text{move}} \times \frac{1}{N(\mathcal{T})} \times p(\text{new rule})$$

where $N(\mathcal{T})$ is the number of internal nodes. The reverse move involves selecting the same node and proposing the original rule:

$$q(\mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)} | \mathcal{T}^*, \mathcal{I}^*) = p_{\text{move}} \times \frac{1}{N(\mathcal{T})} \times p(\text{old rule}).$$

If the prior distribution for splitting rules is symmetric or uniform, the proposal ratio (c) simplifies to 1.

C.4 Acceptance Probability of the Swap Move

If the swap move is chosen, the splitting rules between a randomly selected parent and child node (where both are internal nodes) are swapped. Let $S(\mathcal{T})$ denote the number of parent-child pairs available for swapping. Since the tree topology does not change, $S(\mathcal{T}^{(j-1)}) = S(\mathcal{T}^*)$. The probability of proposing the swap is:

$$q(\mathcal{T}^*, \mathcal{I}^* | \mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)}) = p_{\text{move}} \times \frac{1}{S(\mathcal{T}^{(j-1)})}.$$

The reverse probability is:

$$q(\mathcal{T}^{(j-1)}, \mathcal{I}^{(j-1)} | \mathcal{T}^*, \mathcal{I}^*) = p_{\text{move}} \times \frac{1}{S(\mathcal{T}^*)}.$$

Consequently, the proposal ratio (c) for the swap move is exactly 1. Therefore, the acceptance probability only depends on the likelihood and prior ratios. Similar to the relationship between grow and prune, the change and swap moves preserve the dimensionality of the parameter space.

D Number of Principal Components Across Out-of-Sample Periods

Figure (D.1) displays the number of principal components across the out-of-sample period, based on the proportion of total variation explained by the PCs within each group. One can observe that the number of PCs varies across the out-of-sample period, where the 26th period is Q3:2020 and the 27th period is Q4:2020. Next, since we select 0.4 as the threshold for determining the principal components, Table (D.1) reports the

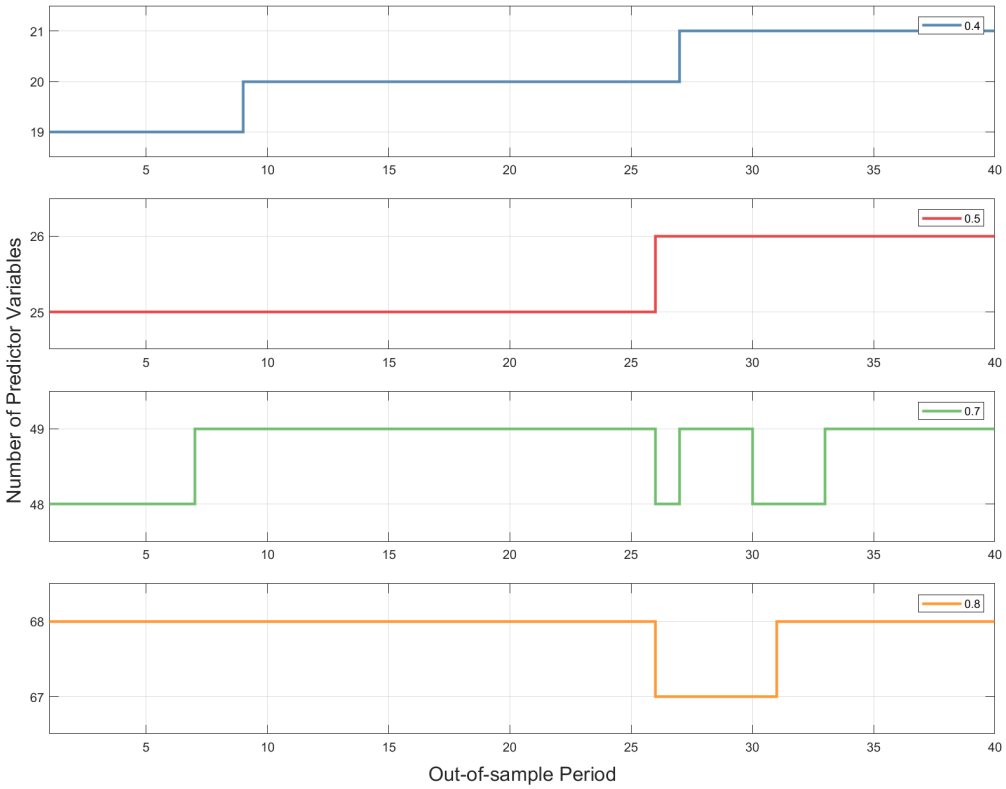


Figure D.1: Number of Principal Components

Note: The figure shows the number of principal components selected over the out-of-sample period based on the proportion of total variance explained within each group. The number of components varies over time, reflecting changes in the underlying data structure, particularly around the COVID-19 period.

specific components retained for each category group.

E Selected Groups for B-CART Model

The average number of groups obtained from the B-CART model across the eight macroeconomic variables and forecast horizons is depicted in Figure D.2. The averages are calculated as the sample mean of groups across the optimal trees estimated over 40 out-of-sample periods. The results show that, on average, only a single group was estimated for most macroeconomic variables, with the exception of the PPI price index,

Table D.1: Predictor Variables across Out-of-Sample Periods

	$1-8$	$9-26$	$27-40$
1	Y(-1)	Y(-1)	Y(-1)
2	Y(-2)	Y(-2)	Y(-2)
3	Y(-3)	Y(-3)	Y(-3)
4	Y(-4)	Y(-4)	Y(-4)
5	NIPA1	NIPA1	NIPA1
6	NIPA2	NIPA2	NIPA2
7	IP1	IP1	IP1
8	EM1	EM1	EM1
9			EM2
10	HOUSE1	HOUSE1	HOUSE1
11	INV1	INV1	INV1
12	PRICE1	PRICE1	PRICE1
13	PRICE2	PRICE2	PRICE2
14	EAPROD1	EAPROD1	EAPROD1
15	EAPROD2	EAPROD2	EAPROD2
16	INT1	INT1	INT1
17	M&C1	M&C1	M&C1
18	M&C2	M&C2	M&C2
19		M&C3	M&C3
20	HBAL1	HBAL1	HBAL1
21	EXCH1	EXCH1	EXCH1
22	STOCK1	STOCK1	STOCK1
23	NHBAL1	NHBAL1	NHBAL1
24	NHBAL2	NHBAL2	NHBAL2
25	SENTI1	SENTI1	SENTI1

Notes: The table reports the predictor variables retained across different out-of-sample periods based on the principal component selection criterion. The selected predictors are largely stable over time, with some variation across sub-periods.

the federal funds effective rate (FFE), and the unemployment rate. Even for these three variables, where some degree of grouping was observed, the average number of groups quickly converge to one as the forecast horizon increases. In other words, the B-CART

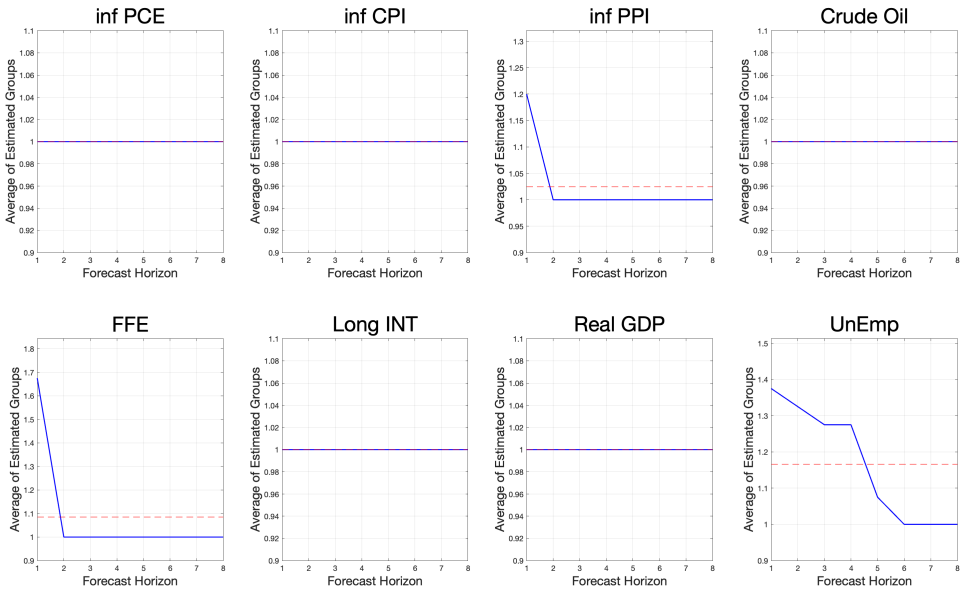


Figure D.2: Posterior Mean of the Regime Count for B-CART

Note: This figure depicts the estimated average number of groups for macroeconomic variables in the B-CART model. The blue line shows the average number of groups across each forecast horizon, while the red dotted line indicates the overall mean across all eight forecast horizons.

model generally did not partition the sample period through the tree structure. This finding implies that jointly estimating variable selection and the optimal tree structure creates a compounding effect. Specifically, incorporating variable selection appears essential for accurately estimating the optimal tree structure when the timing and source of instability among variables are unknown.

F Simulated Application

In this section, we illustrate the in-sample performance of the Bayesian Dirac Classification and Regression Tree (B-DART) model by applying it to a simulated data set corresponding to the tree depicted in Figure 1 of the main text. As stated in Rossi (2013a), although strong in-sample fit does not necessarily guarantee robust out-of-sample forecast accuracy, a poor in-sample fit can undermine out-of-sample forecasts. Therefore, it is crucial to examine the in-sample performance of the B-DART model.

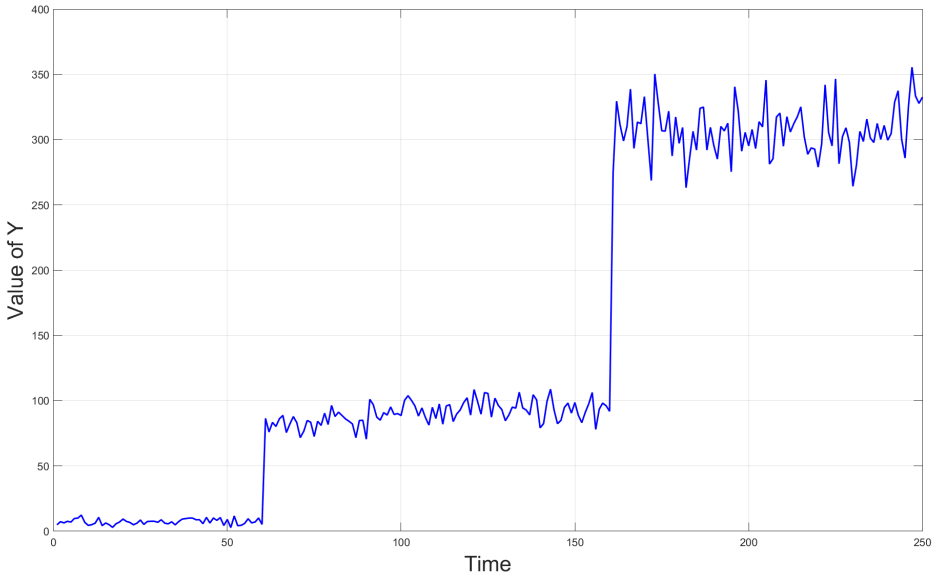


Figure F.1: Simulated Data of Y

Note. This figure depicts the simulated data for 250 periods using the same data generating process as Figure 1 in the main text. Y is split into three groups based on the values of X_1 and X_2.

Moreover, this exercise is useful as it demonstrates how various hyperparameter values in the B-DART model are set, which are later employed in the empirical application of forecasting eight macroeconomic variables in Section 4.

