

Scalable Variational Bayes Inference for Dynamic Variable Selection

Nicolas Bianco, Mauro Bernardi & Daniele Bianchi

To cite this article: Nicolas Bianco, Mauro Bernardi & Daniele Bianchi (29 Apr 2026): Scalable Variational Bayes Inference for Dynamic Variable Selection, Journal of Computational and Graphical Statistics, DOI: [10.1080/10618600.2026.2634826](https://doi.org/10.1080/10618600.2026.2634826)

To link to this article: <https://doi.org/10.1080/10618600.2026.2634826>



© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 29 Apr 2026.



[Submit your article to this journal](#)



Article views: 942



[View related articles](#)



[View Crossmark data](#)

Scalable Variational Bayes Inference for Dynamic Variable Selection

Nicolas Bianco^a , Mauro Bernardi^b, and Daniele Bianchi^c 

^aScientific Computing Center, Karlsruhe Institute of Technology, Karlsruhe, Germany; ^bDepartment of Statistical Sciences, University of Padova, Padova, Italy; ^cSchool of Economics and Finance, Queen Mary University of London, London, UK

ABSTRACT

We develop a variational Bayes approach for dynamic variable selection in high-dimensional regression models with time-varying parameters and predictors that exhibit a predefined group structure. Through comprehensive simulation studies, we demonstrate that our method yields more accurate parameter estimates than existing Bayesian static and dynamic variable selection approaches while maintaining computational efficiency. We illustrate the performance of our approach within the context of a popular problem in economics: forecasting inflation using a large set of macroeconomic predictors. Our approach demonstrates significant improvements in out-of-sample point- and density-forecasting accuracy. A retrospective analysis of the time-varying parameter estimates reveals economically interpretable patterns in inflation dynamics.

ARTICLE HISTORY

Received April 2025
Accepted February 2026

KEYWORDS

Approximate inference;
Bayesian variable selection;
High-dimensional
regressions; Macroeconomic
forecasting; Mean-field
approximation

1. Introduction

Predicting the dynamics of economic variables is critical for policymakers and market participants. For instance, inflation forecasts guide central bankers' monetary and fiscal decisions and help investors hedge against inflation risk. However, the predictive power of individual variables is rarely known *ex ante*. Consequently, variable selection methods have grown in prominence, particularly in predictive regressions (e.g., Ročková and George 2014; Huber, Koop, and Onorante 2021). Although economic predictability is often transient, conventional regression models typically perform static variable selection, that is, the subsets of "active" predictors are assumed to be constant across the training sample.

To address this limitation, we develop a dynamic Bernoulli-Gaussian regression framework that treats variable selection as a latent stochastic process. Our approach specifically addresses the statistical challenge of dynamically incorporating predictors into a time-varying parameter regression model, admitting variables only during periods in which they demonstrate significant predictive power. We propose a variational Bayes estimation method that enables a computationally efficient and statistically consistent identification of relevant predictors in high-dimensional, time-varying parameter regression settings.

Our methodological contribution is threefold. First, the optimal variational density for the latent selection process accommodates flexible smoothness assumptions, regularizing the dynamic inclusion probabilities by filtering noisy estimates that could lead to unstable model identification. Second, we sequentially eliminate predictors that provide no information across the entire training sample, thereby significantly enhancing both estimation accuracy and computational efficiency.

Third, our framework incorporates prior information on group correlations among predictors, enabling decision-makers to leverage domain knowledge of structural relationships among variables for individual variable selection.

We first assess the estimation accuracy of our approach through an extensive simulation study where time-varying parameters exhibit diverse temporal sparsity patterns. Our evaluation proceeds in two stages: First, we benchmark our approach against recent dynamic variable selection methods, including the dynamic spike-and-slab approaches of Koop and Korobilis (2023) and Ročková and McAlinn (2021), as well as static two-component mixture priors (e.g., Ročková and George 2014). Second, we compare our variational Bayes method against an equivalent MCMC algorithm to establish baseline performance differences.

The simulation results demonstrate that our variational Bayes method successfully balances accuracy and computational efficiency. While conventional MCMC becomes computationally prohibitive even for moderate dimensions, our approach scales effectively to high-dimensional settings without compromising estimation accuracy. This computational advantage proves particularly crucial for real-time economic forecasting, where recursive predictions require frequent model updates. Moreover, our method maintains efficiency comparable to competing approximate inference algorithms while achieving significantly superior accuracy across all model dimensions.

The central intuition of our paper is that the precise identification of time-varying relevant predictors is crucial for decision-making in data-rich environments. To empirically validate this claim, we examine a fundamental problem in applied macroeconomics: inflation forecasting using a broad

CONTACT Daniele Bianchi  d.bianchi@qmul.ac.uk  School of Economics and Finance, Queen Mary University of London, London, UK.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/10618600.2026.2634826>.

© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

set of macroeconomic indicators (e.g., Stock and Watson 2006). Our analysis employs a comprehensive dataset of 220+ quarterly macroeconomic variables from 1967:Q3 to 2022:Q2, sourced from FRED-QD and transformed following McCracken and Ng (2020). We evaluate forecasting performance for four key inflation measures—total CPI, core CPI, GDP deflator, and PCE deflator—across multiple horizons.

Our empirical results reaffirm that parsimonious specifications, particularly the local-level model of Stock and Watson (2007), remain strong benchmarks in inflation forecasting. However, our dynamic Bernoulli-Gaussian predictive regression demonstrates consistent and statistically significant outperformance across all competing methods incorporating macroeconomic predictors. This superior performance holds for both point and density forecasts and proves robust across different forecasting horizons.

Unlike conventional time-varying intercept specifications à la Stock and Watson (2007), our framework enables retrospective analysis of evolving relationships between macroeconomic variables and inflation measures. Analysis of posterior dynamic inclusion probabilities yields three principal findings: First, the set of predictors with meaningful explanatory power is strikingly sparse. Second, the composition of relevant predictors shows substantial time variation. Third, economic theories receive nuanced empirical support, with their validity contingent on specific periods.

This paper contributes to the growing literature on Bayesian variable selection in high dimensions (e.g., Ročková and George 2014; Huber, Koop, and Onorante 2021; Ray and Szabó 2022; Mogliani and Simoni 2024). Our work extends these approaches to dynamic settings, complementing several existing frameworks, such as the dynamic variable selection approaches in Koop and Korobilis (2023); Uribe and Lopes (2025); Ročková and McAlinn (2021), and (iii) dynamic shrinkage methods (e.g., Kowal, Matteson, and Ruppert 2019).

2. Model Specification and Inference

Let y_t a scalar response at time t and $\mathbf{x}_{t-1} = (x_{1,t-1}, \dots, x_{p,t-1})'$ a p -dimensional vector of known predictors. A predictive regression model can be specified as

$$y_t = \sum_{j=1}^p \beta_{j,t} x_{j,t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, e^{h_t}), \quad t = 1, \dots, n, \quad (2.1)$$

where $\boldsymbol{\beta}_t = (\beta_{1,t}, \dots, \beta_{p,t})'$ is a vector of regression coefficients and $h_t = \log \sigma_t^2$ is a latent log-volatility process. We assume that predictors can dynamically enter or leave the regression model as time progresses according to

$$\beta_{j,t} = b_{j,t} \gamma_{j,t}, \quad \text{where} \quad b_{j,t} = b_{j,t-1} + v_{j,t} \\ v_{j,t} \sim \mathcal{N}(0, \eta_j^2), \quad (2.2)$$

with $b_{j,0} \sim \mathcal{N}(0, k_0 \eta_j^2)$ the initial state, and $\gamma_{j,t} \in \{0, 1\}$ an indicator variable which identifies if the j -th predictor is included or not in the model at time t .

Leveraging the Markov property of the random walk, the vector $\mathbf{b}_j = (b_{j,0}, \dots, b_{j,n})'$ for $j = 1, \dots, p$ is a Gaussian Markov Random Field (GMRF) $\mathbf{b}_j \sim \mathcal{N}_{n+1}(\mathbf{0}, \eta_j^2 \mathbf{Q}^{-1})$. $\mathbf{Q}$ is a tridiagonal precision matrix with diagonal elements $q_{1,1} = 1 + 1/k_0$, $q_{n+1,n+1} = 1$, and $q_{l,l} = 2$ for $l = 2, \dots, n$, and off-diagonal ones $q_{l,m} = -1$ if $|l - m| = 1$ and 0 elsewhere. The same logic applies to the log-volatility process $\mathbf{h} = (h_0, \dots, h_n)'$ with initial state $h_0 \sim \mathcal{N}(0, k_0 v^2)$, such that $h_t = h_{t-1} + e_t$ with $e_t \sim \mathcal{N}(0, v^2)$. Hence, $\mathbf{h} \sim \mathcal{N}_{n+1}(\mathbf{0}, v^2 \mathbf{Q}^{-1})$.

Equation (2.2) implies that the time-varying regression coefficients $\{\beta_{j,t}\}_{t=1}^n$ are the product of the latent stochastic process $\{b_{j,t}\}_{t=1}^n$ and the indicator $\{\gamma_{j,t}\}_{t=1}^n$. We assume a hierarchical specification for $\gamma_{j,t}$ of the form $\gamma_{j,t} | \omega_{j,t} \sim \text{Bern}(\text{expit}(\omega_{j,t}))$ for $j = 1, \dots, p$, where $\text{expit}(\cdot)$ is the inverse of the logit function. The latent $\boldsymbol{\omega}_j = (\omega_{j,0}, \dots, \omega_{j,n})'$ is a GMRF $\boldsymbol{\omega}_j \sim \mathcal{N}_{n+1}(\mathbf{0}, \xi_j^2 \mathbf{Q}^{-1})$ with $\omega_{j,0} \sim \mathcal{N}(0, k_0 \xi_j^2)$, and it drives the persistence of the inclusion probability $\mathbb{P}(\gamma_{j,t} = 1)$.¹ To keep the prior distribution of the initial states uninformative, we fix the scaling parameter $k_0 \gg 0$. We also assume that k_0 is the same for all GMRFs. This assumption can be relaxed if there is strong prior belief in heterogeneous initial conditions for the latent states.

Finally, we assume inverse-gamma priors for the variances $v^2 \sim \text{IG}(A_v, B_v)$, $\eta_j^2 \sim \text{IG}(A_\eta, B_\eta)$, and $\xi_j^2 \sim \text{IG}(A_\xi, B_\xi)$, where $A_v, B_v, A_\eta, B_\eta, A_\xi, B_\xi$ are positive fixed hyperparameters.

