Adaptive Bayesian Shrinkage of High-Dimensional Panel VARs*

Zhiruo Zhang† Firmin Doko Tchatoka‡

Qazi Haque§

School of Economics and Public Policy, The University of Adelaide, Australia

August 26, 2025

Abstract

This paper develops a Bayesian framework for estimating high-dimensional panel vector autoregressions (PVARs). We propose a novel approach that combines Bayesian shrinkage with adaptive variable selection to effectively tackle over-parameterization and sparsity common in high-dimensional panels. By employing Laplace-based spike-and-slab priors on model coefficients, the framework flexibly captures both cross-sectional inter-dependencies and unit-specific heterogeneity, offering a powerful and robust tool for structured inference. Monte Carlo simulations demonstrate that our method outperforms existing regularization techniques in terms of estimation accuracy and forecasting performance. Empirically, the framework uncovers asymmetric financial contagion within euro area sovereign bond markets and produces stable, reliable forecasts across a multi-country macroeconomic panel. These findings highlight the effectiveness of adaptive shrinkage in modeling heterogeneous and evolving linkages within complex panel data systems.

Key words: Dynamic Interdependency, Cross-sectional Heterogeneity, Bayesian Lasso, Variable Selection, Spike and Slab Prior, Financial Contagion.

JEL classification: C32; C54; C55; E17; F41.

*This research was supported by the Australian Research Council through Discovery Grants DP200101498 and DP210103094. Doko Tchatoka gratefully acknowledges this financial support. Additional support was provided by the Phoenix High Performance Computing (HPC) service at the University of Adelaide. Haque acknowledges funding from the Australian Research Council under Grant DP240100970.

†School of Economics and Public Policy, The University of Adelaide. 10 Pulteney Street, Adelaide SA 5005, Australia. Email: zhiruo.zhang@adelaide.edu.au

‡Corresponding author, School of Economics and Public Policy, The University of Adelaide. 10 Pulteney Street, Adelaide SA 5005, Australia. Email: firmin.dokotchatoka@adelaide.edu.au

§School of Economics and Public Policy, The University of Adelaide. 10 Pulteney Street, Adelaide SA 5005, Australia. Email: qazi.haque@adelaide.edu.au

1 Introduction

Advances in data collection and statistical methods have enabled the widespread use of high-dimensional datasets and the development of models that account for complex data structures.[1] Panel Vector Autoregression (PVAR) models have emerged as a popular tool for analyzing high-dimensional macroeconomic panels, owing to their ability to capture both cross-sectional and dynamic dependencies (see Canova and Ciccarelli, 2013 for a survey). Despite their appeal, PVAR models face substantial challenges—most notably, the risk of overfitting caused by the rapid growth in the number of parameters stemming from complex dynamic and static inter-dependencies among panel units. To ensure tractable estimation in such settings, it is essential to apply dimensionality reduction techniques.[2]

This paper proposes a novel approach that combines Bayesian shrinkage with adaptive variable selection to effectively tackle over-parameterization and sparsity common in high-dimensional panels. Since the seminal contribution of Park and Casella (2008), the Bayesian Lasso has become a widely used tool for regularization and variable selection in high-dimensional settings. Casella et al. (2010) extend this framework by offering a Bayesian interpretation of the fused Lasso through the penalization of successive differences. More recently, Ročková and George (2018) introduce the spike-and-slab Lasso (SSL), which enhances the Bayesian Lasso by explicitly combining shrinkage and variable selection. The SSL assigns each coefficient either to a tightly concentrated "spike" (representing near-zero values) or a more diffuse "slab" (representing non-negligible values), yielding superior performance in sparse linear regression problems. Building on these developments, we adapt the SSL to the panel VAR context, enabling effective estimation in high-dimensional macroeconomic panels with complex inter-dependencies and heterogeneous dynamics.

We generalize this framework to PVARs by developing the *Panel Spike-and-Slab Lasso* (PSSL) technique. This approach enables flexible estimation of both dynamic inter-dependencies and cross-sectional heterogeneity in high-dimensional panel systems. Specifically, we partition the parameter space into blocks associated with each unit and assign each block to one of two Laplace distributions: a tightly concentrated "spike" or a more diffuse "slab," following Ročková and George (2018) and Bai et al. (2022). This blockwise structure allows the model to adaptively distinguish between relevant and irrelevant coefficients, facilitating effective shrinkage

[1] Several studies examine multi-country econometric modeling: Kose et al. (2003) employ a multi-country factor model; Dees et al. (2007) and Feldkircher and Huber (2016) use global Vector Autoregression (VAR) models; and Canova and Ciccarelli (2009) adopt a multi-country VAR framework.

[2] Over-parameterization in PVAR models is commonly addressed through the Minnesota prior, introduced by Doan et al. (1984) and Litterman (1986). Jarociński (2010) extend this approach using hierarchical priors. More recent developments incorporate unit-specific restrictions and explicitly model heterogeneity and interdependence across panel units (Canova and Ciccarelli, 2009, 2013; Koop and Korobilis, 2016, 2019).

and variable selection across heterogeneous units.

Incorporating the spike-and-slab structure into the PVAR setting enhances the model's ability to perform selective shrinkage effectively. Through extensive Monte Carlo simulations spanning low-, moderate-, and high-dimensional settings with varying degrees of sparsity, we demonstrate that the proposed PSSL method outperforms the Stochastic Search Specification Selection (S4) approach of Koop and Korobilis (2016), which builds on the Stochastic Search Variable Selection (SSVS) framework of George et al. (2008). Although both PSSL and S4 rely on spike-and-slab priors, the key distinction lies in their prior specifications: S4 employs normal mixtures, while PSSL uses Laplace-based priors. This difference enables PSSL to impose stronger shrinkage on coefficients near zero while maintaining more moderate shrinkage on relevant signals. Additionally, by jointly estimating the scale parameters of the Laplace distributions, PSSL offers adaptive, data-driven regularization that flexibly adjusts to the underlying sparsity pattern in the data.

We demonstrate the practical value of the proposed PSSL framework through two illustrative applications in macro-financial analysis and forecasting. The first revisits the influential study by Koop and Korobilis (2016) on financial contagion in euro area sovereign bond markets. In contrast to earlier approaches that rely on pre-calibrated priors or fixed shrinkage structures, our method employs fully adaptive, data-driven regularization to uncover the structure of dynamic inter-dependencies and cross-country heterogeneity. By learning the pattern of spillovers directly from the data, PSSL provides a sharper and more interpretable representation of macro-financial transmission channels—particularly valuable in high-dimensional settings, where overfitting and model mis-specification pose significant challenges.

The results reveal several novel features of euro area financial dynamics. We find strong asymmetries in shock propagation: disturbances originating in core economies—such as France, Germany, and the Netherlands—tend to transmit widely to peripheral countries, while feedback from the periphery to the core is comparatively muted. However, this hierarchical pattern is not fixed. During episodes of heightened sovereign risk, spillovers can reverse direction, with core economies becoming recipients of financial stress.

In a second application, we assess the out-of-sample forecasting performance of PSSL using a panel of macroeconomic indicators across multiple countries. The model produces accurate and stable forecasts across various horizons and country conditions, consistently outperforming global shrinkage priors and conventional Bayesian alternatives. Unlike approaches such as the stochastic search specification selection (S4), which rely on fixed mixture structures and group-level pooling, PSSL employs adaptive, coefficient-specific regularization. This flexibility

helps avoid large forecast errors while preserving meaningful heterogeneity. Such robustness is particularly valuable in international macroeconomic forecasting, where structural uncertainty is high and model misspecification can lead to misleading aggregate inferences.