The data generating process (DGP) of the simulated data can be expressed as follows:

$$y_t = \left[\sum_{r=1}^{3} (d_t = r) \mathbf{x}_t' \boldsymbol{\beta}_r \right] + \varepsilon_t, \; \varepsilon_t \sim \mathcal{N}(0, \sum_{r=1}^{R} \mathbb{1}(d_t = r) \sigma_r^2)$$

where

$$\begin{cases} \{\boldsymbol{\beta_1} = (1, 2, 0, \cdots, 0)', \; \sigma_1^2 = 1\} & \text{if } X_1 \leq 0.5 \text{ and } X_2 \leq 0.5 \\ \{\boldsymbol{\beta_2} = (0, 5, 6, 0, \cdots, 0)', \; \sigma_2^2 = 2\} & \text{if } X_1 \leq 0.5 \text{ and } X_2 > 0.5 \\ \{\boldsymbol{\beta_3} = (11, 12, 13, 0, \cdots, 0)', \; \sigma_3^2 = 4\} & \text{if } X_1 > 0.5 \end{cases}$$

Note that the sample size is set to $T = 250$, matching the actual quarterly data used in the empirical application. The simulated data for y_t are shown in Figure F.1. The hyperparameter values of each prior distribution were set to be the same as the empirical

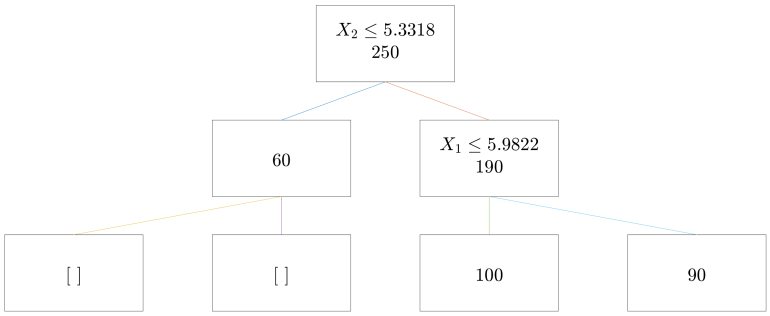


Figure F.2: Estimated Optimal Tree Structure by the B-DART Model

Note: This figure depicts the optimal tree structure estimated by the B-DART model. One can observe that the structure is well estimated and aligns with the data generating process.

application case, except for the case of σ_r^2. In particular, we set $\nu = 5$ and $\lambda = 3$ such that $\mathbb{E}[\sigma_r^2] = (\nu \cdot \lambda/2)/(\nu/2 - 1) = 15/3$.

One characteristic of the Metropolis-Hastings (MH) algorithm used to generate the candidate trees is that the acceptance ratio, i.e. the MH ratio, is significantly lower than typical values. This contrasts with the conventional guideline that a good proposal distribution should yield a MH ratio between 0.4 and 0.8. The reason for this phenomenon is that it is highly likely that the structure of a tree with high conditional marginal likelihood is not persistent. Moreover, the structure of the deeper trees, i.e. the trees that are generated afterwards, depends heavily on the initial tree. To address this, we generate $20,000$ trees while resetting the initial tree every $2,500$ iterations.[21] At the terminal node of every generated tree, the sampling process for Dirac variable selection is performed for $1,000$ iterations following a 10 percent burn-in period.

The optimal tree structure estimated by the B-DART model is shown in Figure F.2. It can be observed that the splits align well with the true data generating process based on X_1 and X_2. Additionally, Figure F.3 shows that the conditional marginal likelihood

[21] Note that reducing the number of generated trees to $10,000$ while resetting it every $2,500$ times does not significantly affect the results. Thus, we later generate $10,000$ trees when working with empirical data to reduce computational burden.

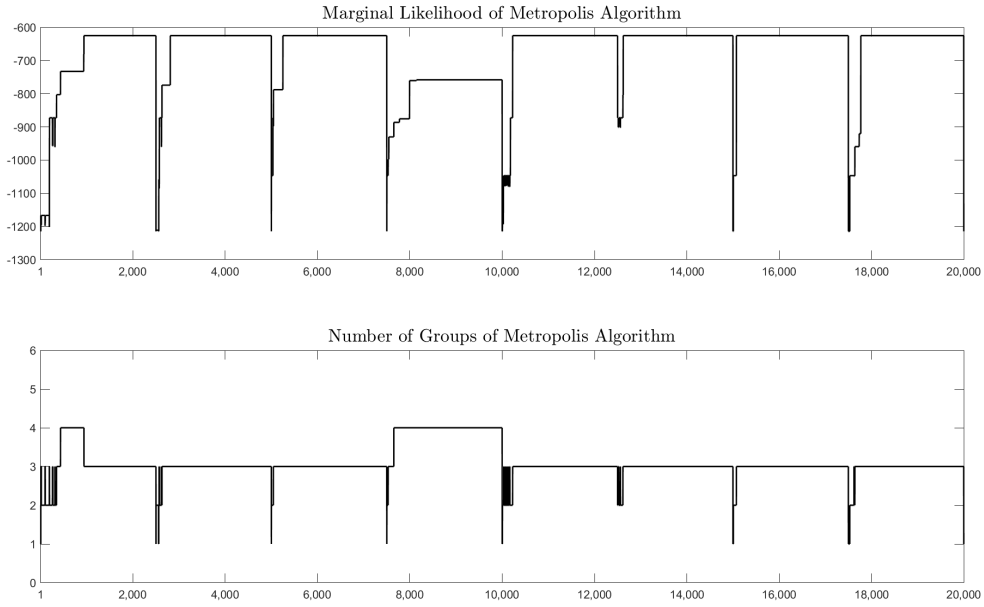


Figure F.3: Marginal Likelihood and Estimated Number of Groups

Note. This figure depicts marginal likelihood and estimated number of groups of the estimated trees across 20,000 iterations. Spikes occur every 2,500 iterations due to resetting the generation of trees.

quickly converges after each reset of the initial tree as mentioned above. The estimated number of groups appears to be stable across iteration cycles, averaging around 3.09, which reflects the stopping rule used.

Given the selected tree structure in Figure F.2, we estimate the parameters for each group given $\boldsymbol{\delta}_r$, i.e. using only the predictors selected by Dirac variable selection. To evaluate the performance of the B-DART model, these estimates are compared to those obtained by applying Dirac variable selection and Bayesian CART (B-CART) model independently. Table F.1 presents the estimated parameter values for each method. Note that for both alternative models, same prior values as in the B-DART model are used.

First, the Dirac variable selection model produces the most heavily biased estimators relative to the true parameters. This demonstrates that ignoring the time-varying variable selection leads to biased parameter estimates and potentially worsens forecast

Table F.1: **Comparison of Results**

Variable	Dirac	B-CART			B-DART		
		Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
β_1	32.8858	1.0798	0.1681	10.8773	1.0330	0.0009	10.9113
β_2	11.4080	1.9473	5.0407	12.3007	1.9183	5.0450	12.1585
β_3	10.1561	0.2167	6.0618	13.2887	0.0025	6.0771	13.1673
β_4	-0.0007	0.0216	0.0303	-0.1005	0.0001	0.0002	-0.0011
β_5	0.0006	-0.0552	0.0347	-0.1654	-0.0010	0.0001	-0.0197
β_6	0.0031	0.0608	0.0210	0.0092	0.0007	0.0003	0.0001
β_7	-0.0014	0.0165	-0.0123	0.0482	0.0005	0.0001	0.0004
β_8	0.0008	0.0631	-0.0737	-0.0311	0.0008	-0.0013	-0.0001
β_9	0.0034	0.1099	0.0411	-0.0855	0.0025	0.0003	-0.0010
β_{10}	-0.0016	0.0008	-0.0299	0.0299	0.0001	-0.0002	0.0001
σ^2	657.1956	1.1033	2.0332	4.1300	1.2270	1.9814	4.0148
ML	-1202.505		-634.092			-625.053	

performance. By contrast, the B-DART model not only produces parameter estimates comparable to those derived from the B-CART model, but also shows a significant improvement in the conditional marginal likelihood. For these reasons, the B-DART model achieves a better in-sample fit compared to the Dirac variable selection model, while also proving to be a sufficient and robust alternative to the B-CART model.

Finally, the posterior distributions obtained from applying Dirac variable selection to the simulated data are depicted in Figure F.4 below. One can observe that X_1, X_2, and X_3 were estimated to be important, i.e. $\hat{\delta}_k \approx 1$. Conversely, the importance of other predictors was close to zero, implying that they follow the degenerate spike distribution.

G Hyperparameter Tuning Process

When generating the trees using the B-DART model, we have found that the tree structure depends strongly on the prior of σ_r^2, i.e.

$$\sigma_r^2 \sim \mathcal{IG}\Big(\frac{\nu}{2}, \frac{\nu\lambda}{2}\Big)$$

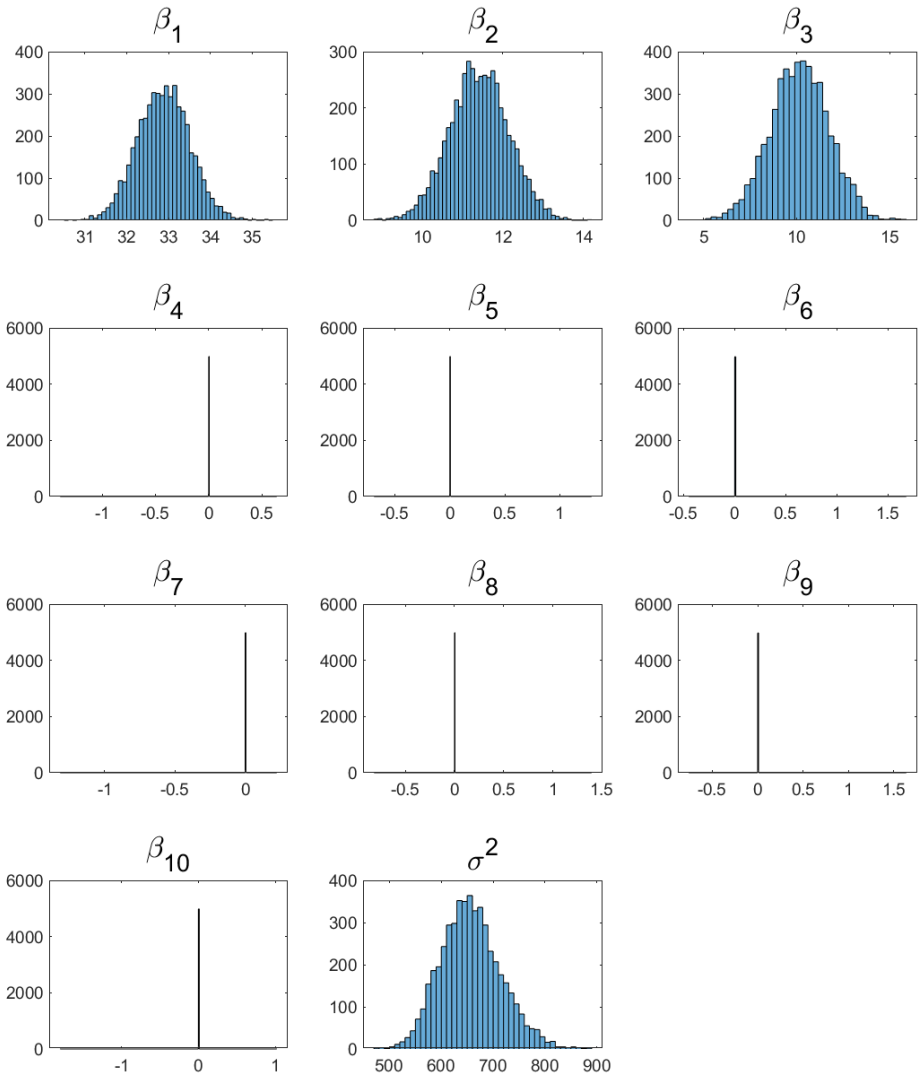


Figure F.4: Posterior Distributions for Simulated Data

Note. This figure depicts the estimated posterior distributions of each parameter for the simulated data. β_1, β_2, and β_3 follow the slab distribution, while the remaining coefficients follow the spike distribution with all mass at zero. However, β_1, β_2 and β_3 are inaccurately estimated as σ^2 is estimated to be extremely large.

Therefore, we apply a tuning procedure to the hyperparameter λ of σ_r^2. Specifically, Chipman et al. (2012) states that the hyperparameters of σ_r^2 can be calibrated using the

Table G.1: Selected Value of Hyper-parameter Tuning

	inf PCE			inf CPI			inf PPI			Crude Oil		
H	AR(1)	B-DART	Dirac	AR(1)	B-DART	Dirac	AR(1)	B-DART	Dirac	AR(1)	B-DART	Dirac
1	3	3	5	1	2	4	1	2	2	2	3	1
2	1	2	5	1	5	3	3	4	4	2	2	3
3	5	5	1	5	5	4	4	2	2	4	1	4
4	1	5	5	3	1	5	2	1	2	4	3	2
5	2	5	1	2	5	4	3	4	1	3	1	4
6	4	3	3	1	3	4	1	5	3	2	1	1
7	2	5	4	2	2	5	3	5	1	4	3	4
8	3	3	5	5	1	3	5	2	5	1	3	5
	FFE			Long INT			rGDP			Unemp		
H	AR(1)	B-DART	Dirac	AR(1)	B-DART	Dirac	AR(1)	B-DART	Dirac	AR(1)	B-DART	Dirac
1	1	1	3	5	1	1	2	4	4	1	5	4
2	4	3	1	1	1	1	5	3	5	2	1	4
3	2	4	2	5	5	5	3	4	3	5	4	5
4	4	1	1	3	5	2	2	5	2	1	5	3
5	3	4	4	2	4	3	1	4	2	4	2	4
6	2	3	3	1	4	4	2	3	3	4	5	5
7	5	3	4	1	4	1	1	1	4	4	1	3
8	5	3	2	5	2	3	3	3	5	1	4	4

Notes: The table reports the selected tuning parameter $l \in \{1, \ldots, 5\}$, which determines the prior scale $\lambda = \hat{\sigma}_r^2/l$, for each model, macroeconomic variable, and forecast horizon $H = 1, \ldots, 8$. Smaller values of l imply a larger prior variance λ and hence weaker shrinkage, whereas larger values indicate stronger shrinkage.