2.1. Variational Bayes Inference

Variational inference requires to minimize a Kullback-Leibler divergence (KL) between the approximating $q(\boldsymbol{\theta})$ and the true posterior $p(\boldsymbol{\theta} | \mathbf{y})$ densities (e.g., Blei, Kucukelbir, and McAuliffe 2017). Ormerod and Wand (2010) show that the problem of minimizing this KL divergence can be equivalently stated as maximizing the variational lower bound (ELBO) denoted by $\underline{p}(\mathbf{y}; q)$:

$$q^*(\boldsymbol{\theta}) = \arg \max_{q(\boldsymbol{\theta}) \in \mathcal{Q}} \log \underline{p}(\mathbf{y}; q) \quad \text{with} \\ \underline{p}(\mathbf{y}; q) = \int q(\boldsymbol{\theta}) \log \left\{ \frac{p(\mathbf{y}, \boldsymbol{\theta})}{q(\boldsymbol{\theta})} \right\} d\boldsymbol{\theta} \quad (2.4)$$

where $q^*(\boldsymbol{\theta}) \in \mathcal{Q}$ represents the optimal variational density and $\mathcal{Q}$ is a space of functions. The choice of the family of distributions $\mathcal{Q}$ is critical and leads to different algorithmic approaches. We consider a Mean-Field Variational Bayes (MFVB) approach which is based on a factorization of the variational density $q(\boldsymbol{\theta}) = \prod_{l=1}^L q_l(\boldsymbol{\theta}_l)$ for a partition $\{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_L\}$ of the parameter vector $\boldsymbol{\theta}$. This factorization implies that the optimal variational density of each component $q(\boldsymbol{\theta}_l)$ can be derived in closed

¹The marginal distribution for $\boldsymbol{\gamma}_j = (\gamma_{j,1}, \dots, \gamma_{j,n})'$ can be retrieved by integrating out $\boldsymbol{\omega}_j$ as,

$$p(\gamma_{j,1}, \dots, \gamma_{j,n}) = \int p(\boldsymbol{\omega}_j) \prod_{t=1}^n p(\gamma_{j,t} | \omega_{j,t}) d\boldsymbol{\omega}_j, \quad (2.3)$$

so that $\gamma_{j,1}, \dots, \gamma_{j,n}$ represent a set of autocorrelated latent states for each $j = 1, \dots, p$.

form as

$$q^*(\boldsymbol{\vartheta}_l) \propto \exp \left\{ \mathbb{E}_{q(\boldsymbol{\vartheta} \setminus \boldsymbol{\vartheta}_l)} \left[\log p(\mathbf{y}, \boldsymbol{\vartheta}) \right] \right\} \quad \text{with} \\ q(\boldsymbol{\vartheta} \setminus \boldsymbol{\vartheta}_l) = \prod_{\substack{j=1 \\ j \neq l}}^L q_l(\boldsymbol{\vartheta}_l) \quad (2.5)$$

where the expectation is taken with respect to the joint approximating density with the l -th element of the partition removed $q(\boldsymbol{\vartheta} \setminus \boldsymbol{\vartheta}_l)$. As a result, the joint density $q^*(\boldsymbol{\vartheta})$ can then be estimated based on a coordinate ascent variational inference (CAVI) algorithm.

Equation (2.5) shows that the factorization of $q(\boldsymbol{\vartheta})$ is crucial. Let $\boldsymbol{\vartheta} = (\mathbf{h}', \mathbf{b}', \boldsymbol{\gamma}', \boldsymbol{\omega}', v^2, \boldsymbol{\eta}^{2'}, \boldsymbol{\xi}^{2'})'$ be the vector that collects the model parameters. We factorize $p(\mathbf{y}, \boldsymbol{\vartheta})$ as:

$$p(\mathbf{y}, \boldsymbol{\vartheta}) = p(\mathbf{y}|\boldsymbol{\vartheta})p(\mathbf{h}|v^2)p(v^2) \prod_{j=1}^p p(\mathbf{b}_j|\eta_j^2)p(\boldsymbol{\omega}_j|\xi_j^2) \\ p(\eta_j^2)p(\xi_j^2) \underbrace{\prod_{t=1}^n p(\gamma_{j,t}|\omega_{j,t})}_{p(\boldsymbol{\gamma}_j|\boldsymbol{\omega}_j)}. \quad (2.6)$$

Two comments are in order. First, we exploit a Polya-Gamma data augmentation to allow for computationally tractable updates of $\boldsymbol{\omega}_j$, as it leads to a known form of the full conditionals (see Polson, Scott, and Windle 2013, and the discussion in Section A.1 in the supplementary for more details). Second, the optimal variational density of the time-varying coefficients $q(\mathbf{b})$ presents a computational bottleneck in high dimensions; assuming that all predictors are correlated requires inversion of a $p(n+1) \times p(n+1)$ matrix. Although this matrix is block-tridiagonal and can be handled efficiently by standard algebra, for large p the computation remains prohibitive, especially for the computing infrastructure typically available to economists and policymakers.

To address this issue, we follow Mogliani and Simoni (2024) and introduce the assumption that predictors may present a known group structure based on shared economic fundamentals, such that the joint density $q(\mathbf{b})$ can be factorized as $q(\mathbf{b}) = \prod_{k=1}^K q(\mathbf{b}_k)$ with K groups. This specification covers different cases: $K = 1$ denotes the most general case, in which all regression coefficients are correlated, whereas $K = p$ implies independence among predictors. By leveraging Polya-Gamma augmentation and the a-priori group correlation structure of the predictors, we can factorize $q(\boldsymbol{\vartheta})$ as,

$$q(\boldsymbol{\vartheta}) = q(\mathbf{h})q(v^2) \prod_{k=1}^K q(\mathbf{b}_k) \prod_{j=1}^p q(\boldsymbol{\omega}_j)q(\eta_j^2)q(\xi_j^2) \prod_{t=1}^n q(\gamma_{j,t})q(z_{j,t}), \quad (2.7)$$

where a characterization of the joint distributions $q(\mathbf{h})$, $q(\mathbf{b}_k)$, and $q(\boldsymbol{\omega}_j)$ is required to provide a global approximation for the latent states.

2.1.1. Optimal Variational Densities

We now discuss the optimal variational densities of $\mathbf{b}_k$, the variable selection indicators $\gamma_{j,t}$, and the stochastic volatility $\mathbf{h}$. A detailed description of the remaining components in (2.7) and the full set of analytical derivations and proofs are available in Section B in the supplementary material. In what follows, we define as $\boldsymbol{\mu}_{q(\mathbf{a})} = \mathbb{E}_q[\mathbf{a}]$ and $\boldsymbol{\Sigma}_{q(\mathbf{a})} = \text{Var}_q[\mathbf{a}]$ (or $\sigma_{q(a)}^2 = \text{Var}_q[a]$ if a is univariate) the expected value and variance of the random variable a , with respect to density q , respectively, and with $\mathbf{a}^{-j}$ the vector $\mathbf{a}$ without its j -th element. Proposition 1 provides the optimal variational density $q^*(\gamma_{j,t})$ for all variables $j = 1, \dots, p_k$ that belong to group $k = 1, \dots, K$ of size p_k .

Proposition 1. *The optimal variational density of $\gamma_{j,t}$ is a Bernoulli $q^*(\gamma_{j,t}) = \text{Bern}(\text{expit}(\omega_{q(\gamma_{j,t})}))$, where $\text{expit}(\cdot)$ is the inverse of the logit function and*

$$\omega_{q(\gamma_{j,t})} = \boldsymbol{\mu}_{q(\omega_{j,t})} - \frac{1}{2} \boldsymbol{\mu}_{q(1/\sigma_t^2)} (\mathbf{x}_{j,t-1}' \mathbb{E}_q[\mathbf{b}_{j,t}^2] \\ - 2 \boldsymbol{\mu}_{q(\mathbf{b}_{j,t})} \mathbf{x}_{j,t-1}' \boldsymbol{\mu}_{q(\varepsilon_{-j,t})}) \\ - \boldsymbol{\mu}_{q(1/\sigma_t^2)} \mathbf{x}_{j,t-1}' \mathbf{x}_{kt-1}^{-j} \text{diag}(\boldsymbol{\mu}_{q(\gamma_{k,t})}^{-j}) [\boldsymbol{\Sigma}_{q(\mathbf{b}_{k,t})}]_{-j,j}, \quad (2.8)$$

with $[\boldsymbol{\Sigma}_{q(\mathbf{b}_{k,t})}]_{-j,j}$ the j -th column without the j -th element of the t -th diagonal block of $\boldsymbol{\Sigma}_{q(\mathbf{b}_k)}$, and $\boldsymbol{\mu}_{q(\varepsilon_{-j,t})} = \mathbf{y}_t - \mathbf{x}_{k,t-1}' \text{diag}(\boldsymbol{\mu}_{q(\gamma_{k,t})}^{-j}) \boldsymbol{\mu}_{q(\mathbf{b}_{k,t})}^{-j} - \sum_{m=1, m \neq k}^K \mathbf{x}_{m,t-1}' \text{diag}(\boldsymbol{\mu}_{q(\gamma_{m,t})}) \boldsymbol{\mu}_{q(\mathbf{b}_{m,t})}$ with $\boldsymbol{\mu}_{q(\gamma_{k,t})}$ being the collection of $\boldsymbol{\mu}_{q(\gamma_{j,t})}$ for the predictors in group k .

We note that $\omega_{q(\gamma_{j,t})}$ depends on the covariance $\boldsymbol{\Sigma}_{q(\mathbf{b}_k)}$. As a result, the probability that a given predictor is selected depends on its correlation with the other predictors in the same group. If predictors are independent, the second line of (2.8) would equal zero. The parameter $\boldsymbol{\mu}_{q(1/\sigma_t^2)}$ is defined in Remark 1 in the supplementary material. Proposition 2 shows that the mean and variance of the time-varying parameters $\mathbf{b}_k$ depend on the full trajectory $t = 1, \dots, n$ of the estimates $\boldsymbol{\mu}_{q(\gamma_{j,t})}$ for all $j = 1, \dots, p_k$ in group k .