Overall, the method delivers competitive performance across diverse macroeconomic environments. Its adaptive, data-driven shrinkage makes it particularly well-suited to the estimation and forecasting challenges inherent in high-dimensional, heterogeneous panel systems.

Related literature on PVAR estimation broadly distinguishes between two types of methods for achieving shrinkage and/or sparsity. The first type comprises Bayesian approaches, such as Bayesian factor models (Canova and Ciccarelli, 2004, 2013; Koop and Korobilis, 2019; Huber et al., 2023) and Bayesian specification selection methods (Koop and Korobilis, 2016; Korobilis, 2016). These approaches are valued for their ability to incorporate shrinkage while flexibly imposing model restrictions. However, despite their popularity, Bayesian factor models may have limited capacity to enforce true sparsity. In particular, they often struggle to shrink coefficients all the way to zero, which can be problematic in settings where some effects are genuinely absent.

In parallel, there is a growing literature on frequentist penalization methods grounded in Lasso-type estimation, building on foundational contributions by Tibshirani (1996), Zhao and Yu (2006), Zou (2006), Yuan and Lin (2006), and Huang et al. (2008). These techniques rely on ℓ_1-norm regularized least squares and have emerged as strong alternatives to Bayesian approaches for high-dimensional VAR estimation. Related contributions, such as Song and Bickel (2011) and Nicholson et al. (2020), apply Bayesian Lasso methods to VAR models, introducing penalty structures that account for time series dependence and group-level sparsity. Together, these studies highlight the potential of Bayesian Lasso techniques to improve estimation efficiency and forecast performance in high-dimensional time series models.

The remainder of the paper is structured as follows. Section 2 introduces our Panel Spike-and-Slab Lasso (PSSL) estimator, which incorporates the two forms of shrinkage described above. Section 3 presents Monte Carlo simulations evaluating the estimator's ability to select relevant restrictions and produce accurate estimates in PVAR settings. Section 4 applies the PSSL to study financial contagion in euro area sovereign bond markets and to perform macroeconomic forecasting. Finally, Section 5 concludes with a summary of key findings and implications.

2 Framework

2.1 Model

We consider a Panel Vector Autoregression (PVAR) model with N cross-sectional units (e.g., countries), each observed over T periods. For each unit $i = 1, \ldots, N$ and time $t = 1, \ldots, T$, let y_{it} be a $G \times 1$ vector of endogenous variables. The PVAR model's equation for unit i is specified as follows:

$$y_{it} = \sum_{j=1}^{N} \sum_{p=1}^{P} A_p^{ij} y_{j,t-p} + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathcal{N}(0, \Sigma_{ii}), \tag{2.1}$$

where $A_p^{ij} \in \mathbb{R}^{G \times G}$ is the matrix of coefficients capturing the effect of the p-th lag of unit j's endogenous variables on those of unit i, and ε_{it} denotes the vector of error terms with unit-specific covariance matrix $\Sigma_{ii} \in \mathbb{R}^{G \times G}$. This specification allows for dynamic interactions both within and across units.

Letting $Y_t = (y_{1t}', \ldots, y_{Nt}')' \in \mathbb{R}^{NG \times 1}$ denote the stacked vector of all units' endogenous variables, the system of PVAR equations can be written in compact form as:

$$Y_t = \mathbf{A}_1 Y_{t-1} + \mathbf{A}_2 Y_{t-2} + \cdots + \mathbf{A}_P Y_{t-P} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma), \tag{2.2}$$

where each $\mathbf{A}_p \in \mathbb{R}^{NG \times NG}$ is a block matrix that collects the coefficient matrices A_p^{ij} for all unit pairs (i, j), and $\varepsilon_t = (\varepsilon_{1t}', \ldots, \varepsilon_{Nt}')' \in \mathbb{R}^{NG \times 1}$ is the stacked vector of error terms. The covariance matrix $\Sigma \in \mathbb{R}^{NG \times NG}$ captures both the within-unit covariances Σ_{ii} and the cross-unit covariances $\Sigma_{ij} = \mathrm{Cov}(\varepsilon_{it}, \varepsilon_{jt})$, defined in a blockwise manner. Equation (2.2) can be further expressed in vectorized form as:

$$Y_t = Z_t \mathbf{A} + \varepsilon_t, \tag{2.3}$$

where $Z_t = I_{NG} \otimes X_t'$, with $X_t = \left(I, Y_{t-1}', \ldots, Y_{t-P}'\right)'$, and $\mathbf{A} = (\mathrm{vec}(A_1)', \ldots, \mathrm{vec}(A_P)')' \in \mathbb{R}^{K \times 1}$ is the stacked vector of all PVAR parameters, with $K = PN^2G^2$. The number of parameters grows rapidly with the number of units, variables, and lags, leading to substantial dimensional complexity. Without any restrictions, the model in equation (2.3) is referred to as the *unrestricted PVAR*, comprising PN^2G^2 parameters for the autoregressive coefficients and $\frac{NG(NG+1)}{2}$ parameters in the symmetric covariance matrix Σ.

To improve tractability and interpretability, researchers have introduced restrictions that reflect two key structural features of the PVAR: dynamic interdependencies (DI) and cross-

sectional heterogeneity (CSH), as emphasized in Canova and Ciccarelli (2013) and Koop and Korobilis (2016). Dynamic interdependencies refer to the possibility that lagged endogenous variables from one unit influence the current outcomes of another. These interactions are encoded in the off-diagonal blocks A_p^{ij}, where $i \neq j$. Cross-sectional heterogeneity, on the other hand, reflects the idea that the own-lag dynamics of each unit, represented by the diagonal blocks A_p^{ii}, may differ across units.

While some studies impose restrictions on *static interdependencies* (SI)—which refer to contemporaneous correlations in the innovations across units, typically captured by the off-diagonal blocks of the covariance matrix Σ—our focus in this paper is solely on the dynamic propagation of shocks and unit-level heterogeneity. Accordingly, we do not impose or estimate SI restrictions, as doing so would require additional structural assumptions that are beyond the scope of this paper. Instead, we allow for contemporaneous dependence flexibly through a fully unrestricted error covariance matrix.

To make these structures explicit, it is useful to represent the full lag-p coefficient matrix as a block matrix:

$$A_p = \begin{bmatrix} A_p^{11} & A_p^{12} & \cdots & A_p^{1N} \\ A_p^{21} & A_p^{22} & \cdots & A_p^{2N} \\ \vdots & \vdots & \ddots & \vdots \\ A_p^{N1} & A_p^{N2} & \cdots & A_p^{NN} \end{bmatrix}, \tag{2.4}$$

where each submatrix A_p^{ij} describes the impact of the p-th lag of unit j's endogenous variables on those of unit i. The diagonal blocks correspond to own-lag effects (and reflect CSH when these differ across i), while the off-diagonal blocks represent dynamic linkages across units (DI).

In practice, it is often reasonable to impose sparsity on the off-diagonal blocks A_p^{ij} for $j \neq i$, since not all units are likely to be directly influenced by all others. For a given unit i, there are $N-1$ such potential DI restrictions at each lag. Across all units, the model admits up to $N(N-1)$ dynamic interdependency restrictions. Similarly, in some empirical settings, it may be appropriate to constrain own-lag structures to be homogeneous across units—for example, by imposing $A_p^{ii} = A_p^{jj}$ for $i \neq j$. The number of such non-cross-sectional heterogeneity restrictions is bounded above by $\frac{N(N-1)}{2}$.

This structured decomposition of the coefficient space allows for flexible modeling of both cross-unit spillovers and unit-specific dynamics, while also enabling the imposition of economically meaningful sparsity or homogeneity restrictions. Such structure is particularly valuable in high-dimensional macroeconomic applications, where parsimony and interpretability are essential for reliable estimation and inference.