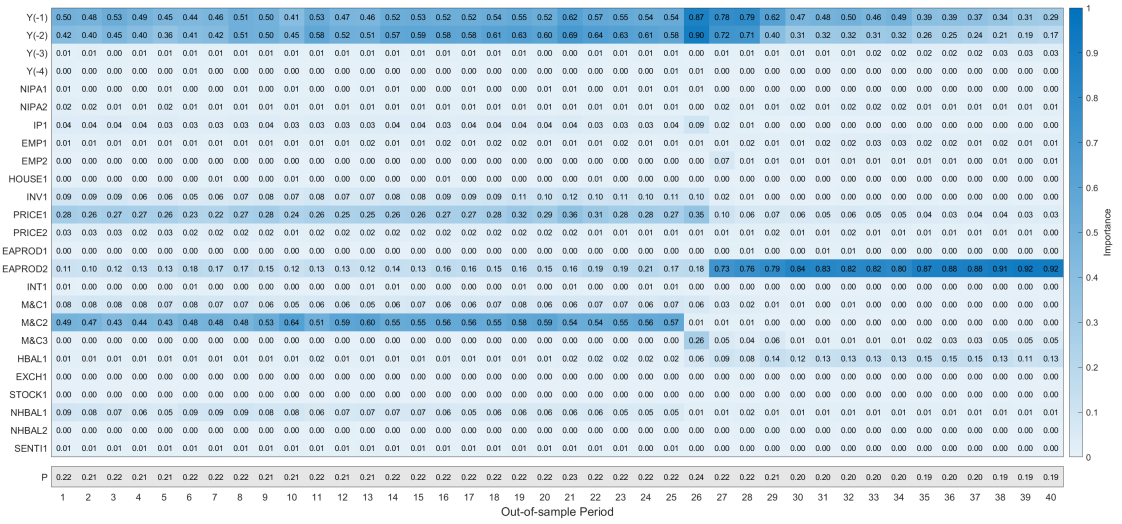
data driven estimate $\hat{\sigma}_r^2$. One method that we apply is the "naive" specification where $\hat{\sigma}_r$ is taken to be the sample standard deviation of the response variable Y. We set $\nu = 5$ and tune λ such that

$$\lambda = \left\{ \frac{\hat{\sigma}_r^2}{l} \right\}_{l=1}^{5}$$

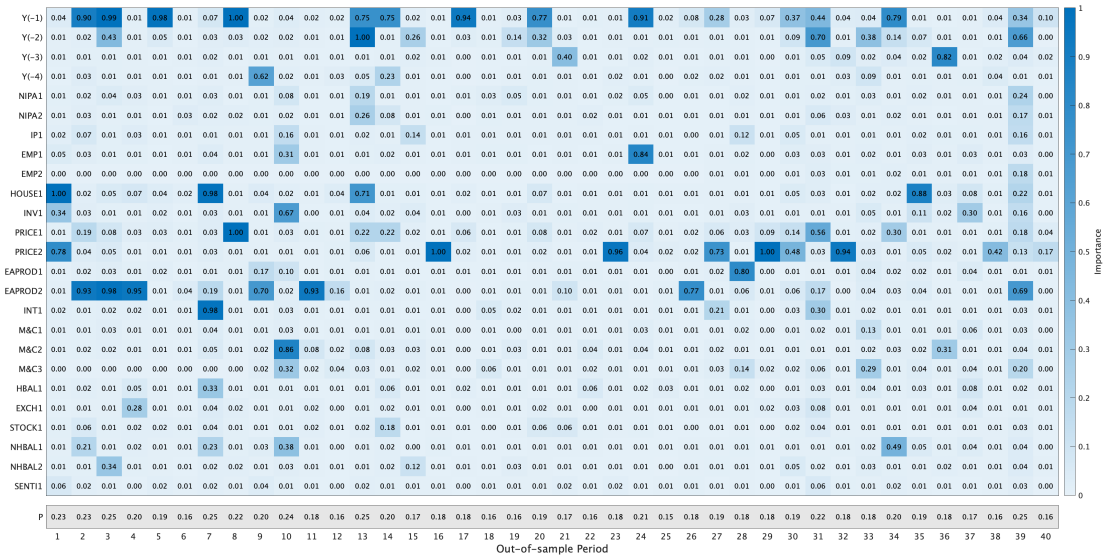
Consequently, we tune the hyper-parameter λ five times for each of the eight forecast horizons for a macroeconomic variable.

The selected values of l for each forecast horizon and macroeconomic variable are reported in Table G.1, based on the RMSE criterion for point forecasts. The table summarizes how the degree of shrinkage, governed by the prior scale λ, varies across models, variables, and forecast horizons. The smaller (larger) values of l imply a larger (smaller) prior variance λ, which leads to weaker (stronger) shrinkage.

H Selected Predictors and Their Importance: Additional Results



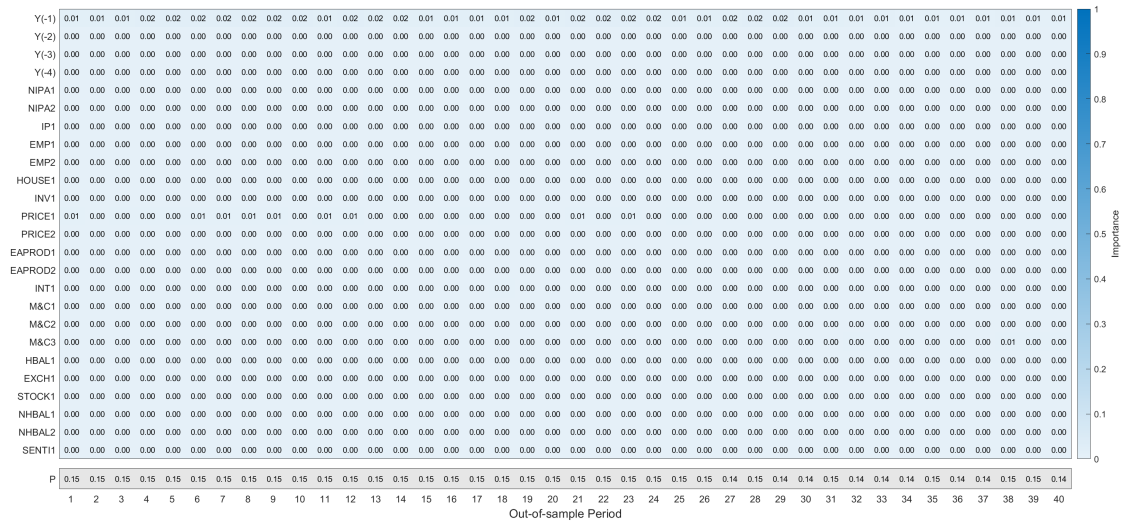
(a) Dirac Model Case



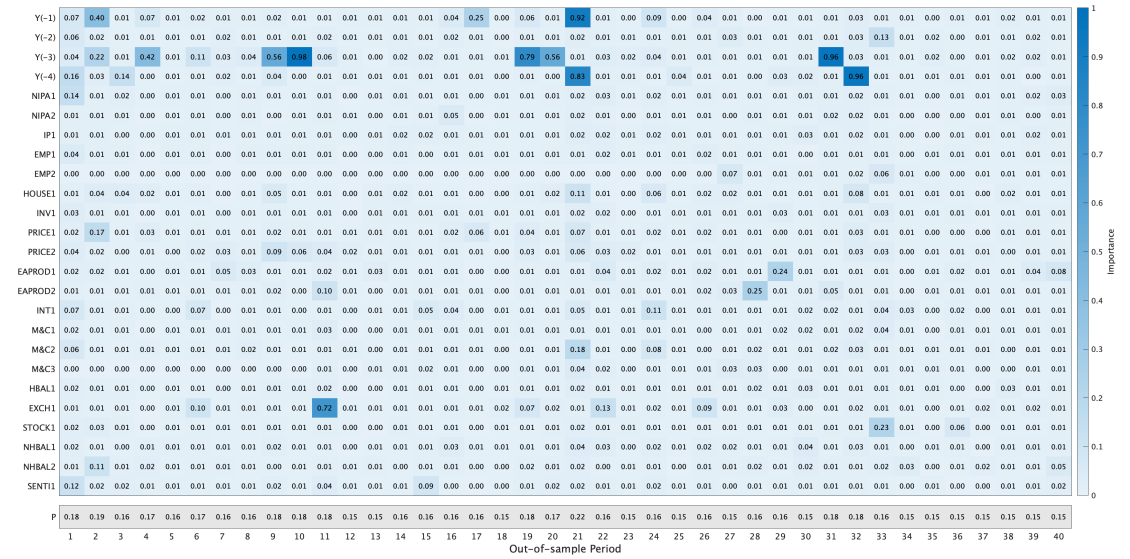
(b) B-DART Model Case

Figure H.1: Importance for Forecasting 1-quarter-ahead PCE Inflation Rate

Note: The figure above depicts the importance of each predictor variable for 40 out-of-sample periods when forecasting the 1-quarter-ahead PCE inflation rate. The heatmap becomes darker as the estimated importance of the predictor variable approaches 1. The last row is the estimated inclusion probability.



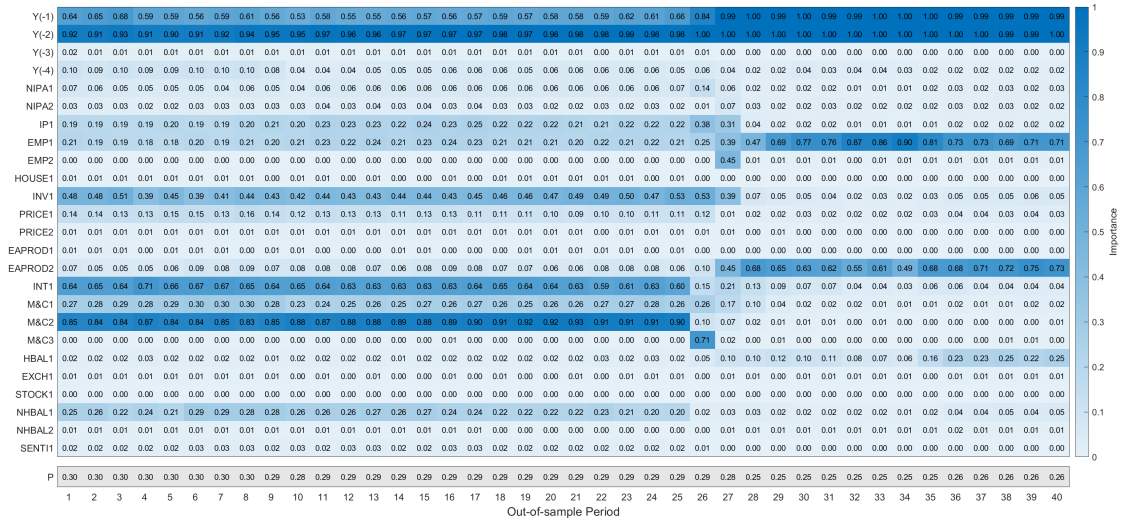
(a) Dirac Model Case



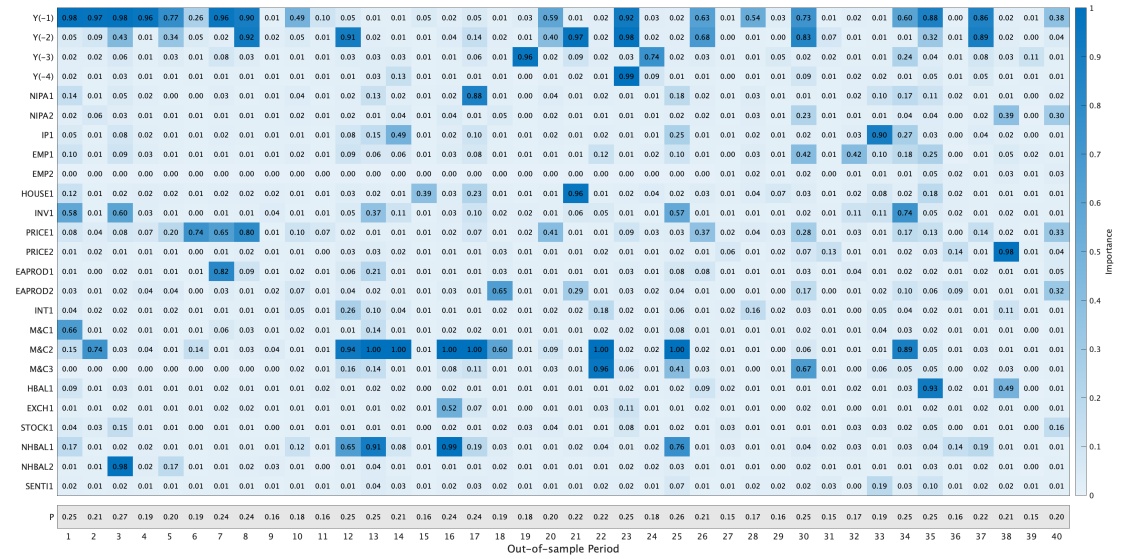
(b) B-DART Model Case

Figure H.2: Importance for Forecasting 6-quarter-ahead PCE Inflation Rate

Note: The figure above depicts the importance of each predictor variable for 40 out-of-sample periods when forecasting the 6-quarter-ahead PCE inflation rate. The heatmap becomes darker as the estimated importance of the predictor variable approaches 1. The last row is the estimated inclusion probability.