Proposition 2. *The optimal variational density for $\mathbf{b}_k = (\mathbf{b}_{k,0}, \dots, \mathbf{b}_{k,n})'$, with $\mathbf{b}_{k,t} = (b_{1,t}, \dots, b_{p_k,t})'$, is a multivariate Gaussian $q^*(\mathbf{b}_k) = \mathcal{N}_{p_k(n+1)}(\boldsymbol{\mu}_{q(\mathbf{b}_k)}, \boldsymbol{\Sigma}_{q(\mathbf{b}_k)})$:*

$$\boldsymbol{\Sigma}_{q(\mathbf{b}_k)} = (\mathbf{D}_k + \mathbf{Q}_k)^{-1}, \quad \boldsymbol{\mu}_{q(\mathbf{b}_k)} = \boldsymbol{\Sigma}_{q(\mathbf{b}_k)} \boldsymbol{\Lambda}_k, \quad (2.9)$$

where $\mathbf{D}_k$ is a block-diagonal matrix with $n+1$ blocks of size p_k having generic block equal to $[\mathbf{D}_k]_t = \boldsymbol{\mu}_{q(1/\sigma_t^2)} \mathbf{x}_{k,t-1}' \mathbf{x}_{k,t-1}' \odot \{\boldsymbol{\mu}_{q(\gamma_{k,t})} \boldsymbol{\mu}_{q(\gamma_{k,t})}' + \text{diag}(\boldsymbol{\mu}_{q(\gamma_{k,t})} (1 - \boldsymbol{\mu}_{q(\gamma_{k,t})}))\}$, $\mathbf{Q}_k$ is a tridiagonal block matrix $\mathbf{Q}_k = \mathbf{Q} \otimes \text{diag}(\boldsymbol{\mu}_{q(1/\eta_1^2)}, \dots, \boldsymbol{\mu}_{q(1/\eta_{p_k}^2)})$, and $\boldsymbol{\Lambda}_k$ stacks p_k dimensional vectors defined as $\boldsymbol{\lambda}_{k,t} = \boldsymbol{\mu}_{q(1/\sigma_t^2)} \text{diag}(\boldsymbol{\mu}_{q(\gamma_{k,t})}) \mathbf{x}_{k,t-1}' (\mathbf{y}_t - \sum_{m=1, m \neq k}^K \mathbf{x}_{m,t-1}' \text{diag}(\boldsymbol{\mu}_{q(\gamma_{m,t})}) \boldsymbol{\mu}_{q(\mathbf{b}_{m,t})})$.

Recall from (2.2) that the dynamic of the regression coefficient $\beta_{j,t}$ is a compounding process of the latent state $b_{j,t}$ and the indicator variable $\gamma_{j,t}$. Proposition 3 builds upon Propositions 1 and 2 and provides the optimal variational density for the time-varying regression coefficients for variables $j = 1, \dots, p_k$ in group $k = 1, \dots, K$.

Proposition 3. Let $q^*(\mathbf{b}_k)$ and $q^*(\gamma_{j,t})$, for $j = 1, \dots, p_k$, be the optimal variational densities presented in Propositions 1 and 2. Define $\boldsymbol{\beta}_k = \Gamma_k \mathbf{b}_k$ with $\Gamma_k = \text{diag}(\boldsymbol{\iota}'_{p_k}, \boldsymbol{\gamma}'_{k,1}, \dots, \boldsymbol{\gamma}'_{k,n})$, where $\boldsymbol{\gamma}_{k,t} = (\gamma_{1,t}, \dots, \gamma_{p_k,t})'$ and $\boldsymbol{\iota}_{p_k}$ is a p_k -dimensional vector of ones. The optimal variational density of $\boldsymbol{\beta}_k$ is defined as

$$q^*(\boldsymbol{\beta}_k) = \sum_{\mathbf{s} \in \mathcal{S}} w_{\mathbf{s}} \mathbf{N}_{p_k(n+1)}(\mathbf{D}_{\mathbf{s}} \boldsymbol{\mu}_{q(\mathbf{b}_k)}, \mathbf{D}_{\mathbf{s}}^{1/2} \boldsymbol{\Sigma}_{q(\mathbf{b}_k)} \mathbf{D}_{\mathbf{s}}^{1/2}), \quad (2.10)$$

where $\mathcal{S}$ is a sequence of $\{0, 1\}$ of length $p_k n$ with cardinality $|\mathcal{S}| = 2^{p_k n}$, the diagonal matrix $\mathbf{D}_{\mathbf{s}} = \text{diag}(1, s_{1,1}, \dots, s_{p_k,n})$, and mixing weights:

$$w_{\mathbf{s}} = \prod_{j=1}^{p_k} \prod_{t=1}^n \mu_{q(\gamma_{j,t})}^{s_{j,t}} (1 - \mu_{q(\gamma_{j,t})})^{1-s_{j,t}}, \quad (2.11)$$

where $\mathbf{s} = (s_{1,1}, \dots, s_{j,t}, \dots, s_{p_k,n}) \in \mathcal{S}$ is an element in $\mathcal{S}$. The mean and variance can be computed analytically as

$$\boldsymbol{\mu}_{q(\boldsymbol{\beta}_k)} = \boldsymbol{\mu}_{q(\Gamma_k)} \boldsymbol{\mu}_{q(\mathbf{b}_k)}, \quad (2.12)$$

$$\boldsymbol{\Sigma}_{q(\boldsymbol{\beta}_k)} = (\boldsymbol{\mu}_{q(\gamma_k)} \boldsymbol{\mu}'_{q(\gamma_k)} + \mathbf{W}_{\mu_{q(\gamma_k)}}) \odot \boldsymbol{\Sigma}_{q(\mathbf{b}_k)} + \mathbf{W}_{\mu_{q(\gamma_k)}} \odot \boldsymbol{\mu}_{q(\mathbf{b}_k)} \boldsymbol{\mu}'_{q(\mathbf{b}_k)}, \quad (2.13)$$

where $\mathbf{W}_{\mu_{q(\gamma_k)}}$ is a diagonal matrix with elements $\mu_{q(\gamma_{j,t})}(1 - \mu_{q(\gamma_{j,t})})$, and $\boldsymbol{\mu}_{q(\gamma_k)}$ the collection of $\mu_{q(\gamma_{j,t})}$ for $j = 1, \dots, p_k$ and $t = 1, \dots, n$.

To approximate the posterior of the stochastic volatility process $\mathbf{h}$, we leverage on a Gaussian variational approximation and assume that the corresponding variational density is $q(\mathbf{h}) = \mathbf{N}_{n+1}(\boldsymbol{\mu}_{q(\mathbf{h})}, \boldsymbol{\Sigma}_{q(\mathbf{h})})$. Proposition 4 builds upon Rohde and Wand (2016) to derive a convenient iterative updating scheme to estimate $\boldsymbol{\mu}_{q(\mathbf{h})}, \boldsymbol{\Sigma}_{q(\mathbf{h})}$.

Proposition 4. Let $\boldsymbol{\varepsilon}^2 = \boldsymbol{\varepsilon} \odot \boldsymbol{\varepsilon}$ with components $[\boldsymbol{\varepsilon}^2]_t = (y_t - \boldsymbol{\beta}'_t \mathbf{x}_t)^2$. Assuming a Gaussian approximation $\mathbf{h} \sim \mathbf{N}_{n+1}(\boldsymbol{\mu}_{q(\mathbf{h})}, \boldsymbol{\Sigma}_{q(\mathbf{h})})$, an iterative optimization algorithm can be set as

$$\begin{aligned} \boldsymbol{\Sigma}_{q(\mathbf{h})}^{\text{new}} &= \left[\nabla_{\boldsymbol{\mu}_{q(\mathbf{h})}, \boldsymbol{\mu}_{q(\mathbf{h})}}^2 S(\boldsymbol{\mu}_{q(\mathbf{h})}^{\text{old}}, \boldsymbol{\Sigma}_{q(\mathbf{h})}^{\text{old}}) \right]^{-1}, \\ \boldsymbol{\mu}_{q(\mathbf{h})}^{\text{new}} &= \boldsymbol{\mu}_{q(\mathbf{h})}^{\text{old}} + \boldsymbol{\Sigma}_{q(\mathbf{h})}^{\text{new}} \nabla_{\boldsymbol{\mu}_{q(\mathbf{h})}} S(\boldsymbol{\mu}_{q(\mathbf{h})}^{\text{old}}, \boldsymbol{\Sigma}_{q(\mathbf{h})}^{\text{old}}), \end{aligned}$$

where

$$\begin{aligned} \nabla_{\boldsymbol{\mu}_{q(\mathbf{h})}} S(\boldsymbol{\mu}_{q(\mathbf{h})}^{\text{old}}, \boldsymbol{\Sigma}_{q(\mathbf{h})}^{\text{old}}) &= -\frac{\boldsymbol{\iota}_n}{2} + \frac{1}{2} \mathbb{E}_q(\boldsymbol{\varepsilon}^2) \odot \mathbf{e}^{-\boldsymbol{\mu}_{q(\mathbf{h})}^{\text{old}} + \boldsymbol{\Sigma}_{q(\mathbf{h})}^{\text{old}}/2} \\ &\quad - \boldsymbol{\mu}_{q(1/v^2)} \mathbf{Q} \boldsymbol{\mu}_{q(\mathbf{h})}^{\text{old}}, \end{aligned}$$

and

$$\begin{aligned} \nabla_{\boldsymbol{\mu}_{q(\mathbf{h})}, \boldsymbol{\mu}_{q(\mathbf{h})}}^2 S(\boldsymbol{\mu}_{q(\mathbf{h})}^{\text{old}}, \boldsymbol{\Sigma}_{q(\mathbf{h})}^{\text{old}}) &= -\frac{1}{2} \text{Diag}(\mathbb{E}_q(\boldsymbol{\varepsilon}^2) \odot \mathbf{e}^{-\boldsymbol{\mu}_{q(\mathbf{h})}^{\text{old}} + \boldsymbol{\Sigma}_{q(\mathbf{h})}^{\text{old}}/2}) \\ &\quad - \boldsymbol{\mu}_{q(1/v^2)} \mathbf{Q}, \end{aligned}$$

denote the first and second derivative of $S(\boldsymbol{\mu}_{q(\mathbf{h})}, \boldsymbol{\Sigma}_{q(\mathbf{h})})$ with respect to $\boldsymbol{\mu}_{q(\mathbf{h})}$ and evaluated at $(\boldsymbol{\mu}_{q(\mathbf{h})}^{\text{old}}, \boldsymbol{\Sigma}_{q(\mathbf{h})}^{\text{old}})$, and $\boldsymbol{\Sigma}_{q(\mathbf{h})}^2 = \text{diag}(\boldsymbol{\Sigma}_{q(\mathbf{h})})$.

Section B in supplementary material reports the optimal variational densities $q^*(z_{j,t})$, $q^*(\omega_j)$, $q^*(\eta_j^2)$, $q^*(\xi_j^2)$, and $q^*(v^2)$. For completeness, it also reports the optimal variational density for the homoskedastic case in which the prior for the variance of the residuals in (2.1) is an inverse-gamma $\sigma^2 \sim \text{IG}(A_\sigma, B_\sigma)$.