2.2 PSSL Prior Specification

Following the definition $\mathbf{A} = (\text{vec}(A_1)', \ldots, \text{vec}(A_P)')' \in \mathbb{R}^{K\times 1}$, where each A_p is an $NG \times NG$ coefficient matrix at lag p and $K = PN^2G^2$, the vector $\text{vec}(A_p)$ stacks the group-level blocks $(A_p^{i1\prime}, \ldots, A_p^{iN\prime})'$ for each lag. These blocks capture both dynamic inter-dependencies (DI) across units and cross-sectional heterogeneity (CSH) in own-lag dynamics, as discussed above.

To flexibly accommodate both types of structural sparsity, we consider a Panel Spike-and-Slab Lasso (PSSL) prior setting. This framework imposes targeted shrinkage at the group level by applying distinct priors to off-diagonal and diagonal coefficient blocks. The spike-and-slab design distinguishes between coefficients likely to be irrelevant (shrunk toward zero) and those expected to be substantively different from zero. This distinction is governed by binary latent variables that probabilistically switch between two regimes: a "spike" component that strongly pulls parameters toward zero, and a "slab" component that allows for more diffuse, less-penalized deviations from zero.

Dynamic Interdependency (DI) Prior. To model potential DI across units, we introduce latent binary indicators $\gamma_{p,ij}^{DI} \in (0, 1)$ for each off-diagonal coefficient block $\left\{\mathbf{A}_p^{ij}\right\}$ with $i \neq j$. These indicators determine whether a given block is included in the model ($\gamma_{p,ij}^{DI} = 1$: presence of DI) or excluded ($\gamma_{p,ij}^{DI} = 0$: no DI). Each $\gamma_{p,ij}^{DI}$ follows a Bernoulli distribution with prior:

$$\gamma_{p,ij}^{DI} \sim \text{Bernoulli}(\theta^{DI}). \tag{2.5}$$

Given $\gamma_{p,ij}^{DI}$, we assign a group-level spike-and-slab prior to the coefficient block $\mathbf{A}_p^{ij}$, modeled as a mixture of two double-exponential (DE) distributions:[3]

$$\pi(\mathbf{A}_p^{ij} \mid \gamma_{p,ij}^{DI}) = (1 - \gamma_{p,ij}^{DI})\, \Phi_0\left(\mathbf{A}_p^{ij} \mid \lambda_0\right) + \gamma_{p,ij}^{DI}\, \Phi_1\left(\mathbf{A}_p^{ij} \mid \lambda_1\right), \tag{2.6}$$

where Φ_0 and Φ_1 are double-exponential (DE) densities with different scales. This prior structure can be written hierarchically as:

$$\mathbf{A}_p^{ij} \mid \tau_{k,ij}^2, \gamma_{p,ij}^{DI} = DE\left(0, \tau_{k,ij}^2\right), \quad \text{if } \gamma_{p,ij}^{DI} = k, \tag{2.7}$$

$$\tau_{k,ij}^2 \mid \lambda_k^2 \sim \exp\left(\frac{\lambda_k^2}{2}\right), \quad k \in \{0, 1\} \tag{2.8}$$

[3]Double-exponential, also known as Laplace.

The DE densities $\mathbf{\Phi}_k\left(\mathbf{A}_p^{ij} \mid \lambda_k\right)$ correspond to:

$$\Phi_k(\mathbf{A}_p^{ij} \mid \lambda_k) = \frac{\lambda_k}{2}\exp(-\lambda_k\|\mathbf{A}_p^{ij}\|_1), \quad k \in \{0,1\}, \quad \lambda_0 > \lambda_1 > 0.$$

The "spike" component Φ_0 has a sharp peak and heavy penalization, inducing strong shrinkage toward zero and effectively eliminating insignificant cross-unit interactions. In contrast, the "slab" component Φ_1 is more diffuse, allowing relevant coefficients to remain relatively unrestricted.

Inference on dynamic interdependencies is based on the posterior inclusion probability,

$$\mathbb{P}\left(\gamma_{p,ij}^{DI} = 1 \mid Y_1, \ldots, Y_N\right) = \int \mathbb{P}\left(\gamma_{p,ij}^{DI} = 1 \mid \mathbf{A}_p^{ij}\right)\, \pi\left(\mathbf{A}_p^{ij} \mid Y_1, \ldots, Y_N\right)\, d\mathbf{A}_p^{ij}. \tag{2.9}$$

which provides a data-driven measure of support for the presence of a dynamic interdependency between unit i and j at lag p.

Cross-Sectional Heterogeneity (CSH) Prior. For diagonal blocks $\left\{\mathbf{A}_p^{ii}\right\}$, which represent the own-lag dynamics of each unit, we distinguish between two forms of cross-sectional shrinkage.

(a) Group Average Shrinkage (GAS): Here, deviations of each unit's own coefficients from a group average $\bar{\mathbf{A}}$ are modeled via:

$$\pi\left(\mathbf{A}_p^{ii} - \bar{\mathbf{A}} \mid \gamma_{p,ij}^{CSH}\right) = (1 - \gamma_{p,ij}^{CSH})\,\Phi_0\left(\cdot \mid \eta_0\right) + \gamma_{p,ij}^{CSH}\,\Phi_1\left(\cdot \mid \eta_1\right), \tag{2.10}$$

with $\eta_0 > \eta_1 > 0$, and:

$$\mathbf{A}_p^{ii} - \bar{\mathbf{A}} \sim DE\left(0, \kappa_{k,ij}^2\right), \quad \text{if } \gamma_{p,ij}^{CSH} = k, \tag{2.11}$$

$$\kappa_{k,ij}^2 \sim \exp\left(\frac{\eta_k^2}{2}\right), \quad k \in \{0,1\}, \tag{2.12}$$

$$\gamma_{p,ij}^{CSH} \sim \text{Bernoulli}(\theta^{CSH}). \tag{2.13}$$

The spike component Φ_0 enforces cross-sectional homogeneity (no CSH) by shrinking deviations from the average toward zero, while the slab allows heterogeneity (CSH) when supported by the data.

(b) Group Value Shrinkage (GVS): Alternatively, the prior may directly penalize pairwise

deviations between diagonal blocks:

$$\pi\left(\mathbf{A}_p^{ii} - \mathbf{A}_p^{jj} \mid \gamma_{p,ij}^{CSH}\right) = (1 - \gamma_{p,ij}^{CSH})\,\Phi_0\left(\cdot \mid \eta_0\right) + \gamma_{p,ij}^{CSH}\,\Phi_1\left(\cdot \mid \eta_1\right), \tag{2.14}$$

again enabling endogenous clustering of units with similar own-lag dynamics.

To impose cross-sectional homogeneity restrictions in a coherent and tractable way, we follow Koop and Korobilis (2016) and define a global selection matrix Γ, constructed as

$$\Gamma = \prod_{i=1}^{N-1} \prod_{j=i+1}^{N} \Gamma_{ij}, \tag{2.15}$$

where each component matrix Γ_{ij} encodes a pairwise no CSH restriction between the diagonal coefficient blocks $\mathbf{A}_p^{ii}$ and $\mathbf{A}_p^{jj}$, governed by the binary selector $\gamma_{p,ij}^{CSH}$. If $\gamma_{p,ij}^{\text{CSH}} = 0$, the prior promotes homogeneity by shrinking $\mathbf{A}_p^{ii}$ and $\mathbf{A}_p^{jj}$ toward each other; in contrast, if $\gamma_{p,ij}^{\text{CSH}} = 1$, it allows for heterogeneity. The product form ensures that all pairwise restrictions are jointly captured and propagated throughout estimation. An illustrative example of Γ_{ij} and its role in the full matrix Γ is provided in Appendix A.2.