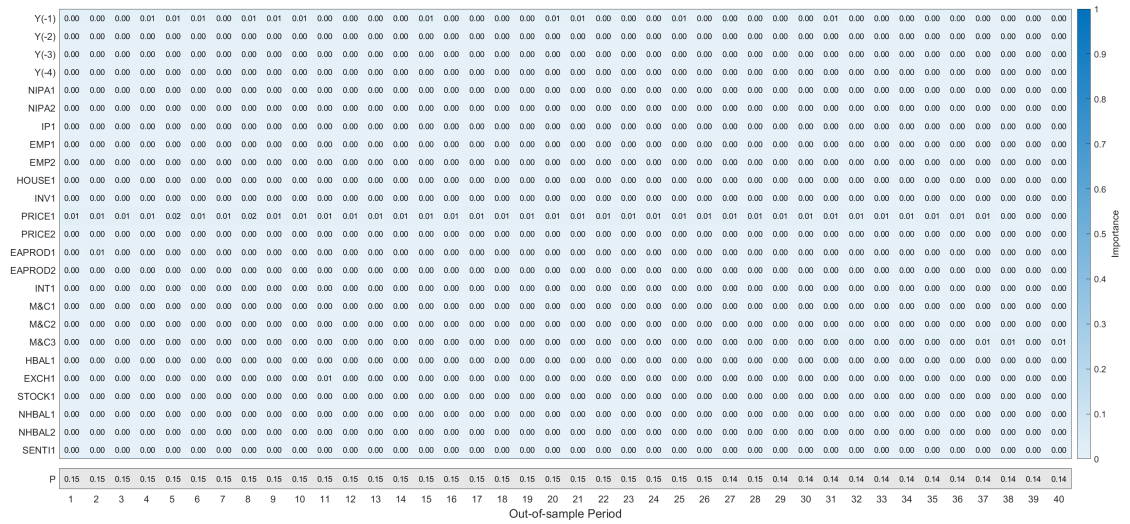
(a) Dirac Model Case



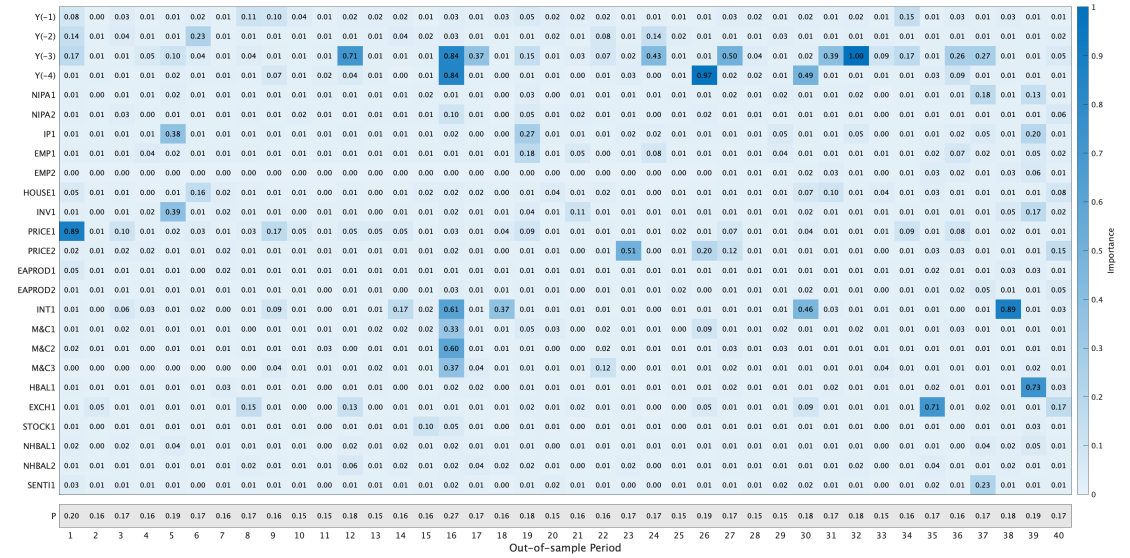
(b) B-DART Model Case

Figure H.3: Importance for Forecasting 1-quarter-ahead CPI Inflation Rate

Note: The figure above depicts the importance of each predictor variable for 40 out-of-sample periods when forecasting the 1-quarter-ahead CPI inflation rate. The heatmap becomes darker as the estimated importance of the predictor variable approaches 1. The last row is the estimated inclusion probability.



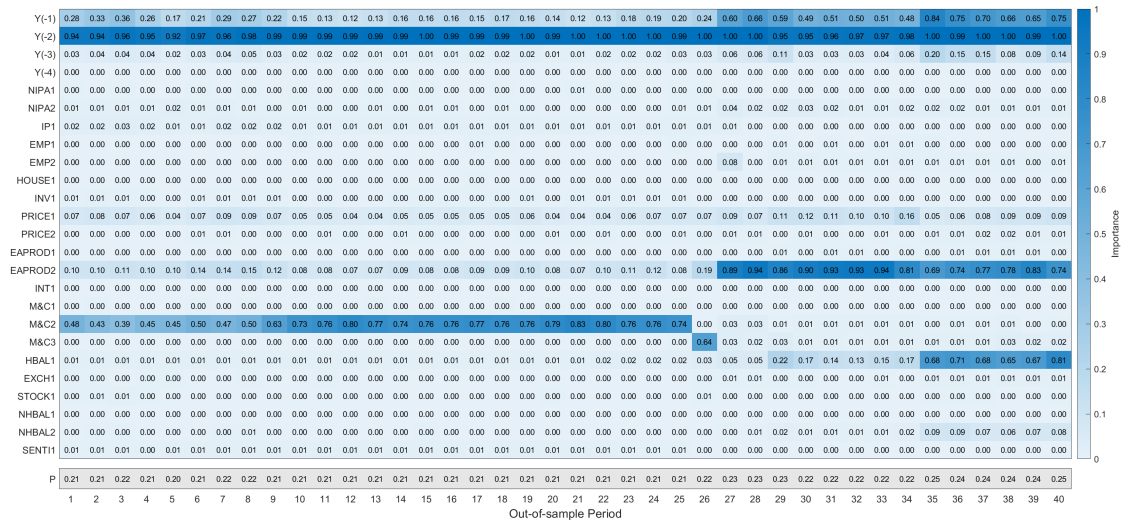
(a) Dirac Model Case



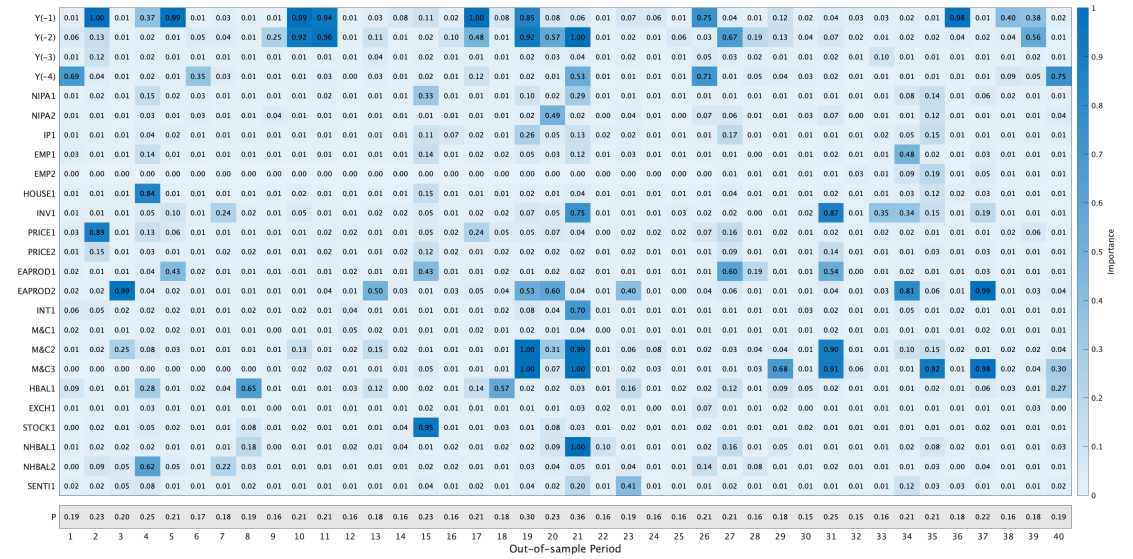
(b) B-DART Model Case

Figure H.4: Importance for Forecasting 6-quarter-ahead CPI Inflation Rate

Note: The figure above depicts the importance of each predictor variable for 40 out-of-sample periods when forecasting the 6-quarter-ahead CPI inflation rate. The heatmap becomes darker as the estimated importance of the predictor variable approaches 1. The last row is the estimated inclusion probability.



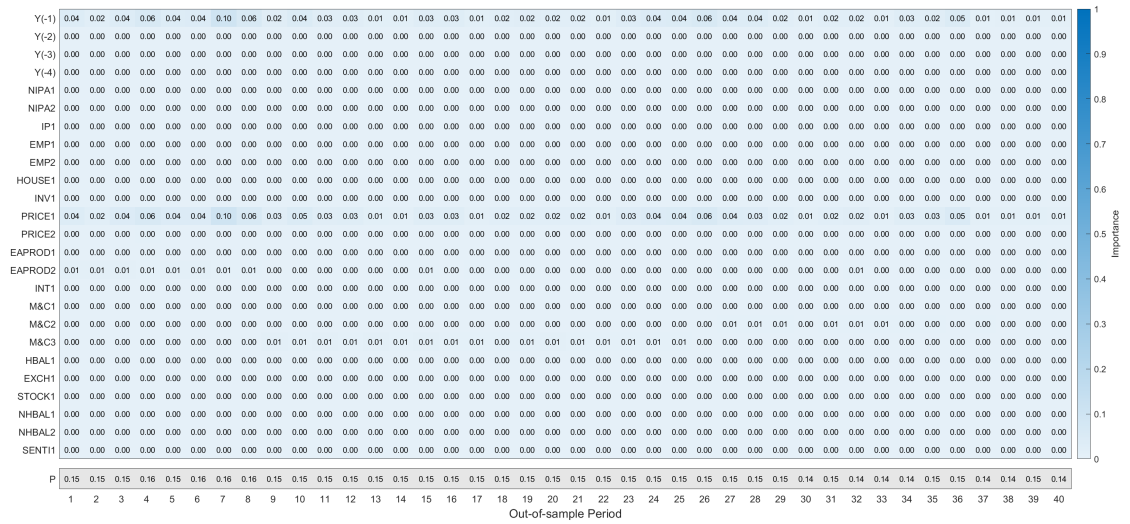
(a) Dirac Model Case



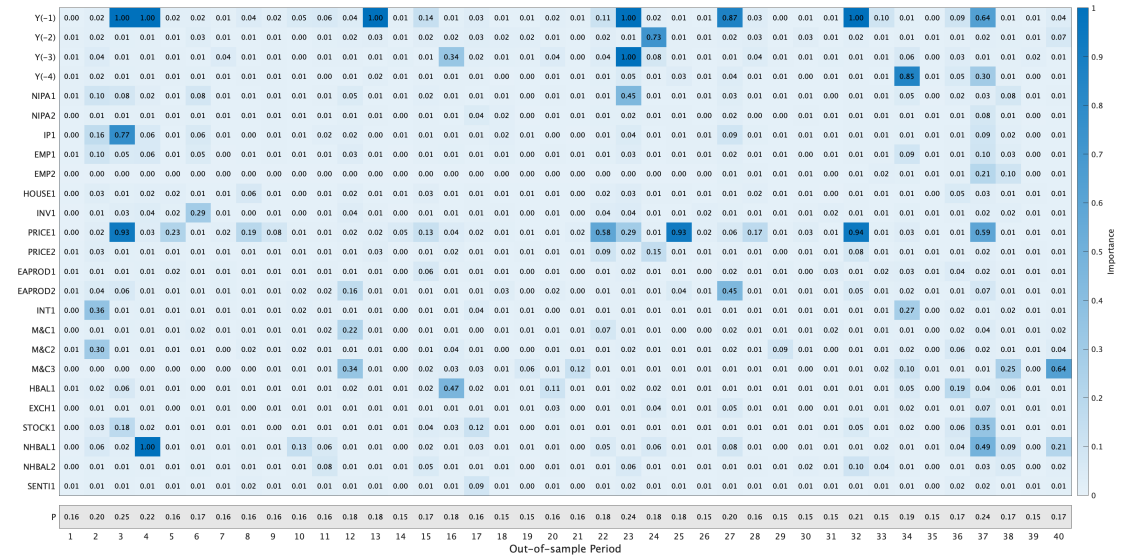
(b) B-DART Model Case

Figure H.5: Importance for Forecasting 1-quarter-ahead PPI Inflation Rate

Note: The figure above depicts the importance of each predictor variable for 40 out-of-sample periods when forecasting the 1-quarter-ahead PPI inflation rate. The heatmap becomes darker as the estimated importance of the predictor variable approaches 1. The last row is the estimated inclusion probability.