2.1.2. Smoothing the Inclusion Probabilities

The optimal variational density in Proposition 1 shows that $q(\gamma_{j,t}) = \text{Bern}(\text{expit}(\omega_{q(\gamma_{j,t})}))$, such that the time trajectory of the posterior inclusion probabilities can be obtained as the mean vector $\mathbb{E}_q(\boldsymbol{\gamma}_j | \mathbf{y}) = \text{expit}(\boldsymbol{\omega}_{q(\gamma_j)})$. This approach could produce erratic posterior inclusion probabilities, especially with noisy dependent variables and predictors. The left panel of Figure 1a shows this case in point based on a simple simulation example. The inclusion probability $\mathbb{P}(\gamma_{j,t} | \mathbf{y})$ can be quite noisy. This is practically inconvenient since $\beta_{j,t} = b_{j,t} \gamma_{j,t}$ for $j = 1, \dots, p_k$, such that the dynamic of $\beta_{j,t}$ inherits any eventual erratic behavior of the inclusion probability.

To address this issue, we propose to regularize the estimates of the time trajectory of $\gamma_{j,t}$ for $t = 1, \dots, n$ based on a parametric approximation. In particular, we define $\{q(\gamma_{j,t})\}_{t=1}^n$ as the closest approximation $\{\tilde{q}(\gamma_{j,t})\}_{t=1}^n$ in terms of KL divergence which leads to a smooth sequence of inclusion probabilities, whose expected values coincide with the non-smooth estimates. Proposition 5 formalizes our smoothing approach.

Proposition 5. A smooth estimate for the trajectory of the inclusion probabilities can be achieved assuming $\tilde{q}(\gamma_{j,t}) = \text{Bern}(\text{expit}(\mathbf{w}'_t \mathbf{f}_j))$ such that the expectation of the joint vector $\boldsymbol{\gamma}_j = (\gamma_{j,1}, \dots, \gamma_{j,n})'$ is equal to $\mathbb{E}_{\tilde{q}}(\boldsymbol{\gamma}_j) = \mathbf{W} \mathbf{f}_j$, where $\mathbf{W}$ is a $n \times k$ B-spline basis matrix. The optimal value of $\mathbf{f}_j$ is the solution of the optimization problem $\hat{\mathbf{f}}_j = \arg \max_{\mathbf{f}_j \in \mathbb{R}^k} \psi(\mathbf{f}_j)$ where $\psi(\mathbf{f}_j) = \sum_{t=1}^n \left[(\omega_{q(\gamma_{j,t})} - \mathbf{w}'_t \mathbf{f}_j) \text{expit}(\mathbf{w}'_t \mathbf{f}_j) + \log(1 + \exp(\mathbf{w}'_t \mathbf{f}_j)) \right]$, such that the gradient is equal to $\nabla_{\mathbf{f}} \psi(\mathbf{f}) = \sum_{t=1}^n \mathbf{w}_t (\omega_{q(\gamma_{j,t})} - \mathbf{w}'_t \mathbf{f}) \frac{\text{expit}(\mathbf{w}'_t \mathbf{f})}{1 + \exp(\mathbf{w}'_t \mathbf{f})}$.

The proof is available after Proposition B.4 in the supplementary material. Note that Proposition 5 holds the same for correlated predictors since the original $q(\gamma_{j,t})$ is obtained by taking into account the correlation structure of $\mathbf{b}_k$ (see Proposition 1). Figure 1b shows that as a by-product of a smoother inclusion probability, the dynamics of the regression coefficient are also regularized.

For the interested reader, Algorithm 2 in the supplementary material illustrates the complete iterative optimization scheme for model estimation. The variational posterior in (2.7) is optimized via Coordinate Ascent Variational Inference (CAVI) (Ormerod and Wand 2010). Starting from an initial guess, CAVI iteratively updates each variational factor $q(\theta_i)$ by computing the expectation in (2.5) while holding all other factors fixed. Each update is available in closed form due to the model's conjugate structure. The algorithm cycles through all parameters until the change in the variational moments falls below a specified threshold. The complete pseudocode is provided in Algorithm 2 in Section B of the supplementary material. The corresponding R code is available online at <https://github.com/whitenoise8/BGDvs>.

2.1.3. Convergence Properties and Computational Efficiency

In this section, we discuss the convergence properties of our variational Bayes algorithm, which can then be exploited to improve computational efficiency further. Consider the variables $j = 1, \dots, p_k$ in group k . Proposition 6 extends Ormerod,

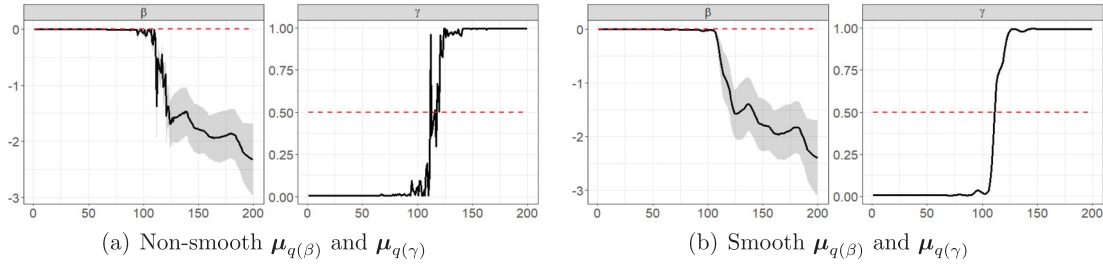


Figure 1. Smoothing $\mu_{q(\beta)}$ and the posterior probability of inclusion $\mathbb{P}(\gamma_{j,t} = 1|y)$ for $t = 1, \dots, n$.

You, and Müller (2017) to the case of dynamic variable selection in a time-varying parameter regression model.

Proposition 6. Assume that the inclusion probabilities for a given variable j in group k at the i -th iteration of the algorithm is such that $\max_{t \in \{1, \dots, n\}} \mu_{q(\gamma_{j,t})}^{(i)} = \epsilon$, for $\epsilon \approx 0$. Let the ω_j covariance update be $\Sigma_{q(\omega_j)}^{(i)} - \Sigma_{q(\omega_j)}^{(i-1)} \geq 0$.² Then we obtain:

- (1) $\mu_{q(\gamma_{j,t})}^{(i+1)} = \text{expit} \left\{ \mu_{q(\omega_{j,t})}^{(i+1)} - \frac{1}{2} \mu_{q(1/\sigma_t^2)}^{(i+1)} x_{j,t-1}^2 \mu_{q(1/\eta_j^2)}^{-1(i+1)} q_{t,t} + O(\epsilon) \right\}$,
 $q_{t,t} = [\mathbf{Q}^{-1}]_{t,t} \geq 0$;
- (2) $\mu_{q(\omega_{j,t})}^{(i+1)} = -1/2 \sum_{k=1}^n s_{t,k} + O(\epsilon)$, $s_{t,k} = [\Sigma_{q(\omega_j)}]_{t,k} \geq 0$;
- (3) $\mu_{q(\omega_{j,t})}^{(i+1)} \leq \mu_{q(\omega_{j,t})}^{(i)}$ decreases after each iteration.

The proof is available in Section C in the supplementary material. From Proposition 6 and Lemma 1 in Ormerod, You, and Müller (2017) two useful numerical results arise: first, for $\epsilon \approx 0$, the following approximation for the update of the inclusion probabilities holds:

$$\mu_{q(\gamma_{j,t})}^{(i+1)} \approx \text{expit} \left\{ \mu_{q(\omega_{j,t})}^{(i+1)} - \frac{1}{2} \mu_{q(1/\sigma_t^2)}^{(i+1)} x_{j,t-1}^2 \mu_{q(1/\eta_j^2)}^{-1(i+1)} q_{t,t} \right\}. \quad (2.14)$$

This implies that for $M^{(i)} = \arg \min_{t \in \{1, \dots, n\}} \mu_{q(\omega_{j,t})}^{(i)} \ll 0$, after i iterations, the sequence $\{\mu_{q(\gamma_{j,t})}^{(i)}\}_{t=1}^n$ is indistinguishable from zero. As a result, the variational densities are concentrated to a point mass at zero $\forall t = 1, \dots, n$. Second, if $\mu_{q(\gamma_{j,t})}^{(i)} \approx 0$, $\forall t$, then all successive updates $i+k \geq i$ imply $\mu_{q(\gamma_{j,t})}^{(i+k)} \approx 0$ since $\mu_{q(\omega_{j,t})}^{(i+k)} \leq \mu_{q(\omega_{j,t})}^{(i)}$ and therefore $M^{(i+k)} \leq M^{(i)}$.

Proposition 6 provides a useful “dimension reduction” strategy whereby we can remove the j -th variable from the set of predictors within the coordinate ascent algorithm. Such an exclusion strategy improves the computational efficiency when p increases but the signal $\bar{p} \leq p$ remains constant, where $\bar{p} = \text{card}(\mathcal{J})$ and $\mathcal{J} = \{j : \sum_{t=1}^n \gamma_{j,t} > 0\}$ collects the indexes of regression coefficients that are included in the model at least for one t . The extended algorithm, which includes Proposition 6, is summarized in Algorithm 3 in the supplementary material.

²A heuristic explanation for this assumption is the following: when the probability $P(\gamma_{j,t} = 1)$ approaches 0 for all t , the feedback that the updates of the j -th indexed variables receive from the observed data diminishes until it becomes negligible, so their approximate posterior converges toward the prior. Since the prior for ω_j has a random walk structure with substantial marginal variance, the posterior variance $\Sigma_{q(\omega_j)}$ increases as this reversion occurs.

Section C.1 in the supplementary material provides a comprehensive visual discussion of the implications of Proposition 6 for a simple simulation set-up. The approximation is exact after fewer than 30 iterations, indicating that the algorithm induces sparsity in the full trajectory of time-varying parameters, as highlighted in (2.14).

2.1.4. Prior Elicitation

We follow Koop and Korobilis (2023) and set $\mu_{q(\gamma_{j,t})}^{(0)} = 1/2$, $\forall t, j$. Next, we follow Ormerod, You, and Müller (2017) and assume weakly-informative inverse-gamma priors for the variances parameters v^2, η_j^2 by setting $A_v = B_v = A_\eta = B_\eta = 0.01$. These are all standard conditionally conjugate priors (Gelman 2006). An important feature of our dynamic variable selection method is that the time-variation of $\gamma_{j,t}$ depends on $\omega_{j,t}$, where the dynamic of the latter is governed by the variance $\xi_j^2 \sim \text{IG}(A_\xi, B_\xi)$.

To shed light on the impact of A_ξ, B_ξ , in Section D in the supplementary material we provide a deeper analysis on the estimated variational mean $\{\mu_{q(\omega_{j,t})}\}_{t=1}^n$ and variance $\Sigma_{q(\omega_j)}$, and the resulting $\{\mu_{q(\gamma_{j,t})}\}_{t=1}^n$, for three alternative limit cases. Specifically, we investigate a series of comparative statics based on the optimal variational densities. This provides a transparent strategy for selecting values of A_ξ and B_ξ .