Joint Prior on Parameters. Bringing the two components together, the prior over the coefficient vector $\mathbf{A}$ is specified as:

$$\pi(\mathbf{A} \mid \gamma) = \prod_{i \neq j}^{R} \prod_{p=1}^{P} \pi\left(\mathbf{A}_p^{ij} \mid \gamma_{p,ij}^{DI}\right) \prod_{i=j}^{C} \prod_{p=1}^{P} \pi\left(\mathbf{A}_p^{ii} - \bar{\mathbf{A}} \mid \gamma_{p,ij}^{CSH}\right), \tag{2.16}$$

where R and C are the number of off-diagonal and diagonal coefficient blocks, respectively.

To complete the model, we impose an improper Jeffreys prior on the error covariance matrix, $\pi(\Sigma^{-1}) \propto |\Sigma|^{-(NG+1)/2}$, yielding the joint prior:

$$\begin{aligned} \pi(\mathbf{A}, \Sigma) = \pi(\Sigma) \prod_{i \neq j}^{R} \prod_{p=1}^{P} \left[(1 - \gamma_{p,ij}^{DI})\Phi_0(\cdot \mid \lambda_0) + \gamma_{p,ij}^{DI}\Phi_1(\cdot \mid \lambda_1)\right] \\ \times \prod_{i=j}^{C} \prod_{p=1}^{P} \left[(1 - \gamma_{p,ij}^{CSH})\Phi_0(\cdot \mid \eta_0) + \gamma_{p,ij}^{CSH}\Phi_1(\cdot \mid \eta_1)\right]. \end{aligned} \tag{2.17}$$

This PSSL prior specification enables scalable and interpretable Bayesian inference in high-dimensional panel VARs by adaptively selecting which interactions matter, which units behave similarly, and where parsimony can be imposed without sacrificing flexibility.

2.2.1 Choice of the Penalty

Figure 2.1 presents a directed acyclic graph (DAG) that illustrates the hierarchical structure of the Bayesian Panel Spike-and-Slab Lasso (PSSL) prior described in the previous section. It visually outlines the flow of information among the model parameters, selection indicators, local scaling parameters, and global penalty hyperparameters. Each arrow in the diagram corresponds to a conditional dependency in the hierarchical prior. The latent selection indicators γ^{DI} and γ^{CSH} determine whether a given coefficient block belongs to the spike or the slab component, with associated local scale parameters τ and κ. These, in turn, are governed by global shrinkage parameters λ and η, which modulate the overall degree of regularization in the prior distribution for the PVAR coefficients.

To estimate the penalty parameters λ_0 and η_0 associated with the spike components, we employ a Monte Carlo Expectation-Maximization (MCEM) algorithm, following Park and Casella (2008). At each iteration k, the Gibbs sampler is used to draw from the conditional posteriors of all parameters, given the current values of $\lambda_0^{(k-1)}$ and $\eta_0^{(k-1)}$. These posterior samples are then used to update the penalty parameters in the E-step via the following expressions:

$$\lambda_0^{(k)} = \sqrt{\frac{2r}{\sum_{p=1}^{P}\sum_{i\neq j}^{R}\mathbb{E}_{\lambda_0^{(k-1)}}\left[\tau_{0,ij}^2 \mid \tilde{Y}_{ij}\right]}}, \quad \eta_0^{(k)} = \sqrt{\frac{2c}{\sum_{p=1}^{P}\sum_{i=j}^{C}\mathbb{E}_{\eta_0^{(k-1)}}\left[\kappa_{0,ij}^2 \mid \tilde{Y}_{ij}\right]}},$$

where $r = N(N-1)$ and $c = N(N-1)/2$ represent the total number of dynamic interdependency (DI) and cross-sectional heterogeneity (CSH) blocks respectively. The expectations on the right-hand side of each equality are approximated by posterior averages of the local scale parameters $\tau_{0,ij}^2$ and $\kappa_{0,ij}^2$ from the Gibbs draws.

We initialize the spike penalty parameters with conservative values, setting $\lambda_0^{(0)} = \eta_0^{(0)} = 1$, following Park and Casella (2008). This choice guards against overly aggressive shrinkage in the early iterations and allows the slab component—designed to capture large coefficients—to dominate when appropriate. This is especially important in high-dimensional settings, where true signals might otherwise be prematurely pulled toward zero.

For the slab components, we fix λ_1 and η_1 to small values, as suggested by Bai et al. (2021), ensuring that large coefficients remain relatively unpenalized and can escape the influence of the spike. This specification allows the model to distinguish between truly negligible and substantively important effects.

To improve stability during early MCEM iterations, we temporarily fix the residual covariance matrix Σ to the OLS estimator $\widehat{\Sigma}^{OLS}$. This prevents erratic updates of Σ that could amplify early shrinkage before the spike penalties are reliably estimated. Once λ_0 and η_0 have

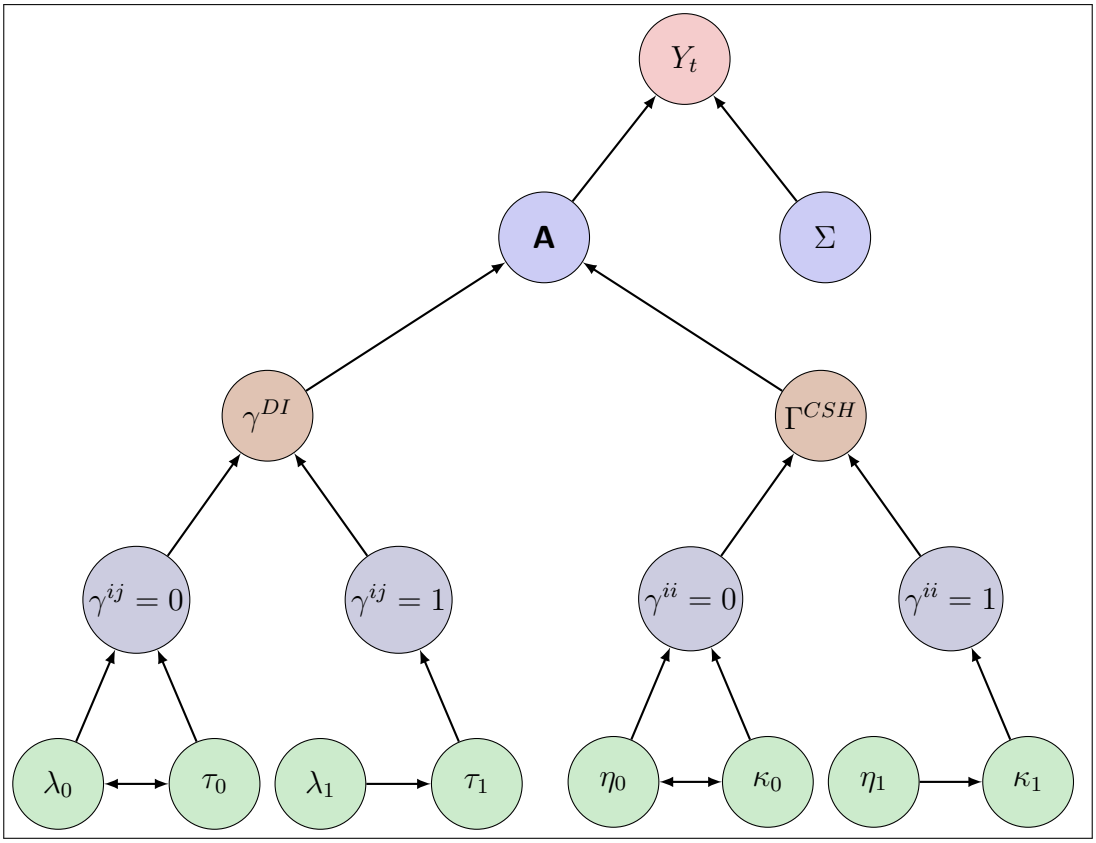


Figure 2.1: Panel Spike-and-Slab Lasso model for inference on PVARs

stabilized, we allow Σ to be updated within the Gibbs sampling step using its full conditional posterior.