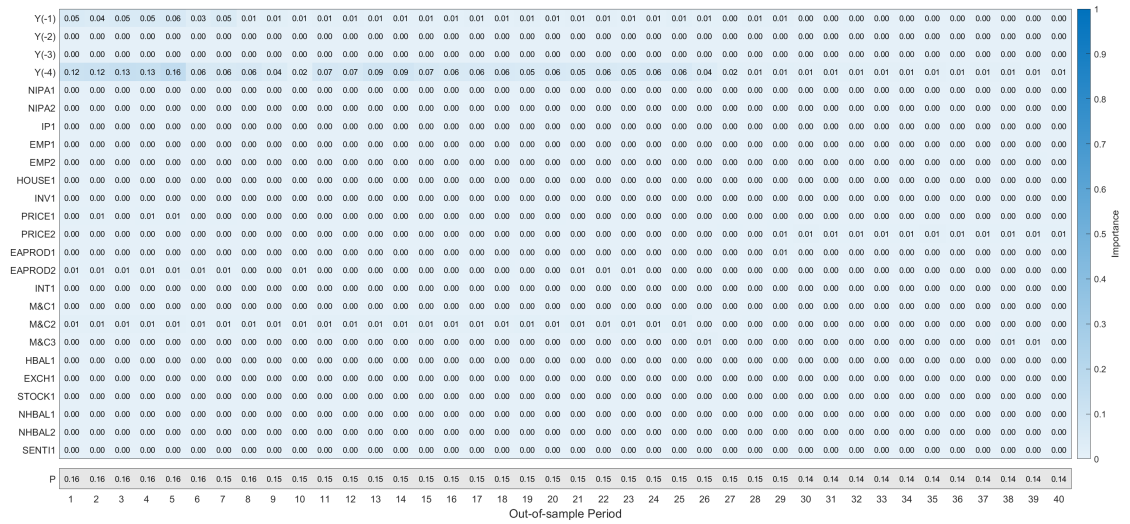
(a) Dirac Model Case



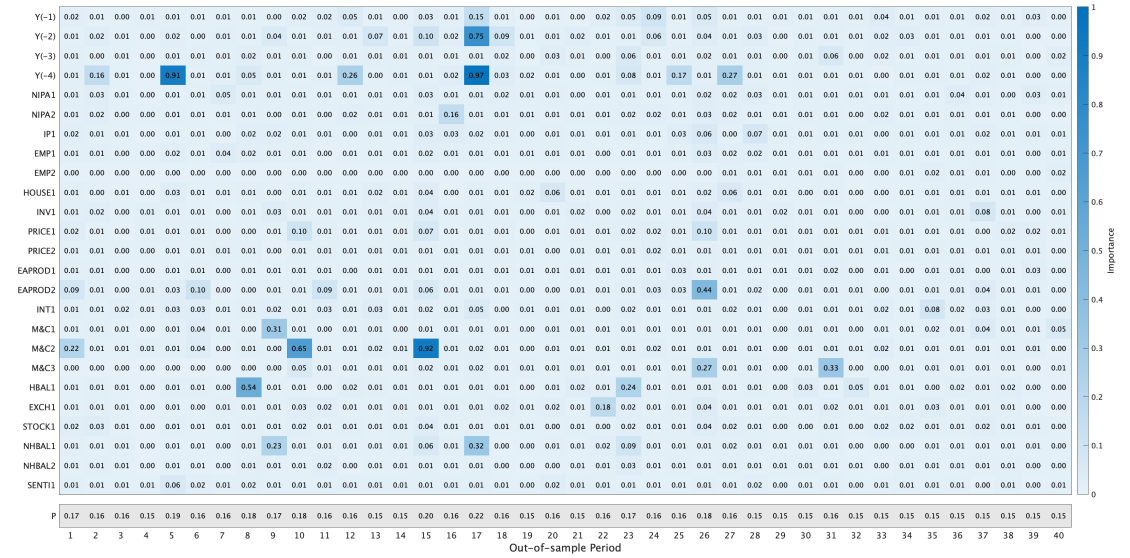
(b) B-DART Model Case

Figure H.6: Importance for Forecasting 4-quarter-ahead PPI Inflation Rate

Note: The figure above depicts the importance of each predictor variable for 40 out-of-sample periods when forecasting the 4-quarter-ahead PPI inflation rate. The heatmap becomes darker as the estimated importance of the predictor variable approaches 1. The last row is the estimated inclusion probability.



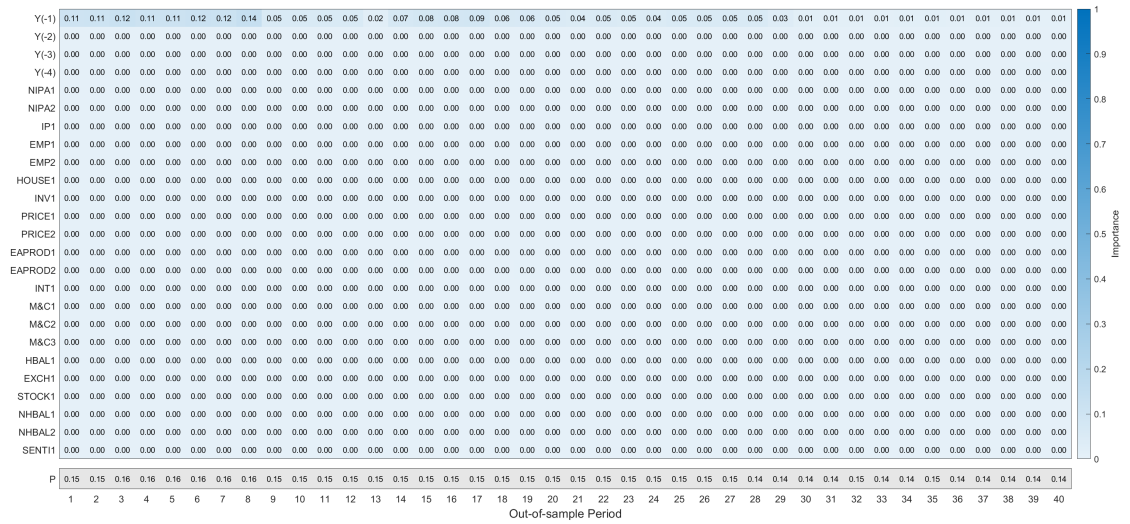
(a) Dirac Model Case



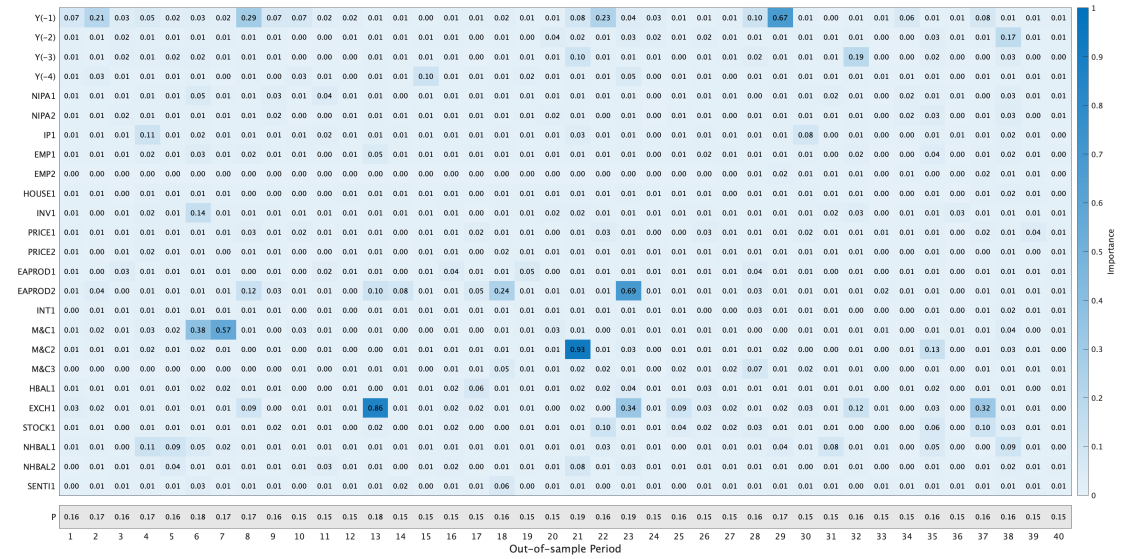
(b) B-DART Model Case

Figure H.7: Importance for Forecasting 2-quarter-ahead Crude Oil Price

Note: The figure above depicts the importance of each predictor variable for 40 out-of-sample periods when forecasting the 2-quarter-ahead Crude Oil Price. The heatmap becomes darker as the estimated importance of the predictor variable approaches 1. The last row is the estimated inclusion probability.



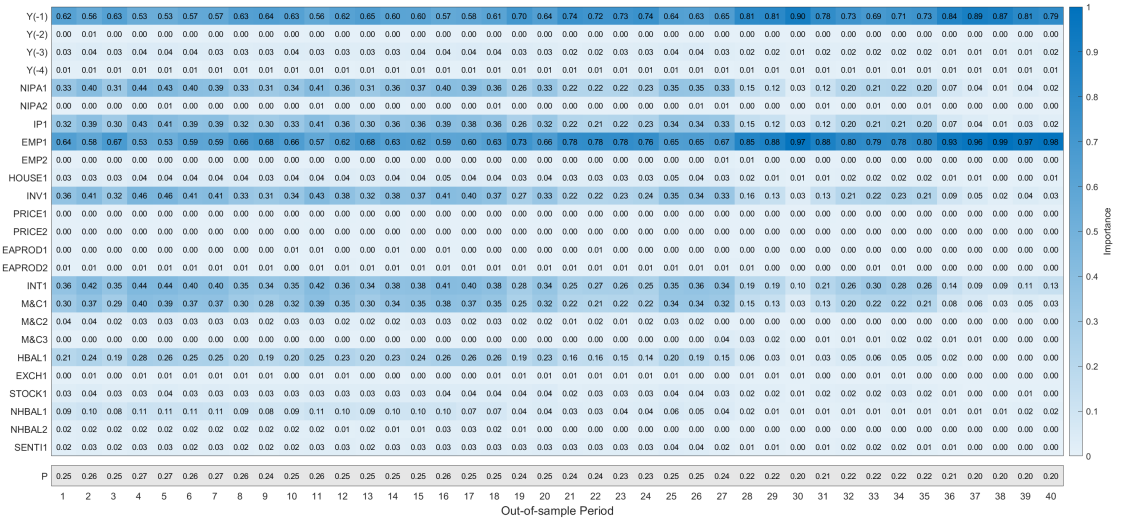
(a) Dirac Model Case



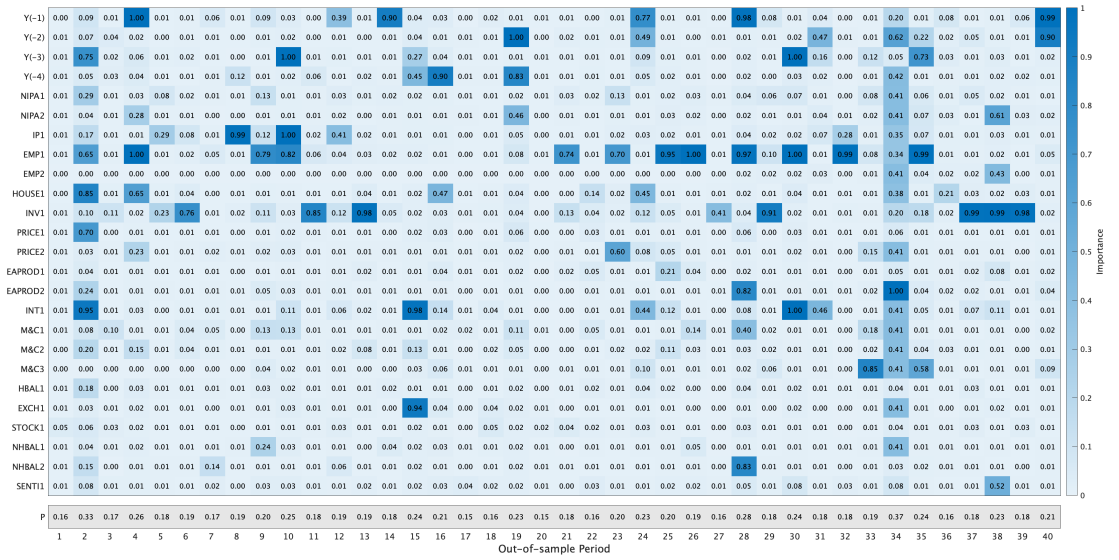
(b) B-DART Model Case

Figure H.8: Importance for Forecasting 5-quarter-ahead Crude Oil Price

Note: The figure above depicts the importance of each predictor variable for 40 out-of-sample periods when forecasting the 5-quarter-ahead Crude Oil Price. The heatmap becomes darker as the estimated importance of the predictor variable approaches 1. The last row is the estimated inclusion probability.