3. Simulation Study

We evaluate our approach through $M = 100$ Monte Carlo replications, sampling from the data-generating process $y_t = \sum_{j=1}^p \beta_{j,t} x_{j,t-1} + \varepsilon_t$ with $\varepsilon_t \sim \text{N}(0, 0.25)$. Mirroring our empirical application, we set $n = 200$ observations and consider $p \in \{10, 50, 100, 200\}$ predictors. The coefficients follow three distinct patterns: $\beta_{1,t}$ serves as a persistent time-varying intercept ($\gamma_{1,t} = 1$ for all t), $\beta_{2:7,t}$ exhibit various sparsity regimes, while $\beta_{8:p,t}$ remain insignificant throughout ($\gamma_{8:p,t} = 0$ for all t). Complete specifications for $\beta_{1:7,t}$ appear in Section E of the supplementary material.

We examine six variants of our framework: The baseline independent predictor model (BG); a version with smoothing on inclusion probabilities (BGS); two alternative hyperprior specifications for $\xi_j^2 \sim \text{IG}(A_\xi, B_\xi)$ with $B_\xi = 1$ (BGS(1)) and $B_\xi = 10$ (BGS(10)); a constant volatility restriction (BGH); a fully correlated predictor version (BG joint); and a grouped structure variant (BG group). These specifications differ primarily in their variational factorization, as detailed in Section 2.1.

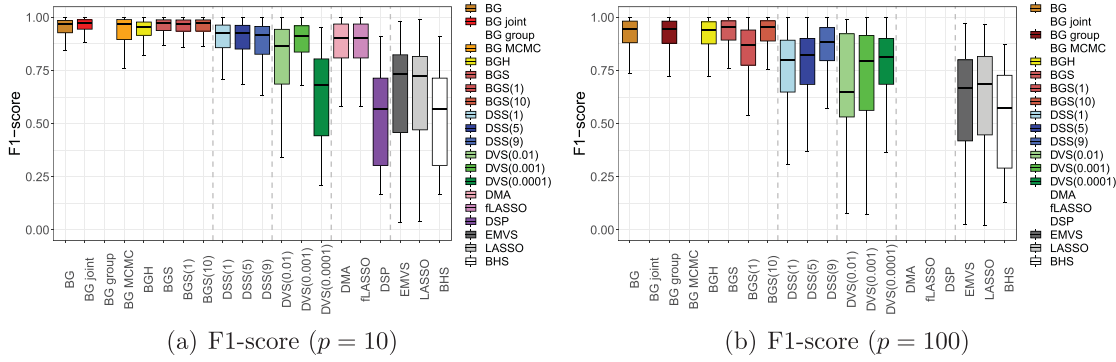


Figure 2. Aggregate F1 scores for parameters $\beta_{2,7,t}$ exhibiting dynamic sparsity patterns. Results shown for independent predictors across dimensions $p = 10, 100$. for $p = 10$ BG group coincides with BG joint and therefore it is not shown, while for $p = 100$ some methods are excluded for computational reason.

For comparison, we benchmark against established variable selection methods including the static spike-and-slab EMVS (Ročková and George 2014), lasso LASSO (Tibshirani 1996), dynamic model averaging DMA (Raftery, Kárný, and Ettler 2010), fused lasso fLASSO (Tibshirani et al. 2005), and dynamic spike-and-slab variants DVS (Koop and Korobilis 2023) and DSS (Ročková and McAlinn 2021). We also consider non-sparse shrinkage approaches: the horseshoe prior BHS (Carvalho, Polson, and Scott 2010) and dynamic shrinkage process DSP (Kowal, Matteson, and Ruppert 2019). All static methods employ rolling-window estimation with 100 observations.

We assess variable selection accuracy using the F1 score, which balances precision and recall as: $F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$, where TP, FP, and FN denote counts of true positives, false positives, and false negatives, respectively. The metric ranges from 0 (no true positives) to 1 (perfect classification with zero false positives/negatives), providing a rigorous evaluation metric (e.g., Bernardi, Bianchi, and Bianco 2024).

3.1. Independent Predictors

We first consider the scenario with independently generated predictors $\{x_{j,t}\}_{j=1}^p \sim N(0, 1)$ at each time point $t = 1, \dots, n$. In this setting, we implement the group factorization for BG group using randomly assigned groups of ten variables. Figure 2 displays the aggregate F1 scores for dynamically sparse parameters $(\beta_{2,7,t})$ with $p \in \{10, 100\}$. Section E.1 in the supplementary material reports the results for $p \in \{50, 200\}$.

The results reveal three key findings. First, static methods (LASSO, EMVS, fLASSO) prove inadequate for time-varying sparsity. Second, the dynamic spike-and-slab prior (Koop and Korobilis 2023) shows significant sensitivity to the variance ratio $\zeta \in \{0.01, 0.001, 0.0001\}$. In contrast, Ročková and McAlinn (2021)’s method remains robust to varying marginal importance weights $\Theta \in \{0.1, 0.5, 0.9\}$, though both DVS and DSS degrade with increasing p . Third, while the dynamic approaches DMA and DSP match newer methods (DSS, DVS) in low dimensions ($p \leq 20$), they cannot be estimated for $p > 20$.

Our variational Bayes approach outperforms all competitors, with its advantages increasing with dimensionality. As expected in this uncorrelated setting, BG group performs similarly to

BG.³ Additional results in Section E.1 in the supplementary material confirm all methods handle consistently (non-)zero parameters adequately, highlighting our framework’s superior detection of complex time-varying sparsity patterns.

3.2. Correlated Predictors

We now examine a scenario where predictors follow $\mathbf{x}_t \sim N_p(\mathbf{0}_p, \Sigma_x)$, with Σ_x estimated from $p \in \{10, 50, 100, 200\}$ randomly selected macroeconomic series in our empirical application (Section 4). This design replicates the correlation structures of the macroeconomic data used. For BG group, we maintain random assignment to groups of ten variables, deliberately avoiding correlation-informed grouping to test robustness. Figure 3 reports the results for $p \in \{10, 100\}$ while the results for $p \in \{50, 200\}$ are reported in Section E.1 in the supplementary material.

In Section E.2 in the supplementary material, we constructed an informed group structure using the empirical correlation matrix of the p covariates: we use the functions `corclust` and `cvtree` in the `klAR` R-package to define groups such that minimum within-group correlation is 0.1 and between-group correlation is minimized. We also consider random groups of sizes 5, 10, and 20, as well as the fully independent factorization. Figure E.3 shows that for $p = 100, 200$ the choice of grouping does not substantially affect the performance of the variable selection in terms of F1-score.

The results reveal two key insights: First, correlation-induced selection bias reduces F1 scores relative to the independent case, with the deterioration accelerating for $p > 50$. Second, BG group maintains comparable performance to BG despite its agnostic grouping, suggesting robustness to imperfect group specifications. These results align with theoretical expectations regarding correlated covariates and demonstrate the practical applicability of our method to macroeconomic data structures.

3.3. Sparse versus Dense Data Generating Processes

E.3 in the supplementary material reports a series of additional simulation results whereby we fix $p = 100$ and consider increasingly dense settings with $p_{\text{non-zero}} = 20, 50, 80$

³For $p = 10$, BG joint and BG group coincide; we report only BG group.

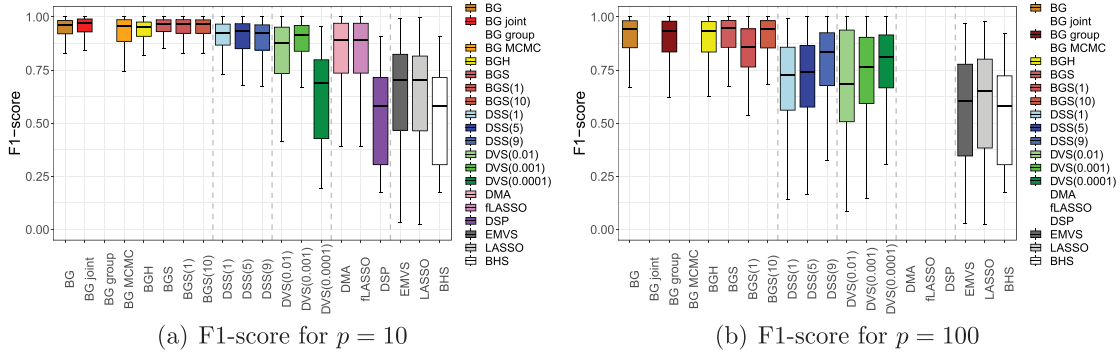


Figure 3. Aggregate F1 score for parameters $\beta_{2,7,t}$ exhibiting dynamic sparsity patterns over the sample period with correlated predictors. Results are reported for dimensions $p = 10, 100$. For $p = 10$ BG group coincides with BG joint and therefore it is not shown, while for $p = 100$ some methods are excluded for computational reason.

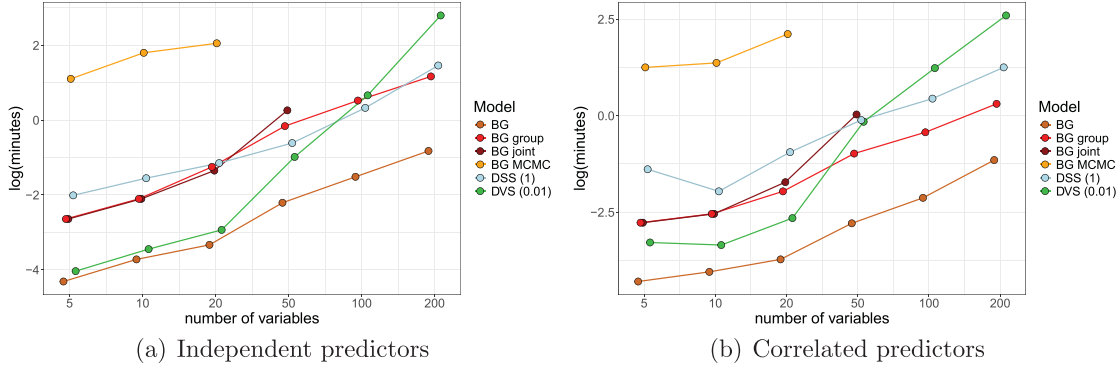


Figure 4. Computation time (log minutes) for dynamic variable selection methods.

non-zero time-varying coefficients. Figure E.4 shows the results. Our methods (BG, BG joint, BG group, BGS) maintain high F1-scores (close to 1.0) even as the signal becomes dense, with 80 out of 100 coefficients being non-zero. This suggests that our approach does not artificially impose sparsity when the true data-generating process is dense. In contrast, competing dynamic variable selection methods, such as DSS and DVS, exhibit greater variability and lower median F1-scores as density increases. Static methods (EMVS, LASSO) also exhibit degraded performance in denser settings. These results indicate that our method can recover the correct signal with sufficient accuracy across different sparsity regimes.