Overall, this MCEM-based procedure allows the global shrinkage parameters to adapt to the data in a principled way while preserving the interpretability inherent in spike-and-slab variable selection. Full details of the conditional posterior distributions and the Gibbs sampling algorithm used in this procedure are provided in Appendix B–C.

2.3 Adaptive Shrinkage in PSSL

Ročková and George (2018) demonstrate that any sparsity-inducing Bayesian prior $\pi(\mathbf{A})$ can be equivalently expressed within a penalized likelihood framework, where the log-marginal prior $\log \pi(\mathbf{A})$ functions as a penalty term. This equivalence provides a unifying perspective bridging Bayesian and regularization approaches. Building on this insight and following Camehl (2023), we leverage this duality to represent the posterior mode of the PVAR model under the PSSL prior as the solution to the following penalized likelihood optimization problem:

$$\underset{\mathbf{A}}{\operatorname{argmin}} \left\{ \frac{1}{T} \operatorname{tr}\left[(Y_t - Z_t\mathbf{A})'\Omega(Y_t - Z_t\mathbf{A}) \right] + \operatorname{pen}(\Lambda, \mathbf{A}, \bar{\mathbf{A}}) \right\}, \tag{2.18}$$

where $\Omega = \Sigma^{-1}$ is the precision matrix and $\text{tr}(\cdot)$ denotes the trace. The coefficient vector $\mathbf{A} = (\text{vec}(A_1)', \text{vec}(A_2)', \ldots, \text{vec}(A_P)')'$ stacks all VAR coefficient matrices across lags. For each lag p, the matrix A_p is structured into block components $\mathbf{A}_p^{ij}$ capturing the effect of variable group j on group i.

The penalty function $\text{pen}(\Lambda, \mathbf{A}, \bar{\mathbf{A}})$ encodes prior shrinkage and captures two essential features: (i) *Dynamic Interdependencies (DI)* across cross-sectional units, where off-diagonal elements $\mathbf{A}_p^{ij}$ for $i \neq j$ are shrunk toward zero; (ii) *Cross-Sectional Heterogeneity (CSH)* among own-lag coefficients $\mathbf{A}_p^{ii}$, which are either shrunk toward a group mean $\bar{\mathbf{A}}$ or a cross-sectional pair $\mathbf{A}_p^{jj}$.

2.3.1 DI Penalty and Posterior Inclusion Probability

The DI penalty promotes sparsity in the off-diagonal blocks, which capture interdependencies across units.Following Ročková and George (2018), we define the DI penalty as the logarithm of the ratio between the conditional prior density evaluated at $\mathbf{A}$ and its value at zero:

$$\text{pen}_{DI}(\mathbf{A}|\gamma^{DI}) = \log\left(\frac{\pi(\mathbf{A}|\gamma^{DI})}{\pi(\mathbf{0}|\gamma^{DI})}\right), \tag{2.19}$$

which is centered such that $\text{pen}_{DI}(\mathbf{0} \mid \gamma^{DI}) = 0$.

Due to the conditional independence across blocks $\mathbf{A}_p^{ij}$, the penalty decomposes into a sum of individual terms:

$$\text{pen}_{DI}(\mathbf{A}_p^{ij}|\gamma_{ij}^{DI}) = -\lambda_1|\mathbf{A}_p^{ij}| + \log\left(\frac{\mathbb{P}_{\gamma_{ij}^{DI}}(\mathbf{0})}{\mathbb{P}_{\gamma_{ij}^{DI}}(\mathbf{A}_p^{ij})}\right), \tag{2.20}$$

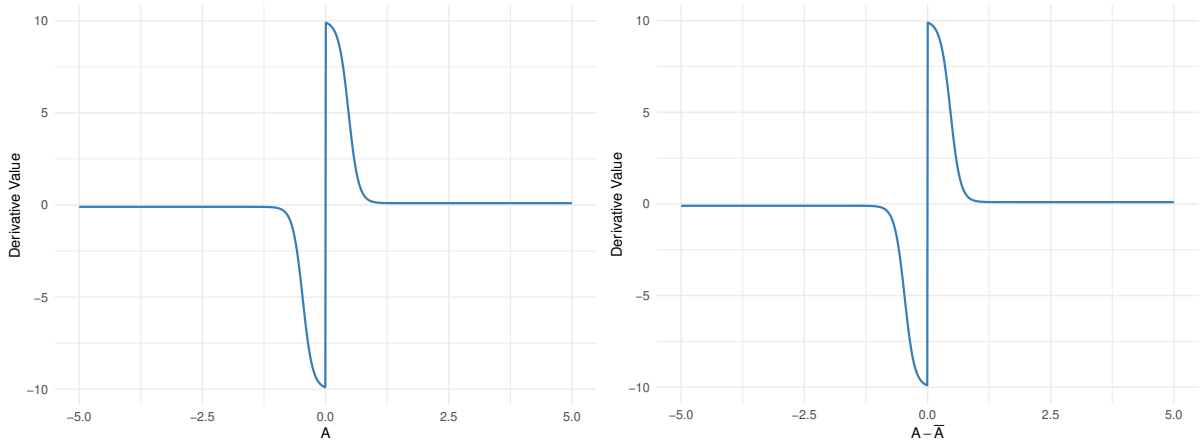
where the first term is the baseline ℓ_1-shrinkage and the second reflects adaptive penalization via the posterior inclusion probability. The total DI penalty is:

$$\text{pen}_{DI}(\mathbf{A}|\gamma^{DI}) = -\lambda_1|\mathbf{A}| + \sum_{i\neq j}^{R}\sum_{p=1}^{P}\log\left(\frac{\mathbb{P}_{\gamma_{ij}^{DI}}(\mathbf{0})}{\mathbb{P}_{\gamma_{ij}^{DI}}(\mathbf{A}_p^{ij})}\right), \tag{2.21}$$

where $|\mathbf{A}| = \sum_{i\neq j}^{R}\sum_{p=1}^{P}|\mathbf{A}_p^{ij}|$[4]. The posterior inclusion probability is given by:

$$\mathbb{P}_{\gamma_{ij}^{DI}}(\mathbf{A}_p^{ij}) = \frac{\gamma_{ij}^{DI}\Phi_1(\mathbf{A}_p^{ij})}{(1-\gamma_{ij}^{DI})\Phi_0(\mathbf{A}_p^{ij}) + \gamma_{ij}^{DI}\Phi_1(\mathbf{A}_p^{ij})}. \tag{2.22}$$

[4]The complete derivation is provided in Appendix A.3.1.



(a) Derivative of $pen_{DI}(A)$ (b) Derivative of $pen_{CSH}(A)$

Figure 2.2: Plot of the derivative of the penalty functions

Using Laplace densities for spike Φ_0 and slab Φ_1, this simplifies to:

$$\mathbb{P}_{\gamma_{ij}^{DI}}(\mathbf{A}_p^{ij}) = \left[1 + \frac{\lambda_0}{\lambda_1}\left(\frac{1-\gamma_{ij}^{DI}}{\gamma_{ij}^{DI}}\right)\exp(-|\mathbf{A}_p^{ij}|(\lambda_0 - \lambda_1))\right]^{-1}. \tag{2.23}$$

This expression reflects two key forms of adaptivity. First, as $|\mathbf{A}_p^{ij}|$ increases, the exponential term decays, driving the inclusion probability toward one. This shifts the coefficient into the slab component, thereby reducing shrinkage. Second, the difference between λ_0 and λ_1 governs the sharpness of the spike–slab contrast: a larger gap leads to more decisive transitions. As a result, large coefficients are more likely to escape shrinkage, while small ones are subject to stronger regularization.