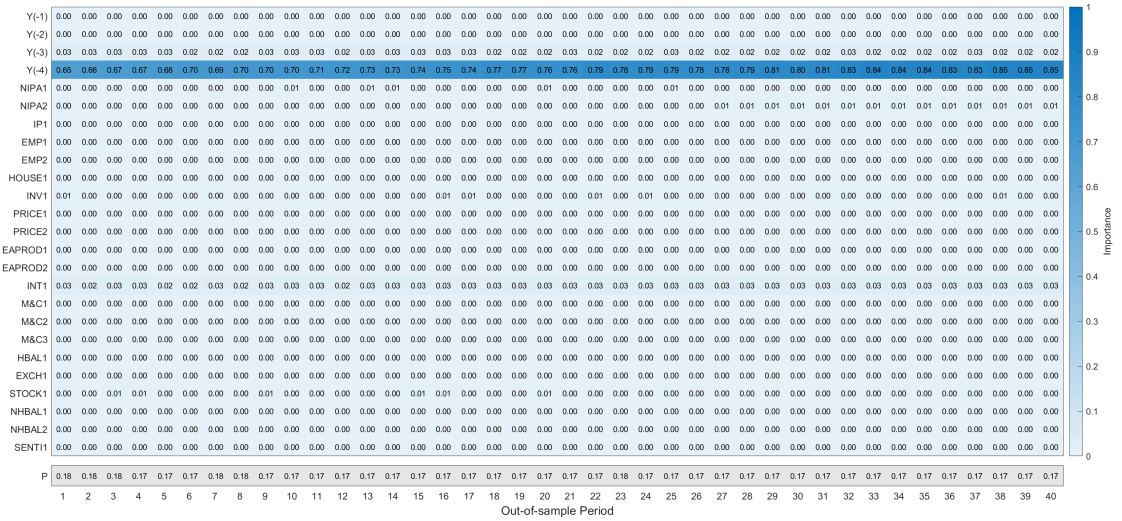
(a) Dirac Model Case



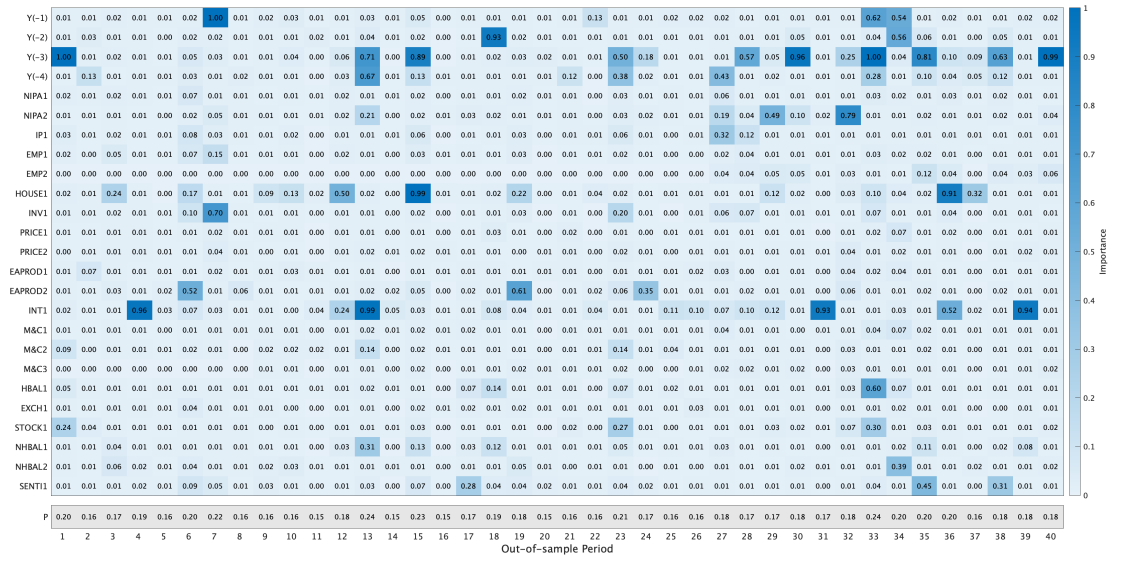
(b) B-DART Model Case

Figure H.9: Importance for Forecasting 2-quarter-ahead Federal Funds Effective Rate

Note: The figure above depicts the importance of each predictor variable for 40 out-of-sample periods when forecasting the 2-quarter-ahead Federal Funds Effective Rate. The heatmap becomes darker as the estimated importance of the predictor variable approaches 1. The last row is the estimated inclusion probability.



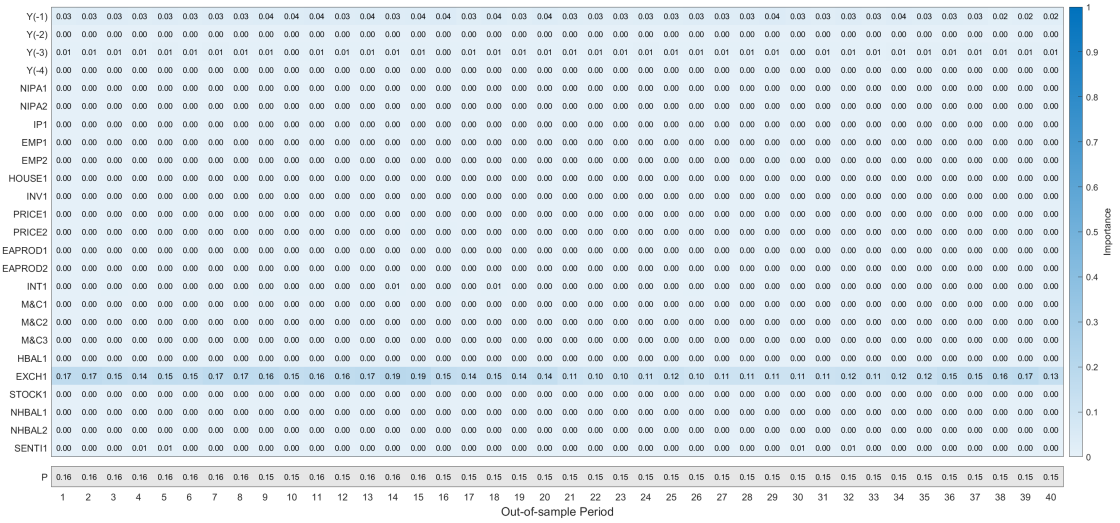
(a) Dirac Model Case



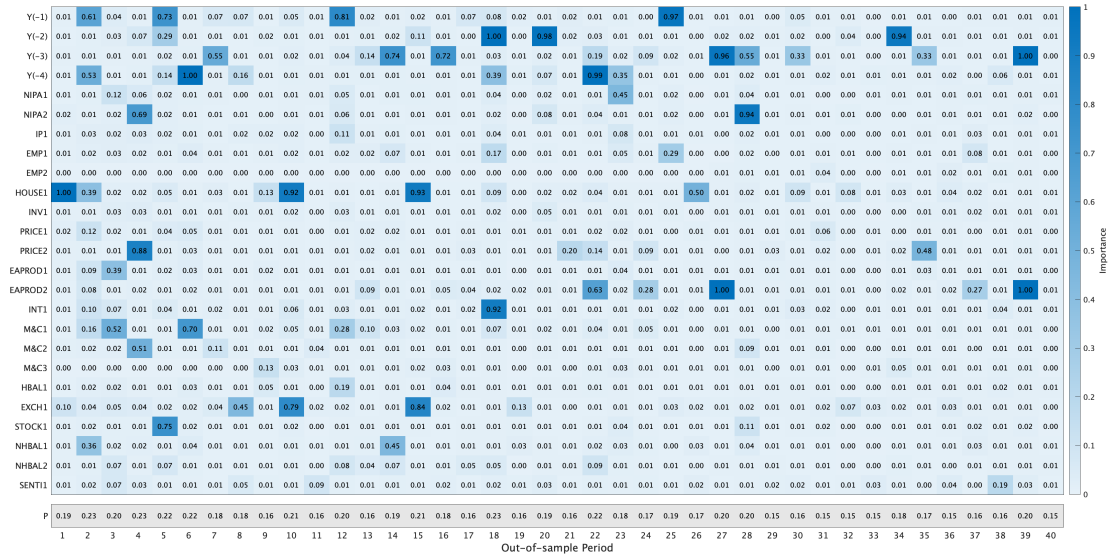
(b) B-DART Model Case

Figure H.10: Importance for Forecasting 8-quarter-ahead Federal Funds Effective Rate

Note: The figure above depicts the importance of each predictor variable for 40 out-of-sample periods when forecasting the 8-quarter-ahead Federal Funds Effective Rate. The heatmap becomes darker as the estimated importance of the predictor variable approaches 1. The last row is the estimated inclusion probability.