3.4. Computational Efficiency

We assess computational efficiency for $p \in \{5, 10, 20, 50, 100, 200\}$ with $n = 200$ observations. Figure 4 compares log-scale computation times (minutes) for key methods: BG, BG group, BG joint, BG MCMC, DSS (1), and DVS (0.01). Table E.2 in the supplementary material provides computational times for all remaining models.⁴

Three key findings emerge: First, while DVS shows efficiency for $p < 50$ due to its variational implementation, our BG method achieves 5-10 $\times$ speed gains at higher dimensions with independent predictors, and 100 $\times$ gains with correlated

predictors. Second, the fully correlated BG joint becomes computationally prohibitive for $p > 50$, whereas the group-structured BG group maintains feasible computation costs, outperforming DVS at $p = 200$. Third, MCMC estimation requires 2.5–3 orders of magnitude more time than variational Bayes, becoming impractical for $p > 20$. The results demonstrate our method's superior scalability, particularly in data-rich environments where correlation structures and high dimensionality pose dual challenges.

3.5. Variational Bayes versus MCMC

The computational efficiency of variational Bayes (VB) comes with inherent tradeoffs in posterior approximation quality that merit careful consideration. Where MCMC samples from the exact joint posterior $p(\boldsymbol{\theta}|\mathbf{y})$, VB approximates it through an optimized factorized density $q(\boldsymbol{\theta})$. This formulation, particularly under mean-field assumptions, systematically produces under-dispersed posterior approximations (Blei, Kucukelbir, and McAuliffe 2017), a well-documented limitation of variational methods.

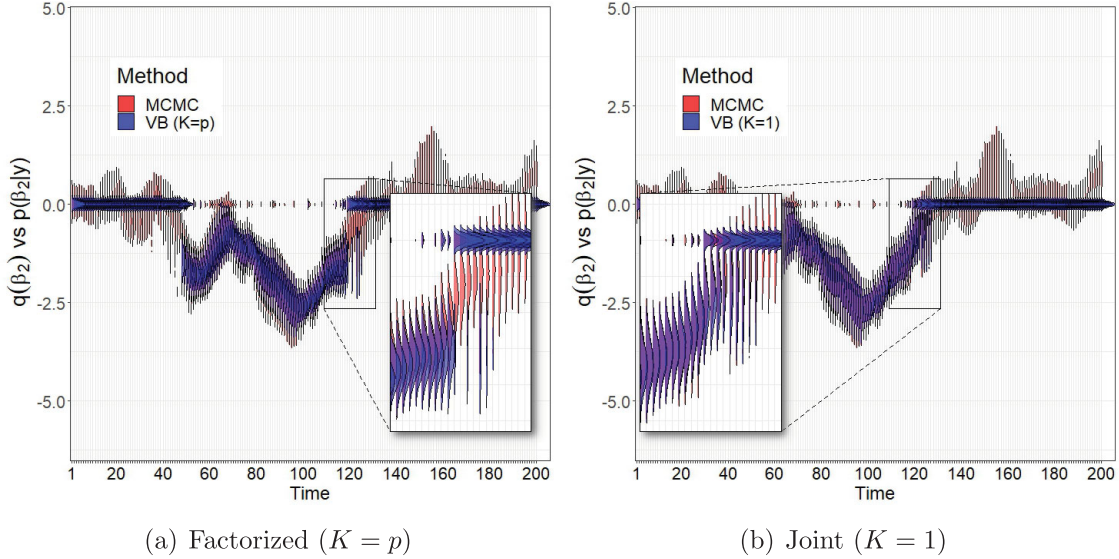
We evaluate approximation accuracy, we follow Wand et al. (2011) and use an L_1 -based metric $\text{ACC}(\boldsymbol{\theta}) = [1 - 0.5 \int |q(\boldsymbol{\theta}) - p(\boldsymbol{\theta}|\mathbf{y})| d\boldsymbol{\theta}] \times 100\%$, where $p(\boldsymbol{\theta}|\mathbf{y})$ represents the ground-truth MCMC posterior. Table 1 reports the accuracy of $q^*(\beta_{j,t})$ with respect to its MCMC counterpart $p(\beta_{j,t}|\mathbf{y})$, aggregated by sparsity pattern type: $\beta_{1,t}$ (intercept), $\beta_{2,t}$ (one switch), $\beta_{3,t}$ (two switches), and $\beta_{4,t}$ (short-lived significance). The results show that VB achieves 84–92% approximation accuracy under full

⁴We reimplemented all methods in R when they were originally coded in different languages, and we executed all implementations on the same 2.5 GHz Intel Xeon W-2175 processor with 32 GB of RAM.

Table 1. The table shows the accuracy of $q^*(\beta_{jt})$ with respect to its MCMC $p(\beta_{jt}|y)$ counterpart.

Parameter	Independent variables						Correlated variables					
	$p = 5$		$p = 10$		$p = 20$		$p = 5$		$p = 10$		$p = 20$	
	BG	BG joint	BG	BG joint	BG	BG joint	BG	BG joint	BG	BG joint	BG	BG joint
$\beta_{1,t}$	88.15	91.84	88.96	91.94	88.99	91.53	87.17	91.4	87.53	91.13	87.78	90.4
$\beta_{2,t}$	88.68	92.53	90.12	93.25	89.81	92.65	87.12	91.67	87.84	92.08	87.16	90.56
$\beta_{3,t}$	84.74	89.01	85.73	89.51	84.93	87.83	83.78	88.23	84.34	88.61	84.11	87.23
$\beta_{4,t}$	94.67	94.84	95.05	95.19	95.44	95.64	94.43	94.59	95.05	95.12	93.78	93.93

This is aggregated based on the type of sparsity dynamic; that is, $\beta_{1,t}$ (intercept), $\beta_{2,t}$ (one switch), $\beta_{3,t}$ (two switches), and $\beta_{4,t}$ (short-lived significance).

**Figure 5.** Posterior comparison (VB: blue, MCMC: red) for $\beta_{2,t}$ during regime transitions, demonstrating VB's variance underestimation in (a) and improved uncertainty characterization in (b).

factorization ($K = p$), with the joint specification ($K = 1$) improving to 89–96% across different coefficient types. This accuracy gradient indicates that preserving dependency structures within the variational family improves the quality of posterior approximations.

The practical consequences of these approximations become evident in Figure 5, which compares posterior densities for a transitioning coefficient $\beta_{2,t}$. The factorized VB approximation ($K = p$) produces markedly sharper transitions between inclusion/exclusion states, but it underestimates posterior variance during regime switches. In contrast, the joint specification ($K = 1$) better preserves the uncertainty characteristics of the MCMC ground truth. This underestimation of variance stems directly from Proposition 3, in which the factorized approximation tends to collapse into a single mixture component.

Notably, these theoretical limitations have minimal operational impact on variable selection performance. The VB approximation preserves decision-critical information—particularly the inclusion probabilities $\mathbb{P}(\gamma_{j,t} = 1)$ —explaining its comparable selection accuracy to MCMC in Figures 2–3. This practical robustness, combined with VB's orders-of-magnitude speed advantage (see Figure 4), establishes its clear superiority for high-dimensional applications. The method's scalability enables implementations in which MCMC is computationally prohibitive, particularly in recursive forecasting contexts that require frequent model re-estimation. To summarize, while theoretical accuracy tradeoffs exist, VB's preserved decision

quality and computational tractability make it the preferred approach for variable selection in data-rich environments.

4. Empirical Analysis of Inflation Predictability

We assess the empirical performance of our dynamic variable selection approach in an inflation-forecasting application, using a comprehensive set of macroeconomic predictors (Stock and Watson 2006). This context is particularly appropriate given the well-documented parameter instability in inflation dynamics, especially for out-of-sample forecasting (Huber, Koop, and Onorante 2021).

Our dataset comprises 229 quarterly macroeconomic variables from the FRED-QD database (McCracken and Ng 2020), spanning 1967:Q3 to 2022:Q2. Following standard practice, we transform all predictors to ensure stationarity. The variables are organized into 14 economically meaningful categories: (1) National Income and Product Accounts, (2) Industrial Production, (3) Employment and Unemployment, (4) Housing, (5) Inventories, Orders and Sales, (6) Prices, (7) Earnings and Productivity, (8) Interest Rates, (9) Money and Credit, (10) Household Balance Sheets, (11) Exchange Rates, (12) Other Indicators, (13) Stock Markets, and (14) Non-Household Balance Sheets. We supplement these with two lags of each response variable. The correlation structure of these predictors is analyzed in Section F.1 of the supplementary material. The categories inform our variational factorization from (2.7), allowing us to

Table 2. This table reports the out-of-sample root mean squared prediction error across models.

Horizon	Inflation	Benchmarks			Static selection (rolling)				Dynamic selection						
		TVI	AR (2)	TVAR (2)	PCA (5)	HS	EMVS	GLP	BGS	BG group	BGH	BG	DSS (5)	DSS (9)	DVS
1	CPIAUCSL	2.38	2.54	2.48	3.38	3.29	4.67	3.13	2.49	2.43	2.51	2.75	3.38	3.44	3.31
1	CPILFESL	1.08	1.02	1.08	2.75	1.26	1.23	1.92	0.96	0.97	0.99	1.02	1.26	1.27	1.07
1	GDPCTPI	1.13	1.06	1.12	1.91	1.56	1.29	1.48	1.03	1.01	1.04	1.02	1.33	1.33	1.17
1	PCECTPI	1.75	1.82	1.78	2.40	2.41	2.17	2.14	1.77	1.73	1.84	1.80	1.90	1.92	2.12
2	CPIAUCSL	1.65	2.28	1.79	3.04	3.30	3.56	2.61	1.59	1.63	1.71	1.61	2.14	2.16	2.84
2	CPILFESL	0.78	0.86	0.92	2.37	1.28	1.21	1.45	0.57	0.58	0.59	0.57	0.83	0.84	0.71
2	GDPCTPI	0.69	0.99	0.92	1.76	1.37	1.19	1.28	0.62	0.66	0.61	0.62	1.11	1.12	1.02
2	PCECTPI	1.18	1.73	1.41	2.23	2.48	2.34	1.85	1.22	1.22	1.29	1.21	1.48	1.48	1.98
4	CPIAUCSL	0.92	1.83	0.92	2.53	3.21	2.00	2.06	0.99	0.94	0.97	1.07	1.44	1.44	1.25
4	CPILFESL	0.68	0.96	0.66	2.29	1.31	1.07	1.41	0.37	0.38	0.44	0.37	0.77	0.77	0.70
4	GDPCTPI	0.53	1.03	0.82	1.74	1.43	1.36	1.43	0.39	0.39	0.39	0.39	1.00	1.01	0.81
4	PCECTPI	0.69	1.46	0.70	2.00	2.92	1.87	1.56	0.74	0.71	0.72	0.79	1.12	1.13	1.03
8	CPIAUCSL	0.43	1.34	0.45	2.04	1.94	1.92	2.05	0.47	0.47	0.46	0.47	0.64	0.64	0.62
8	CPILFESL	0.19	0.87	0.28	1.99	1.17	1.32	1.15	0.21	0.19	0.21	0.21	0.28	0.28	0.24
8	GDPCTPI	0.26	0.90	0.34	1.48	1.11	1.33	1.26	0.26	0.26	0.26	0.26	0.44	0.45	0.36
8	PCECTPI	0.32	1.02	0.38	1.64	1.49	1.49	1.44	0.36	0.35	0.38	0.37	0.51	0.51	0.45

The sample period is from 1967Q3 to 2022Q3. The first prediction is generated in 1997Q3.

incorporate economic structure into the estimation while maintaining computational efficiency.