The partial derivative of the DI penalty with respect to $|\mathbf{A}_p^{ij}|$ is:

$$\frac{\partial \text{pen}_{DI}(\mathbf{A}|\gamma^{DI})}{\partial|\mathbf{A}_p^{ij}|} = -\lambda_\gamma^*(\mathbf{A}_p^{ij}), \tag{2.24}$$

where

$$\lambda_\gamma^*(\mathbf{A}_p^{ij}) = \lambda_1 w(\mathbf{A}_p^{ij}) + \lambda_0(1 - w(\mathbf{A}_p^{ij}))$$

represents the weighted average of λ_0 and λ_1 with weight $w(\mathbf{A}_p^{ij})$, indicating the degree of shrinkage[5].

Figure 2.2a shows the shrinkage pattern implied by Equation (2.24). The gradient begins steep near zero, indicating aggressive shrinkage for small coefficients, and flattens as $|\mathbf{A}_p^{ij}|$ increases, thereby selectively preserving significant dependencies.

[5]Detailed derivations for Equation 2.23 and Equation 2.24 are provided in Appendix A.3.2 and Appendix A.4.

2.3.2 CSH Penalty and Posterior Inclusion Probability

The cross-sectional homogeneity penalty (no CSH) aims to exploit structural similarities across units in the same group. In particular, we assume that the diagonal blocks $\mathbf{A}_p^{ii}$ for units $i \in \{1, \ldots, N\}$ are similar and can be shrunk toward a group average matrix $\bar{\mathbf{A}}_p$. For each unit, deviations $\Delta\mathbf{A}_p^{ii} = \mathbf{A}_p^{ii} - \bar{\mathbf{A}}_p$ are penalized.

Following the same spike-and-slab logic, the CSH penalty for each diagonal block is:

$$\text{pen}_{CSH}(\Delta\mathbf{A}_p^{ii}|\gamma_{ii}^{CSH}) = -\eta_1|\Delta\mathbf{A}_p^{ii}| + \log\left(\frac{\mathbb{P}_{\gamma_{ii}^{CSH}}(\mathbf{0})}{\mathbb{P}_{\gamma_{ii}^{CSH}}(\Delta\mathbf{A}_p^{ii})}\right), \tag{2.25}$$

which parallels the DI formulation.

The posterior inclusion probability under CSH shrinkage is:

$$\mathbb{P}_{\gamma_{ii}^{CSH}}(\Delta\mathbf{A}_p^{ii}) = \left[1 + \frac{\eta_0}{\eta_1}\left(\frac{1-\gamma_{ii}^{CSH}}{\gamma_{ii}^{CSH}}\right)\exp\left(-|\Delta\mathbf{A}_p^{ii}|(\eta_0 - \eta_1)\right)\right]^{-1}. \tag{2.26}$$

This probability behaves similarly to its DI counterpart. For units with strong deviations from the group trend (large $|\Delta\mathbf{A}_p^{ii}|$), the inclusion probability is high, reducing shrinkage. Conversely, smaller deviations are aggressively shrunk to zero. This adaptively enforces homogeneity across diagonal blocks without rigidly forcing equality.

The corresponding partial derivative of the CSH penalty with respect to $|\Delta\mathbf{A}_p^{ii}|$ is:[6]

$$\frac{\partial\text{pen}_{CSH}(\Delta\mathbf{A}|\gamma^{CSH})}{\partial|\Delta\mathbf{A}_p^{ii}|} = -\eta_\gamma^*(\Delta\mathbf{A}_p^{ii}) = -\left[\eta_1 w(\Delta\mathbf{A}_p^{ii}) + \eta_0(1 - w(\Delta\mathbf{A}_p^{ii}))\right]. \tag{2.27}$$

As shown in Figure 2.2b, this penalty generates a smooth transition from high shrinkage (near the origin) to low shrinkage (farther from the group mean), balancing structure and flexibility.

3 Monte Carlo Experiments

We conduct a series of Monte Carlo simulations to evaluate the performance of the proposed PSSL method. The first three data-generating processes (DGPs) are designed to test the model's ability to correctly identify the underlying structural features of the panel VAR system. In these controlled settings, the true restrictions—dynamic interdependencies (DI), cross-sectional heterogeneity (CSH), or both—are known by construction. This allows us to assess whether the

[6]The full derivation of the CSH penalty term and its derivative is provided in Appendix A.3.2 and Appendix A.4.

PSSL approach can accurately recover the true structure and to examine how the imposition of incorrect restrictions affects estimation accuracy and forecasting performance.

Each DGP is evaluated under low-, moderate-, and high-dimensional settings, defined by increasing levels of sparsity in the coefficient matrix $\mathbf{A}$. Specifically, we vary the cross-sectional dimension with $N = 3$, 5, and 10, while keeping the time dimension fixed at $T = 100$. The number of variables per unit is set to $G = 2$, and we include a single lag ($P = 1$) in the VAR specification. These configurations allow us to assess how model performance scales with dimensionality.

To benchmark the performance of the PSSL approach, we compare it against several well-established alternatives: the *Stochastic Search Variable Selection* (S4) method of Koop and Korobilis (2016), the Bayesian Lasso (BL), and the Bayesian Fused Lasso (BFL). These methods represent three prominent shrinkage paradigms in Bayesian sparse modeling: (i) a mixture normal prior (S4), (ii) a single Laplace prior (BL and BFL), and (iii) a mixture Laplace prior (PSSL)[7].

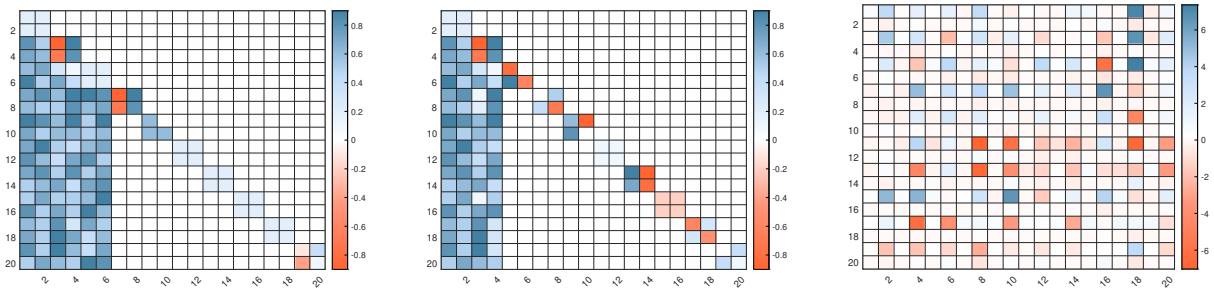
As discussed in Koop and Korobilis (2016), S4 employs a mixture normal distribution with a calibrated shrinkage factor, typically resulting in mild and smooth shrinkage. However, this leads to what is referred to as soft sparsity, wherein both noise and weak signals are likely to be retained due to the distribution's light tails. In contrast, the Laplace prior used in BL and BFL imposes stronger shrinkage around zero, more aggressively promoting sparsity, while its heavy tails allow large coefficients to escape shrinkage. BFL further extends this framework by incorporating a fused penalty that encourages similarity among adjacent coefficients, thereby promoting local smoothness or block-wise structure.[8]

We also include a frequentist benchmark: the penalized panel VAR estimator proposed by Camehl (2023). Simulation results across different dimensional settings—$(N, G, T) = (3, 2, 100)$, $(5, 2, 100)$, and $(10, 2, 100)$—are reported in Section 3.2.