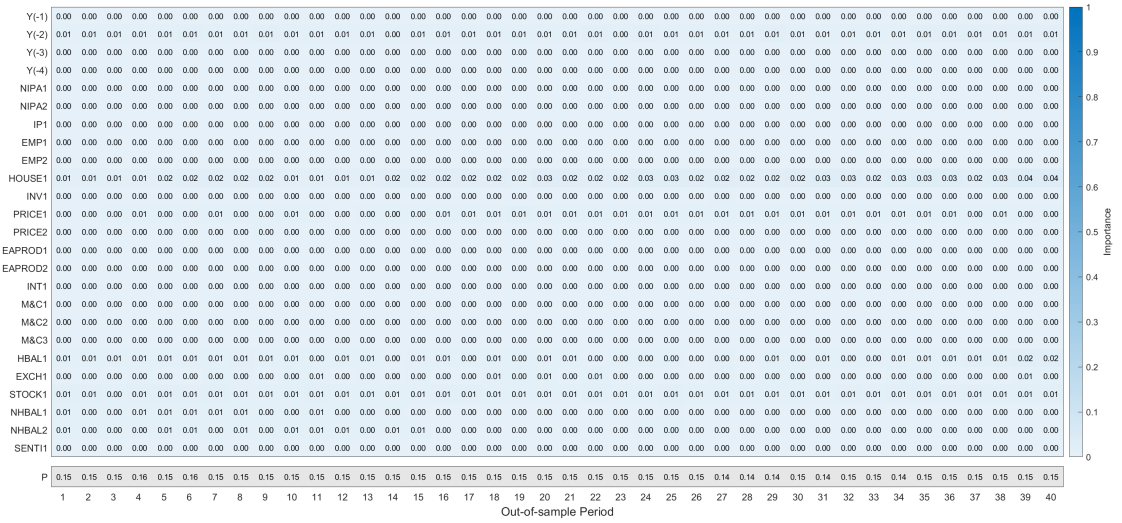
(a) Dirac Model Case



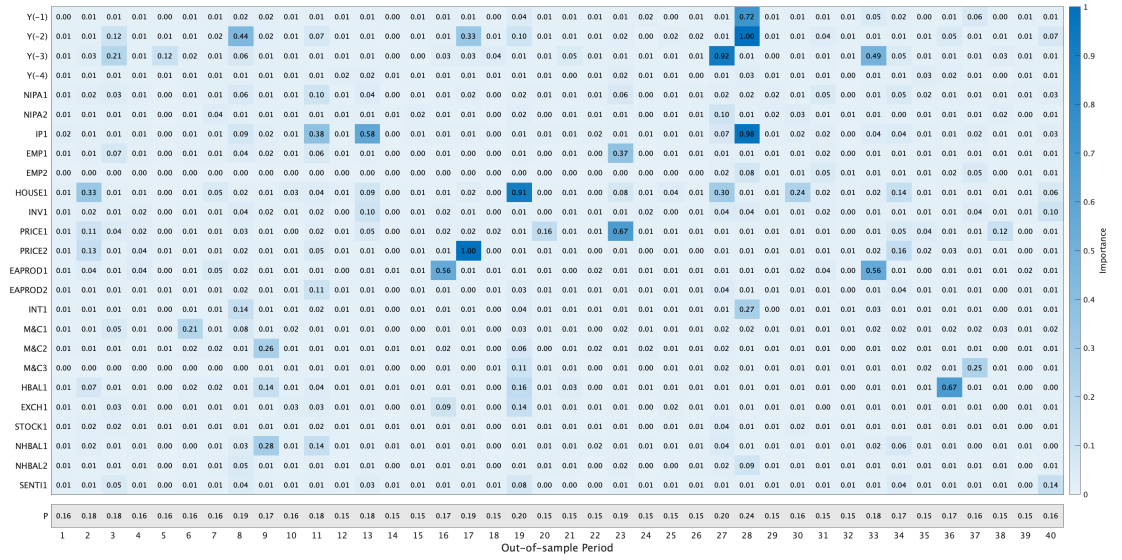
(b) B-DART Model Case

Figure H.11: Importance for Forecasting 5-quarter-ahead 10-year Maturity Interest Rate

Note: The figure above depicts the importance of each predictor variable for 40 out-of-sample periods when forecasting the 5-quarter-ahead 10-year Maturity Interest Rate. The heatmap becomes darker as the estimated importance of the predictor variable approaches 1. The last row is the estimated inclusion probability.



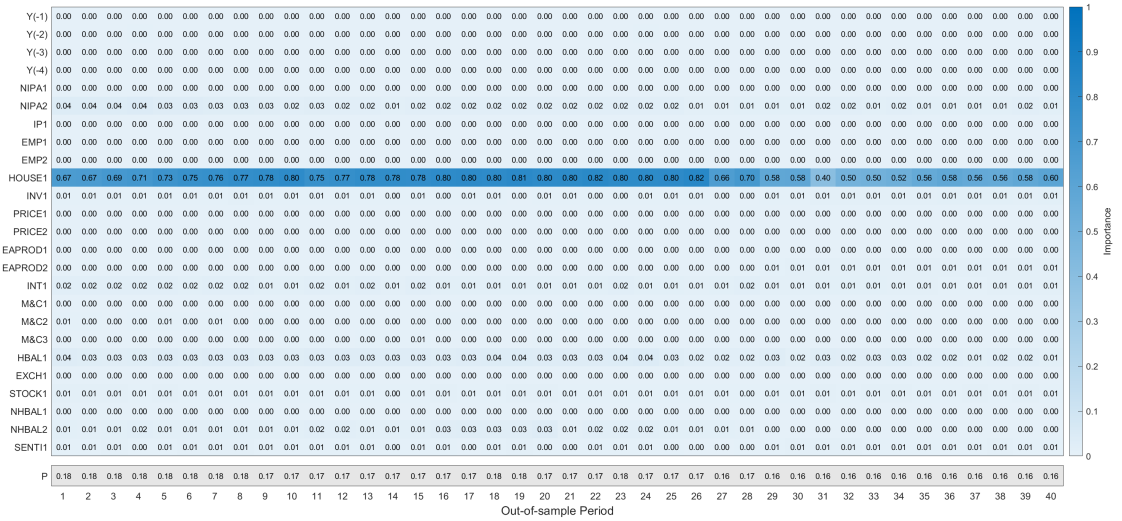
(a) Dirac Model Case



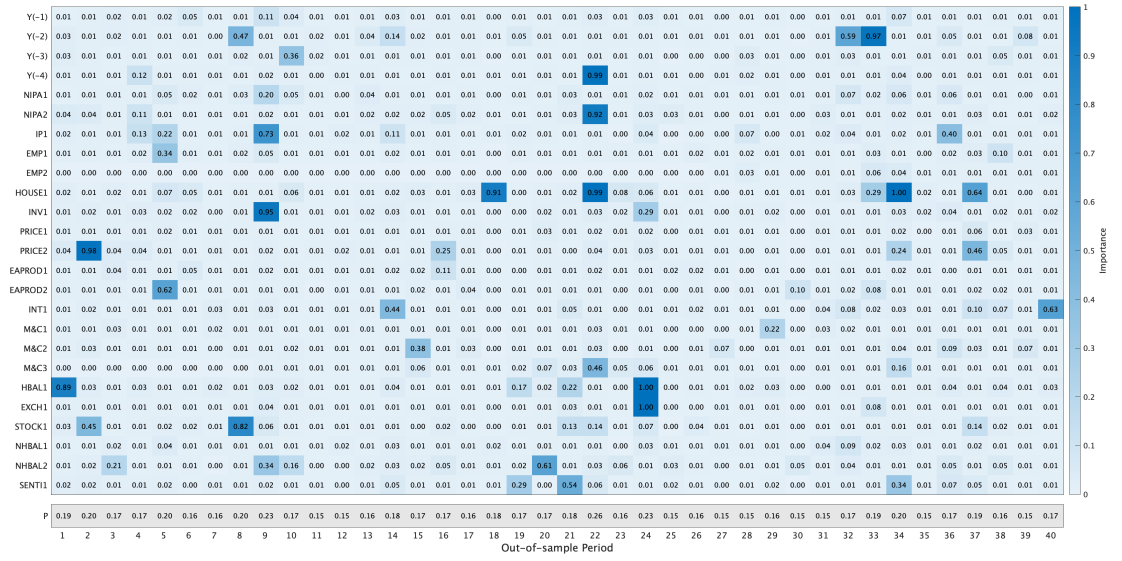
(b) B-DART Model Case

Figure H.12: Importance for Forecasting 6-quarter-ahead 10-year Maturity Interest Rate

Note: The figure above depicts the importance of each predictor variable for 40 out-of-sample periods when forecasting the 6-quarter-ahead 10-year Maturity Interest Rate. The heatmap becomes darker as the estimated importance of the predictor variable approaches 1. The last row is the estimated inclusion probability.



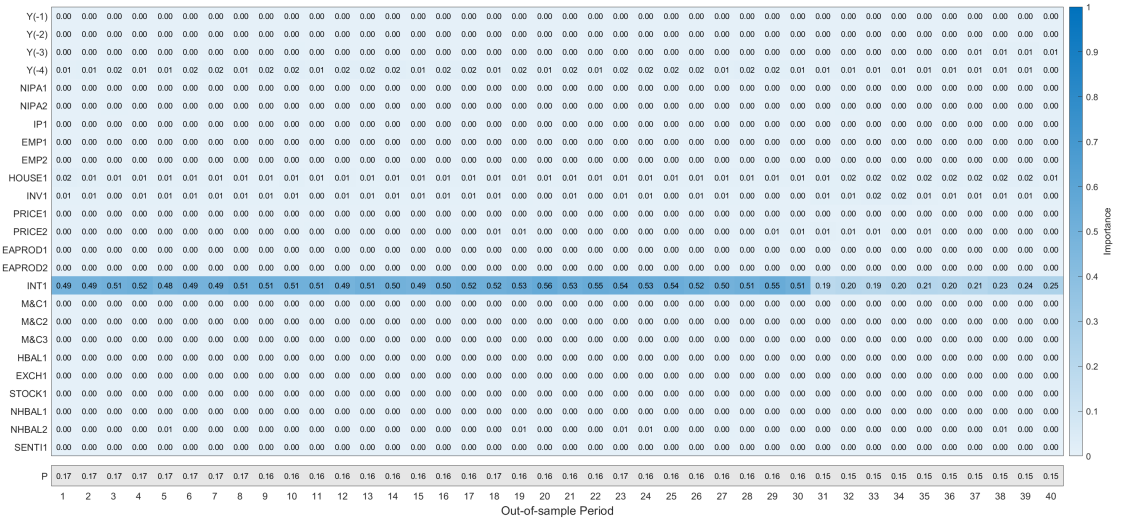
(a) Dirac Model Case



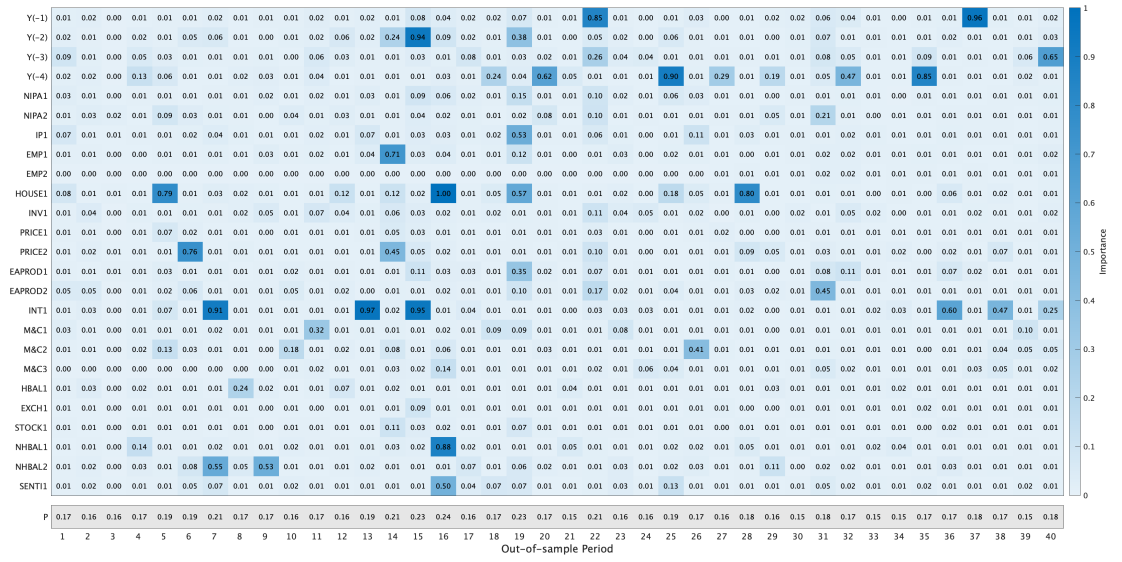
(b) B-DART Model Case

Figure H.13: Importance for Forecasting 3-quarter-ahead Real GDP Growth Rate

Note: The figure above depicts the importance of each predictor variable for 40 out-of-sample periods when forecasting the 3-quarter-ahead Real GDP Growth Rate. The heatmap becomes darker as the estimated importance of the predictor variable approaches 1. The last row is the estimated inclusion probability.