We examine four distinct inflation measures as response variables: the total Consumer Price Index (CPIAUCSL), the core Consumer Price Index (CPILFESL), the GDP deflator (GDPCTPI), and the Personal Consumption Expenditures deflator (PCECTPI). The parenthetical codes correspond to the original variable identifiers in the FRED-QD database. For each price series P_t , we compute h -quarter ahead annualized inflation rates as $y_{t+h} = (400/h) \ln(P_{t+h}/P_t)$.

4.1. Out-of-Sample Forecasting Performance

We evaluate h -quarter ahead point and density forecasts using four specifications of our approach: the grouped dynamic Bernoulli-Gaussian model (BG group), baseline independent coefficients (BG), constant volatility variant (BGH), and smoothed inclusion probability specification (BGS). All models employ uninformative hyperparameters ($A_v = 0.01$, $B_v = 0.01$; $A_\eta = 0.01$, $B_\eta = 0.01$; $A_\xi = 2$, $B_\xi = 5$) following sensitivity analyses in [Section D in the supplementary material](#).

We compare against three classes of alternative approaches. Benchmark models include the local-level specification TVI (Stock and Watson 2007), standard AR(2), and time-varying AR(2) (TVAR(2)) (Koop and Korobilis 2023). Both TVI and TVAR(2) account for stochastic volatility. Static selection methods comprise the horseshoe prior BHS, spike-and-slab EMVS (Ročková and George 2014), Dirac specification GLP (Giannone, Lenza, and Primiceri 2021), and five-factor regression F5. Dynamic alternatives consist of DVS (Koop and Korobilis 2023) with $\underline{c} = 0.0001$ – which simulation shows works comparatively well (see [Figure 2](#)) – and DSS (Ročková and McAlinn 2021) with $\Theta \in \{0.5, 0.9\}$.

[Table 2](#) reports root mean squared errors $\text{RMSE}_i = \sqrt{n^{-1} \sum_{t=1}^n \hat{e}_{i,t}^2}$, where $\hat{e}_{i,t}$ denotes the forecast error for model i at time t . Our recursive evaluation spans 1997:Q3 to 2022:Q3 using 30-year rolling windows, with results color-coded from white (best RMSE) to dark red (worst RMSE) for visual comparison.

The results reveal several important patterns. First, the strong performance of TVI confirms Stock and Watson (2007)’s insight regarding persistent inflation components—particularly during the Great Moderation (1985–2007) when well-anchored inflation expectations diminished the direct influence of macroeconomic predictors. Second, all static selection methods underperform, suggesting that inflation-predictor relationships evolve significantly across economic regimes. Third, while dynamic alternatives DSS and DVS improve upon static approaches, our BGS specification achieves superior forecasting accuracy with average RMSEs of 1.56, 1.00, 0.62, and 0.33 for horizons $h = 1, 2, 4, 8$, respectively. The minimal improvement from economic groupings (BG group yields RMSEs of 1.65, 1.00, 0.65, and 0.33), suggesting relevant inflation signals frequently cut across economic categories, often emerging simultaneously from diverse sectors during specific periods.

In [Section F.2 of the supplementary material](#), we assess the statistical significance of our variational Bayes approach’s forecasting gains using Diebold-Mariano (DM) tests (Diebold and Mariano 2002). For brevity, we report results for the Total CPI (CPIAUCSL), the GDP deflator (GDPCTPI), and the PCE deflator (PCECTPI) for $h = 1, 2, 4$ quarters ahead. The tests evaluate $\mathcal{H}_0 : \text{MSE}^C = \text{MSE}^R$ against $\mathcal{H}_a : \text{MSE}^C > \text{MSE}^R$, where MSE^C and MSE^R represent the mean squared errors of column (comparison) and row (reference) models, respectively. The results demonstrate that our BG, BGS, and BG group specifications significantly outperform all other macroeconomic predictor-based approaches at conventional significance levels, while achieving point forecasts statistically indistinguishable from the TVI specification.

This finding substantiates two key insights: First, our dynamic variable selection framework successfully distills the predictive content from a large macroeconomic dataset without sacrificing the robustness of simpler benchmarks. Second, the maintained accuracy relative to TVI confirms our method’s ability to adaptively discard irrelevant predictors while preserving the core inflation dynamics captured by the local-level model.

Table 3. This table reports the out-of-sample log-predictive score calculated across models.

Horizon	Inflation	Benchmarks			Static selection (rolling)				Dynamic selection						
		TVI	AR (2)	TVAR (2)	PCA (5)	HS	EMVS	GLP	BGS	BG group	BGH	BG	DSS (5)	DSS (9)	DVS
1	CPIAUCSL	-2.5	-2.3	-2.3	-2.8	-3.0	-4.2	-3.0	-2.6	-2.6	-2.4	-2.6	-3.53	-3.65	-5.87
1	CPILFESL	-1.6	-1.7	-1.1	-2.5	-1.6	-2.4	-2.2	-1.4	-1.4	-1.3	-1.6	-2.18	-2.17	-1.55
1	GDPCTPI	-2.2	-1.7	-1.5	-2.6	-2.1	-2.5	-1.7	-1.7	-1.9	-1.6	-1.7	-2.49	-2.37	-3.27
1	PCECTPI	-2.4	-2.0	-2.0	-2.5	-2.5	-2.8	-2.5	-2.1	-2.2	-2.1	-2.3	-2.95	-3.13	-4.77
2	CPIAUCSL	-3.1	-2.3	-2.3	-2.8	-3.0	-3.7	-3.0	-2.2	-2.2	-2.3	-2.3	-2.87	-2.95	-5.00
2	CPILFESL	-6.2	-1.7	-0.8	-2.3	-1.9	-2.2	-1.9	-0.9	-1.1	-1.3	-0.9	-1.94	-1.91	-1.22
2	GDPCTPI	-5.3	-1.6	-1.3	-2.6	-2.4	-2.5	-1.8	-1.2	-1.4	-1.3	-1.2	-2.49	-2.34	-2.02
2	PCECTPI	-3.7	-2.1	-1.9	-2.6	-2.7	-3.0	-2.5	-1.9	-1.9	-2.1	-2.0	-2.61	-2.67	-3.59
4	CPIAUCSL	-4.0	-2.2	-2.1	-2.6	-2.9	-3.2	-2.7	-1.9	-1.9	-1.9	-1.9	-2.61	-2.64	-2.55
4	CPILFESL	-5.0	-1.8	-0.5	-2.2	-2.0	-1.7	-2.1	-0.5	-0.9	-1.2	-0.6	-1.65	-1.54	-0.88
4	GDPCTPI	-7.3	-1.7	-0.8	-2 > 3	-2.2	-2.6	-2.1	-0.8	-1.0	-1.2	-0.9	-2.42	-2.46	-1.23
4	PCECTPI	-4.9	-2.0	-1.5	-2.5	-2.6	-2.9	-2.4	-1.7	-1.6	-1.7	-1.5	-2.40	-2.29	-2.10
8	CPIAUCSL	-7.5	-1.9	-0.9	-2.2	-2.4	-3.6	-2.2	-1.2	-1.3	-1.4	-1.2	-2.87	-2.97	-1.32
8	CPILFESL	-2.9	-1.4	-0.1	-1.8	-1.8	-1.9	-1.6	-0.9	-0.7	-0.8	-0.6	-1.26	-1.29	-0.10
8	GDPCTPI	-7.7	-1.6	-0.5	-1.8	-1.9	-2.6	-2.1	-1.1	-1.1	-1.4	-1.2	-2.87	-2.98	-0.62
8	PCECTPI	-7.3	-1.6	-0.7	-2.0	-2.2	-3.2	-2.0	-1.0	-1.0	-1.3	-1.1	-2.42	-2.38	-0.86

The sample period is from 1967Q3 to 2022Q3. The first prediction is generated in 1997Q3.

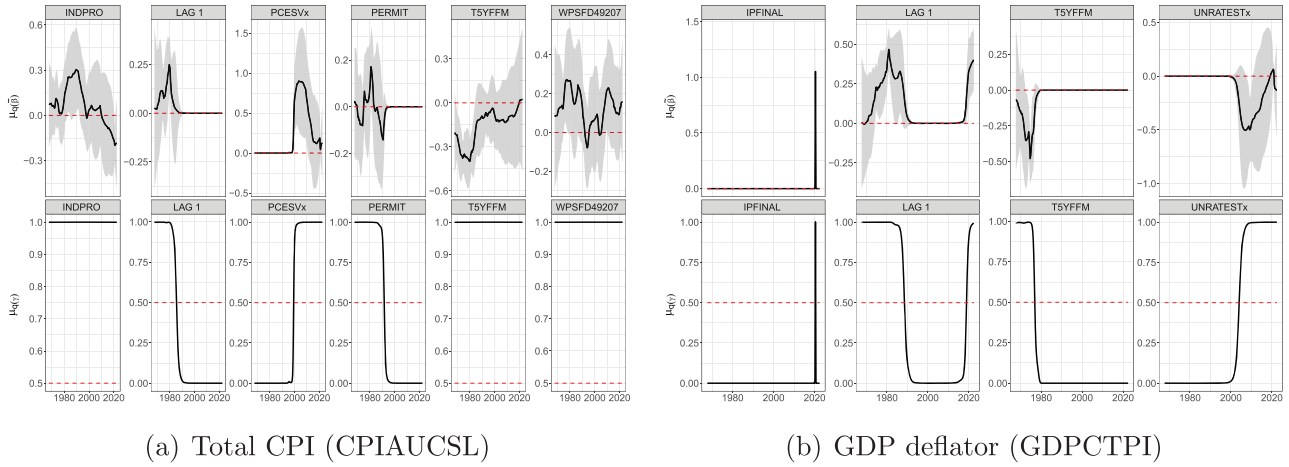
**Figure 6.** Time-varying coefficient estimates and inclusion probabilities for four inflation measures.