In addition to the three controlled DGPs, we include a fourth simulation designed to mimic the structure of real-world data rather than relying on a stylized, pre-specified model. Specifically, we generate pseudo-data calibrated to the empirical setting of Camehl (2023), with the goal of capturing the types of dependencies and heterogeneity typically encountered in practice. This exercise allows us to assess whether the proposed method remains effective in an applied

[7]While the SI (static interdependency) restriction is important for structural analysis and has been emphasized by Canova and Ciccarelli (2009) and Koop and Korobilis (2016), we do not impose it here. Our focus is on forecasting rather than structural inference, and we follow the forecasting-oriented framework of Korobilis (2016)

[8]The "fused" component in BFL adds a penalty on the differences between $\sum |\beta_j - \beta_{j-1}|$, in addition to the standard Lasso penalty on the coefficients themselves. This contrasts with the standard Bayesian Lasso (BL), which only penalizes the absolute values of the coefficients, encouraging sparsity but not smoothness across neighboring effects.



(a) DGP 1 (b) DGP 2 (c) DGP 3

Figure 3.1: Autoregressive Coefficients of the Simulated DGPs

Note: The figure shows autoregressive coefficients of the simulated DGPs. Zero coefficients are colored in white, and non-zero coefficients are in pink to blue shades, where the darkest shade corresponds to the highest coefficient value.

context, where the data-generating process is inherently more complex, reflecting the conditions under which researchers actually work.

3.1 Experiment Design

Figure 3.1 illustrates the specific autoregressive coefficient patterns used in each DGP, offering a visual summary of their structural designs.

DGP1 (DI and CSH Restrictions). DGP1 is designed to evaluate the model's ability to recover both dynamic interdependencies (DI) and cross-sectional heterogeneity (CSH). Units are arranged hierarchically: unit 1 influences all others but is not influenced in return; unit 2 affects those below it in the hierarchy, and so on. From unit 4 onward, all units are isolated and exert no influence on others.

This structure is reflected in the coefficient matrix, where off-diagonal entries in the lower triangular portion (for units 1–3) are nonzero, while all other off-diagonal elements are set to zero. These zeros guide the DI restriction search. Simultaneously, the diagonal blocks include both homogeneous and heterogeneous coefficients across units, necessitating a CSH restriction search. All nonzero coefficients are randomly drawn from the interval $[-0.9, 0.9]$, inducing heterogeneity in magnitudes and directions.

DGP2 (DI Restriction Only). DGP2 builds on DGP1 by eliminating cross-sectional homogeneity. All diagonal entries in the coefficient matrix are set to distinct values, i.e., $A_{1,1} \neq A_{2,2} \neq \cdots \neq A_{N,N}$, so that only DI remains to be identified. The true value of the CSH shrinkage parameter is $\gamma^{\text{CSH}} = 1$, and an effective estimation procedure should recover $\hat{\gamma}^{\text{CSH}}$ close to one. The off-diagonal structure is retained from DGP1, preserving the hierarchical influence pattern across countries.

DGP3 (CSH Restriction Only). DGP3 features a dense coefficient matrix with weak but

widespread dynamic inter-dependencies: most off-diagonal entries are nonzero but small in magnitude. This setup diminishes the need for a DI restriction search, as pervasive spillovers exist by design and should not be heavily shrunk.

In contrast, the diagonal blocks mirror those in DGP1: some units share identical own-dynamics, while others differ. This variation justifies a CSH restriction search. The true value of the DI shrinkage parameter is $\gamma^{\text{DI}} = 1$, so accurate estimation should yield $\hat{\gamma}^{\text{DI}}$ near one, reflecting minimal regularization of the off-diagonal elements.

3.2 Results

We evaluate the performance of the *PSSL* framework along three key dimensions—estimation accuracy, variable selection accuracy, and forecasting accuracy—using both the fully calibrated experimental data (DGP1–DGP3) and the pseudo-realistic simulated data designed to mimic the structure of actual macroeconomic data (DGP4).

3.2.1 Estimation Accuracy in DGP1–DGP3

We use the *root mean squared error* (RMSE) and the *mean absolute deviation* (MAD) to evaluate the accuracy of the estimated autoregressive coefficient matrix $\mathbf{A}$. In the context of our PSSL framework, these criteria are defined respectively as:

$$\text{RMSE} = \sqrt{\frac{1}{K}\sum_{p=1}^{P}\sum_{i,j=1}^{N}\sum_{l,k=1}^{G}\left(\tilde{\alpha}_{l,k}^{i,j} - \alpha_{l,k}^{i,j}\right)^2}, \tag{3.1}$$

$$\text{MAD} = \frac{1}{K}\sum_{p=1}^{P}\sum_{i,j=1}^{N}\sum_{l,k=1}^{G}\left|\tilde{\alpha}_{l,k}^{i,j} - \alpha_{l,k}^{i,j}\right|, \tag{3.2}$$

where $K = PN^2G^2$ denotes the total number of coefficients, $\tilde{\alpha}$ represents the estimated values, and α denotes the true coefficients.

Table 3.1 reports the RMSE and MAD across all three DGPs and various values of N. In addition to our proposed Group Value Shrinkage (GVS) and Group Average Shrinkage (GAS) methods, we also report the performance of the Bayesian Lasso (BL), Bayesian Fused Lasso (BFL), and Stochastic Search Variable Selection (S4) methods.The results are normalized relative to the performance of the GAS estimator; that is, values greater than one indicate worse performance compared to GAS. Overall, the findings provide clear evidence of the superior estimation accuracy of the GAS method across a range of settings, particularly in the presence of dynamic inter-dependencies (DI) and cross-sectional heterogeneity (CSH) restrictions.

Table 3.1: Relative Performance of Different Methods with respect to GAS Across DGPs

	Panel A: DGP 1						
	Criterion	**Proposed Methods**		**Other Bayesian Methods**			**Lasso**
		GAS	**GVS**	**BL**	**BFL**	**S4**	$\boldsymbol{\lambda, c, \alpha, \Sigma}$
$N = 3$	RMSE	1.00	3.33	1.48	1.95	2.23	1.48
	MAD	1.00	2.58	1.57	2.36	2.94	1.88
$N = 5$	RMSE	1.00	0.96	21.61	21.58	8.15	3.76
	MAD	1.00	0.98	18.63	20.00	7.60	2.76
$N = 10$	RMSE	1.00	3.34	13.83	13.82	6.11	2.04
	MAD	1.00	1.91	12.30	12.18	5.71	1.44
	Panel B: DGP 2						
	Criterion	**Proposed Methods**		**Other Bayesian Methods**			**Lasso**
		GAS	**GVS**	**BL**	**BFL**	**S4**	$\boldsymbol{\lambda, c, \alpha, \Sigma}$
$N = 3$	RMSE	1.00	7.63	6.09	5.49	4.19	3.69
	MAD	1.00	4.65	5.15	5.16	3.79	3.44
$N = 5$	RMSE	1.00	1.08	27.35	27.47	9.89	5.23
	MAD	1.00	1.04	24.82	24.93	9.04	4.23
$N = 10$	RMSE	1.00	2.95	10.83	9.88	4.89	1.97
	MAD	1.00	1.56	7.97	7.42	3.86	1.36
	Panel C: DGP 3						
	Criterion	**Proposed Methods**		**Other Bayesian Methods**			**Lasso**
		GAS	**GVS**	**BL**	**BFL**	**S4**	$\boldsymbol{\lambda, c, \alpha, \Sigma}$
$N = 3$	RMSE	1.00	1.04	0.47	0.45	1.67	3.83
	MAD	1.00	0.96	0.57	0.56	1.99	3.39
$N = 5$	RMSE	1.00	1.03	3.68	3.92	2.39	1.86
	MAD	1.00	0.94	5.13	5.48	3.61	2.87
$N = 10$	RMSE	1.00	10.00	2.02	5.13	5.80	5.84
	MAD	1.00	3.77	1.95	5.97	5.74	4.99

Note: Each panel reports RMSE and MAD for a given DGP across different values of N, normalized to the GAS model. A value below one indicates superior performance relative to GAS. GVS refers to group value shrinkage; BL and BFL are Bayesian shrinkage variants; S4 is the stochastic selection method; and Lasso is the frequentist counterpart. The Lasso setting $\lambda, c, \alpha, \Sigma$ refers to DI and CSH penalization also with a penalization in Σ.