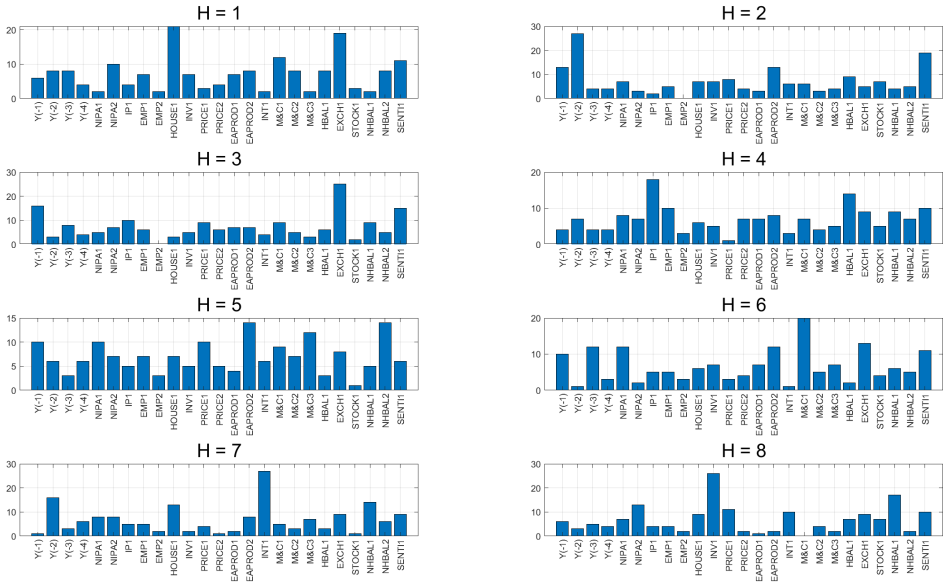
(a) Dirac Model Case



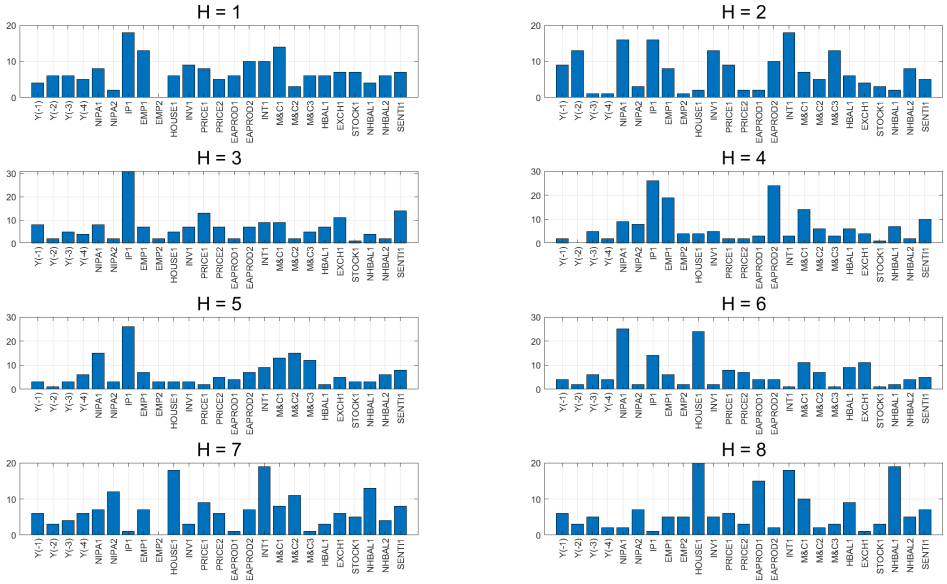
(b) B-DART Model Case

Figure H.14: Importance for Forecasting 5-quarter-ahead Real GDP Growth Rate

Note: The figure above depicts the importance of each predictor variable for 40 out-of-sample periods when forecasting the 5-quarter-ahead Real GDP Growth Rate. The heatmap becomes darker as the estimated importance of the predictor variable approaches 1. The last row is the estimated inclusion probability.



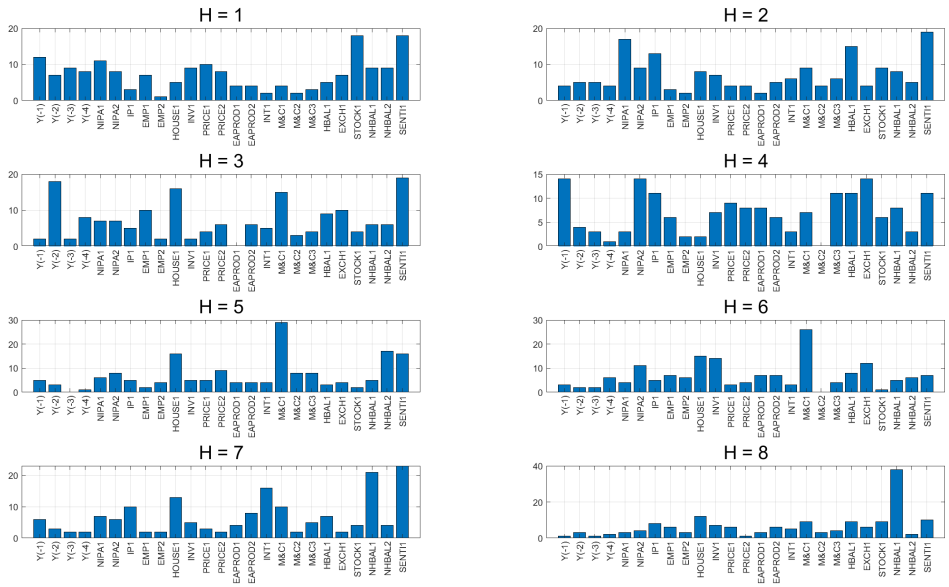
(a) PCE Inflation Case



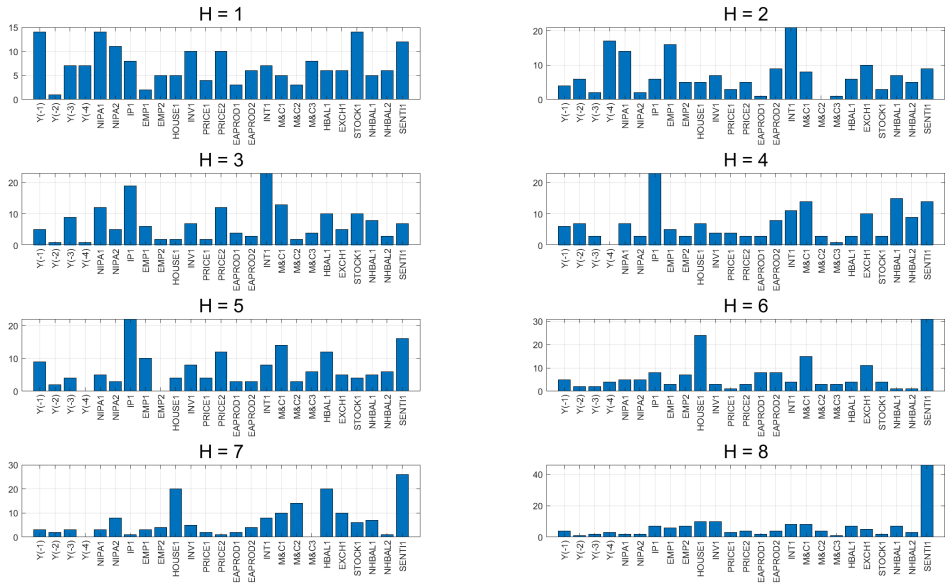
(b) CPI Inflation Case

Figure H.15: Frequency of Predictors used as the Splitting Criteria

Note: Frequency of predictors selected as splitting variables in the B-DART model. The upper and lower panel each corresponds to the PCE inflation and CPI inflation, and the frequencies are computed across all posterior draws and forecast horizons. Higher values indicate predictors that are more frequently used in tree splits.



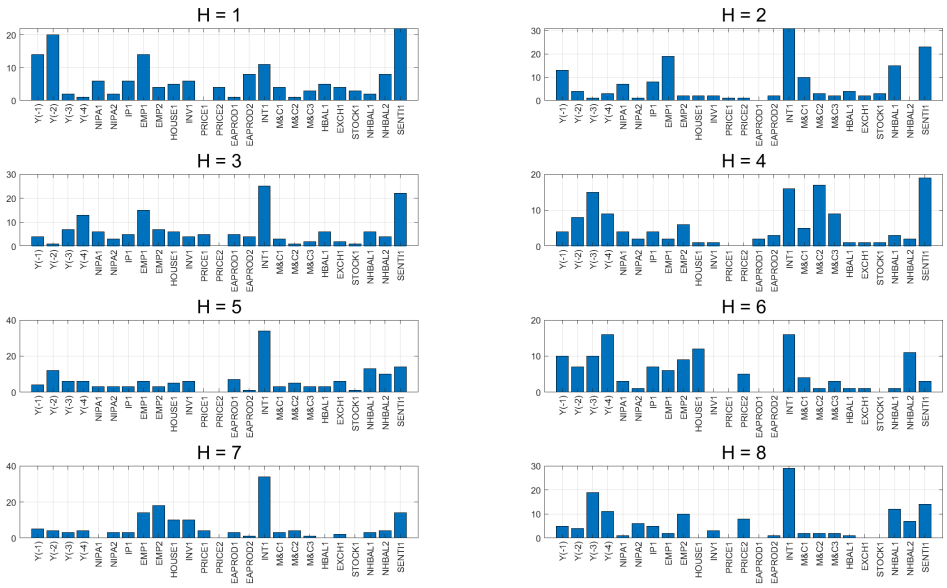
(a) PPI Inflation Case



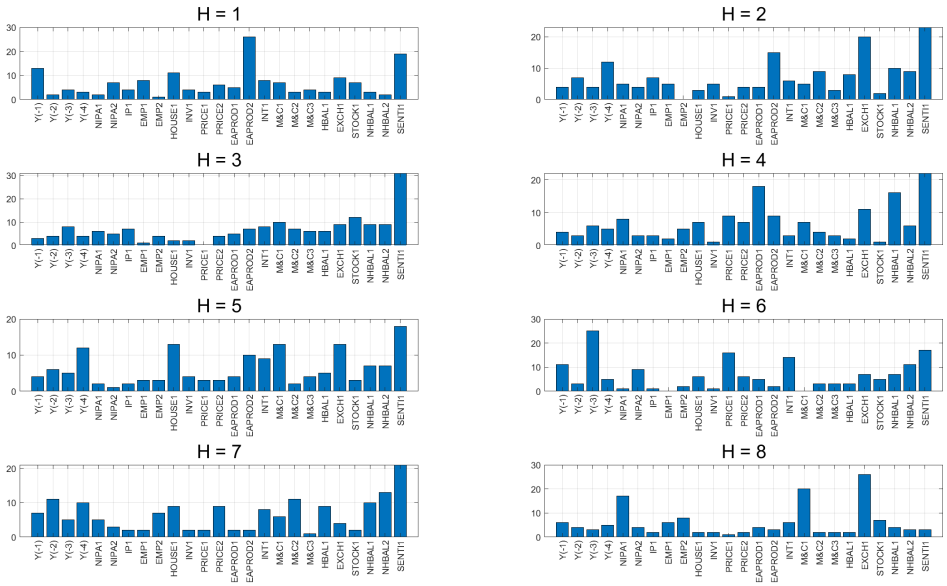
(b) Crude Oil Price Case

Figure H.16: Frequency of Predictors used as the Splitting Criteria

Note: Frequency of predictors selected as splitting variables in the B-DART model. The upper and lower panel each corresponds to the PPI inflation and crude oil price, and the frequencies are computed across all posterior draws and forecast horizons. Higher values indicate predictors that are more frequently used in tree splits.



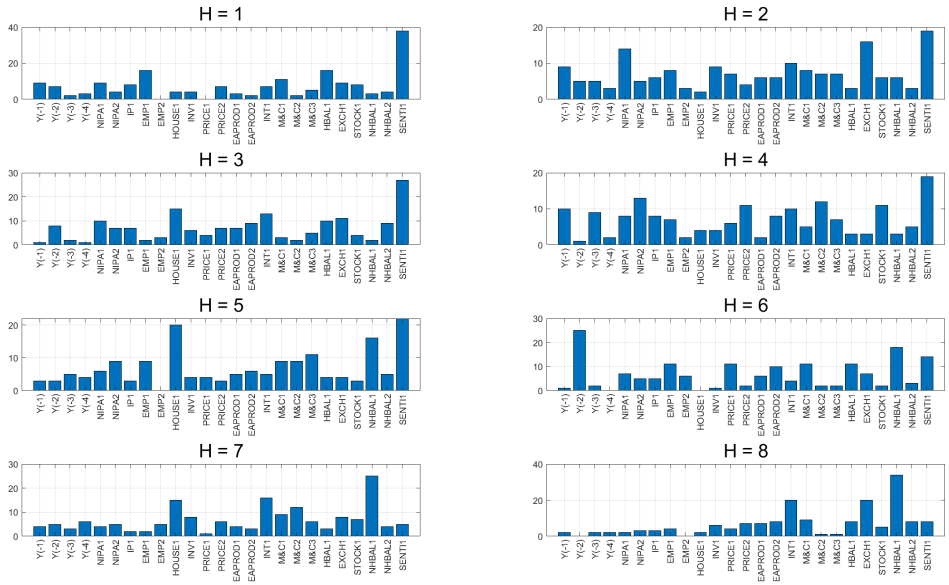
(a) Federal Funds Effective Rate Case



(b) 10-year Maturity Interest Rate Case

Figure H.17: Frequency of Predictors used as the Splitting Criteria

Note: Frequency of predictors selected as splitting variables in the B-DART model. The upper and lower panel each corresponds to the Federal Funds Effective Rate and 10-year Maturity Interest Rate, and the frequencies are computed across all posterior draws and forecast horizons. Higher values indicate predictors that are more frequently used in tree splits.



(a) Real GDP Growth Rate Case

Figure H.18: Frequency of Predictors used as the Splitting Criteria

Note: Frequency of predictors selected as splitting variables in the B-DART model. The upper and lower panel each corresponds to the real GDP growth rate, and the frequencies are computed across all posterior draws and forecast horizons. Higher values indicate predictors that are more frequently used in tree splits.