Table 3 presents log-predictive scores $LS_i = n^{-1} \sum_{t=1}^n \log(S_{i,t})$, where $\log(S_{i,t})$ denotes the time- t predictive likelihood for model i . Higher scores indicate better performance in density forecasting. Three key patterns emerge from these results.

First, the strong performance of TVAR(2) in density forecasting, particularly at longer horizons, underscores the critical role of accurately characterizing inflation uncertainty. While TVI produces accurate point forecasts, its density forecasts prove excessively concentrated around the conditional mean, systematically underestimating true inflation uncertainty.

Second, rolling-window implementations of static shrinkage and selection methods (BHS, EMVS, GLP) perform similarly to dynamic alternatives DSS and DVS at shorter horizons ($h \leq 4$), suggesting that adaptive methods may initially overfit when estimating predictive densities.

Most significantly, our BG, BGS, and BG group specifications demonstrate superior density forecasting performance across all horizons. This robust performance indicates that our approach successfully captures both gradual inflation dynamics and abrupt changes in volatility regimes. The consistent outperformance across all four inflation measures further confirms the method's ability to adapt to different conditions while maintaining well-calibrated uncertainty estimates.

4.2. Retrospective Analysis of Inflation Drivers

Our dynamic variable selection framework provides both accurate forecasts and interpretable estimates of time-varying regression parameters $\beta_{j,t}$ that reveal the macroeconomic drivers of inflation. Focusing on one-quarter ahead forecasts, Figure 6 presents coefficient estimates $\mu_q(\beta_{j,t})$ and inclusion probabilities $\mu_q(\gamma_{j,t})$ from the BGS model, showing only predictors active for significant periods. For brevity, we report results for the Total CPI and the GDP deflator. The remaining plots for PCE deflator and Core CPI are available in Section F.3 of the supplementary material.

The analysis yields several important findings about inflation dynamics. First, the sparse sets of active predictors support the unobserved components view advanced by Stock and Watson (2007) and Harvey, Trimbur, and Van Dijk (2007), where inflation variation primarily reflects latent local trends rather than observable economic drivers. This sparsity is particularly evident during the Great Moderation period from the mid-1980s through 2007, when inflation expectations became well-anchored.

Second, while certain predictors, such as industrial production (INDPRO), 5-year Treasury rates (T5YFFM), and producer

prices (WPSFD49207), appear consistently across different inflation measures, their timing and magnitude of influence vary. Lagged inflation is particularly significant during the 1970s oil shocks and the Volcker disinflation period.

Third, the estimates reveal distinct regimes in monetary policy transmission. While 5-year Treasury rates maintain negative predictive relationships with CPI and PCE inflation throughout the sample, they lose predictive power for GDP deflator inflation following the financial deregulation of the early 1980s. Simultaneously, supply-side factors such as industrial production exhibit persistent positive effects, supporting cost-push theories of inflation.

Fourth, the results provide nuanced evidence about traditional inflation relationships. Short-term unemployment gains predictive power for the GDP deflator following the 2008 financial crisis, offering conditional support for the Phillips curve, consistent with recent evidence of its nonlinear, slack-dependent nature (Blanchard 2016). The framework also successfully captures transient shocks, such as the 2020 spikes in industrial production (IPFINAL) and consumption (PCECC96) that anticipated the subsequent inflation surge.

5. Concluding Remarks

We develop a variational Bayes method for dynamic variable selection in high-dimensional time-varying parameter regressions. The approach delivers significant computational efficiency gains over MCMC while maintaining accuracy. In inflation forecasting applications with 220+ macroeconomic predictors, our method outperforms existing alternatives and provides interpretable time-varying estimates of inflation drivers. The results highlight the importance of dynamic selection when modeling economic time series.

Disclosure statement

The authors report there are no competing interests to declare.

ORCID

Nicolas Bianco  <http://orcid.org/0000-0001-5430-7137>
Daniele Bianchi  <http://orcid.org/0000-0002-2321-2108>

References

- Bernardi, M., Bianchi, D., and Bianco, N. (2024), "Variational Inference for Large Bayesian Vector Autoregressions," *Journal of Business & Economic Statistics*, 42, 1066–1082. DOI: [10.1080/07350015.2023.2290716](https://doi.org/10.1080/07350015.2023.2290716). [6]
- Blanchard, O. (2016), "The Phillips Curve: back to the '60s?" *American Economic Review*, 106, 31–34. DOI: [10.1257/aer.p20161003](https://doi.org/10.1257/aer.p20161003). [11]
- Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017), "Variational Inference: A Review for Statisticians," *Journal of the American Statistical Association*, 112, 859–877. DOI: [10.1080/01621459.2017.1285773](https://doi.org/10.1080/01621459.2017.1285773). [2,7]
- Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010), "The Horseshoe Estimator for Sparse Signals," *Biometrika*, 97, 465–480. DOI: [10.1093/biomet/asq017](https://doi.org/10.1093/biomet/asq017). [6]
- Diebold, F. X., and Mariano, R. S. (2002), "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 20, 134–144. DOI: [10.1198/073500102753410444](https://doi.org/10.1198/073500102753410444). [9]
- Gelman, A. (2006), "Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper)," *Bayesian Analysis*, 1, 515–534. DOI: [10.1214/06-BA117A](https://doi.org/10.1214/06-BA117A). [5]
- Giannone, D., Lenza, M., and Primiceri, G. E. (2021), "Economic Predictions with Big Data: The Illusion of Sparsity," *Econometrica*, 89, 2409–2437. DOI: [10.3982/ECTA17842](https://doi.org/10.3982/ECTA17842). [9]
- Harvey, A. C., Trimbur, T. M., and Van Dijk, H. K. (2007), "Trends and Cycles in Economic Time Series: A Bayesian Approach," *Journal of Econometrics*, 140, 618–649. DOI: [10.1016/j.jeconom.2006.07.006](https://doi.org/10.1016/j.jeconom.2006.07.006). [10]
- Huber, F., Koop, G., and Onorante, L. (2021), "Inducing Sparsity and Shrinkage in Time-Varying Parameter Models," *Journal of Business & Economic Statistics*, 39, 669–683. DOI: [10.1080/07350015.2020.1713796](https://doi.org/10.1080/07350015.2020.1713796). [1,2,8]
- Koop, G., and Korobilis, D. (2023), "Bayesian Dynamic Variable Selection in High Dimensions," *International Economic Review*, 64, 1047–1074. DOI: [10.1111/iere.12623](https://doi.org/10.1111/iere.12623). [1,2,5,6,9]
- Kowal, D. R., Matteson, D. S., and Ruppert, D. (2019), "Dynamic Shrinkage Processes," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 81, 781–804. DOI: [10.1111/rssb.12325](https://doi.org/10.1111/rssb.12325). [2,6]
- McCracken, M., and Ng, S. (2020), "FRED-QD: A Quarterly Database for Macroeconomic Research No. w26872." Technical Report, National Bureau of Economic Research. [2,8]
- Mogliani, M., and Simoni, A. (2024), "Bayesian bi-Level Sparse Group Regressions for Macroeconomic Forecasting," *arXiv preprint arXiv:2404.02671*. <https://arxiv.org/abs/2404.02671>. [2,3]
- Ormerod, J. T., and Wand, M. P. (2010), "Explaining Variational Approximations," *The American Statistician*, 64, 140–153. DOI: [10.1198/tast.2010.09058](https://doi.org/10.1198/tast.2010.09058). [2,4]
- Ormerod, J. T., You, C., and Müller, S. (2017), "A Variational Bayes Approach to Variable Selection," *Electronic Journal of Statistics*, 11, 3549–3594. DOI: [10.1214/17-EJS1332](https://doi.org/10.1214/17-EJS1332). [5]
- Polson, N. G., Scott, J. G., and Windle, J. (2013), "Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables," *Journal of the American Statistical Association*, 108, 1339–1349. DOI: [10.1080/01621459.2013.829001](https://doi.org/10.1080/01621459.2013.829001). [3]
- Raftery, A. E., Kárný, M., and Ettler, P. (2010), "Online Prediction under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill," *Technometrics: a Journal of Statistics for the Physical, Chemical, and Engineering Sciences*, 52, 52–66. DOI: [10.1198/TECH.2009.08104](https://doi.org/10.1198/TECH.2009.08104). [6]
- Ray, K., and Szabó, B. (2022), "Variational Bayes for High-Dimensional Linear Regression with Sparse Priors," *Journal of the American Statistical Association*, 117, 1270–1281. DOI: [10.1080/01621459.2020.1847121](https://doi.org/10.1080/01621459.2020.1847121). [2]
- Rohde, D., and Wand, M. P. (2016), "Semiparametric Mean Field Variational Bayes: General Principles and Numerical Issues," *Journal of Machine Learning Research*, 17, 1–47. [4]
- Ročková, V., and George, E. I. (2014), "EMVS: The em Approach to Bayesian Variable Selection," *Journal of the American Statistical Association*, 109, 828–846. DOI: [10.1080/01621459.2013.869223](https://doi.org/10.1080/01621459.2013.869223). [1,2,6,9]
- Ročková, V., and McAlinn, K. (2021), "Dynamic Variable Selection with Spike-and-Slab Process Priors," *Bayesian Analysis*, 16, 233–269. [1,2,6,9]
- Stock, J. H., and Watson, M. W. (2006), "Forecasting with Many Predictors," *Handbook of Economic Forecasting*, 1, 515–554. [2,8]
- (2007), "Why Has us Inflation Become Harder to Forecast?," *Journal of Money, Credit and Banking*, 39, 3–33. DOI: [10.1111/j.1538-4616.2007.00014.x](https://doi.org/10.1111/j.1538-4616.2007.00014.x). [2,9,10]
- Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58, 267–288. DOI: [10.1111/j.2517-6161.1996.tb02080.x](https://doi.org/10.1111/j.2517-6161.1996.tb02080.x). [6]
- Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005), "Sparsity and Smoothness via the Fused Lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67, 91–108. DOI: [10.1111/j.1467-9868.2005.00490.x](https://doi.org/10.1111/j.1467-9868.2005.00490.x). [6]
- Uribe, P., and Lopes, H. (2025), "Dynamic Sparsity on Dynamic Regression Models," *Brazilian Journal of Probability and Statistics*, 39, 366–386. [2]
- Wand, M. P., Ormerod, J. T., Padoan, S. A., and Frühwirth, R. (2011), "Mean Field Variational Bayes for Elaborate Distributions," *Bayesian Analysis*, 6, 847–900. DOI: [10.1214/11-BA631](https://doi.org/10.1214/11-BA631). [7]