In DGP1 and DGP2, where sparsity is a defining feature, GAS consistently outperforms all competing methods. Its advantage becomes especially pronounced as N increases ($N = 5$ and $N = 10$), where BL and BFL perform poorly due to their use of uniform, non-adaptive

shrinkage. Both methods lack the flexibility to selectively penalize coefficients according to underlying structural patterns, which often results in suboptimal estimation in sparse settings.

The S4 approach, which incorporates a stochastic search mechanism to explicitly target dynamic inter-dependencies (DI) and cross-sectional heterogeneity (CSH) structures, generally outperforms both BL and BFL. However, it still lags behind GAS—and, in most cases, GVS—in terms of estimation accuracy. This limitation likely arises from its sensitivity to prior scaling and the need for careful calibration, which can reduce its practical effectiveness.

A comparison between GAS and GVS underscores the importance of flexible shrinkage in high-dimensional settings. GVS imposes exact homogeneity by shrinking coefficients toward a common group value, which can be overly restrictive when there is moderate within-group variation. In contrast, GAS shrinks toward the group average, accommodating controlled heterogeneity within blocks. This flexibility proves particularly beneficial in partially homogeneous environments—such as DGP1 and DGP3—where allowing some variation across units leads to improved estimation accuracy.

Notably, in DGP3—which features a dense coefficient matrix with only weak sparsity—GAS remains competitive, despite the data-generating process deviating from the sparse benchmark for which it is better suited. This robustness highlights the adaptability of GAS: it delivers strong performance not only under sparse configurations but also in moderately dense environments, making it a reliable and versatile choice across a wide range of empirical settings.

3.2.2 Selection Accuracy in DGP1–DGP3

We use the *sensitivity* (SE), *specificity* (SP), and *accuracy* (AC) measures—commonly employed in the classification literature—to evaluate the performance of our proposed variable selection method. These metrics are computed based on the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). These selection measure are defined as:

$$\text{SE} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{SP} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad \text{AC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{3.3}$$

In this formulation, sensitivity (SE) measures the proportion of true nonzero coefficients correctly identified, specificity (SP) captures the proportion of true zeros accurately excluded, and accuracy (AC) reflects the overall proportion of correct classifications. High values of SE and SP—ideally approaching one—indicate strong variable selection performance.

Table 3.2 reports sensitivity (SE), specificity (SP), and overall classification accuracy (AC) for all models for $N = 10$. In DGP1 and DGP2, GAS and GVS consistently achieve the highest overall accuracy—approximately 64% and 59%, respectively. These results indicate a well-

Table 3.2: Selection Accuracy of Different Methods Across DGPs: $N = 10$

	DGP1			**DGP2**			**DGP3**		
	SE	**SP**	**AC**	**SE**	**SP**	**AC**	**SE**	**SP**	**AC**
Proposed Methods									
GAS	94.4%	49.3%	64.5%	85.4%	49.7%	59.0%	52.8%	53.0%	52.8%
GVS	94.2%	48.7%	64.1%	85.2%	50.5%	59.5%	52.5%	52.0%	52.5%
Other Bayesian Methods									
BL	86.3%	41.6%	56.7%	77.1%	45.8%	53.9%	52.7%	54.0%	52.7%
BFL	86.3%	41.2%	56.4%	80.0%	45.8%	54.7%	51.2%	14.0%	51.1%
S4	93.2%	47.9%	63.2%	84.3%	48.6%	57.9%	52.7%	43.0%	52.6%
Lasso									
$\lambda, c, \alpha, \Sigma$	40.8%	35.5%	37.3%	31.6%	44.2%	41.0%	49.6%	57.5%	49.6%

Note: The table presents SE, SP, and AC for different estimation methods across four DGPs, using N = 10. The best accuracy values for each DGP are highlighted in bold. BL and BFL are Bayesian shrinkage variants; S4 is the stochastic selection method; CC is the factor model without shrinkage; and Lasso is the frequentist counterpart. The Lasso setting $\lambda, c, \alpha, \Sigma$ refers to DI and CSH penalization also with a penalization in Σ.

calibrated trade-off between correctly identifying nonzero coefficients (high SE) and excluding zero coefficients (moderate SP). Their strong performance underscores the effectiveness of the spike-and-slab priors in capturing the key structural features of the coefficient matrix.

The S4 method ranks closely behind GAS and GVS, delivering high sensitivity but somewhat lower specificity, which modestly reduces its overall accuracy. In contrast, the Bayesian Lasso (BL), Bayesian Fused Lasso (BFL), and frequentist Lasso perform substantially worse. Their reliance on global, non-adaptive shrinkage results in systematic misclassification of both zero and nonzero coefficients, leading to poorer performance on both SE and SP.

Although the proposed method is particularly well-suited for sparse settings, its design is inherently flexible and data-driven, allowing it to adapt to a wide range of underlying structures. This is evident in DGP3, where the coefficient matrix is relatively dense with only weak sparsity. Even in this setting, both GAS and GVS maintain overall classification accuracy around 52–53%, performing on par with or better than competing methods. While accuracy is naturally lower than in DGP1 and DGP2—where the true structure is more sparse—the results confirm that the spike-and-slab approach retains its effectiveness. Its adaptive selection mechanism continues

to balance inclusion and exclusion of coefficients based on the observed data, demonstrating robustness and reliability across varying levels of sparsity[9] .

subsubsectionForecasting Accuracy in DGP1–DGP3

To evaluate the out-of-sample forecasting accuracy of each method, we partition the total sample of length T into a training set of size T_{train} and a hold-out test set of size $T_{\text{test}} = T - T_{\text{train}}$. We fix the test window to $T_{\text{test}} = 13$, implying $T_{\text{train}} = T - 13$ for all simulations.

For each Monte Carlo replication, we compute the h-step-ahead *mean squared forecast error* (MSFE) for variable l in unit i as:

$$\text{MSFE}_l^i(h) = \frac{1}{T_{\text{test}} - h + 1} \sum_{t=T_{\text{train}}}^{T-h} \left(\hat{Y}_{l,t+h}^i - Y_{l,t+h}^i \right)^2, \tag{3.4}$$

where $\hat{Y}_{l,t+h}^i$ denotes the h-step-ahead forecast made at time t using information available up to T_{train}, and $Y_{l,t+h}^i$ is the oberved value at $t + h$. Forecasts are generated recursively for each $t = T_{\text{train}}, \ldots, T - h$ and for forecast horizons $h = 1, \ldots, h_{\max}$. In our analysis, we set the maximum forecast horizon to $h_{\max} = 12$.

Tables 3.3 and 3.4 report the relative mean squared forecast errors (MSFE) for short horizons ($h = 1, 3$), medium horizon ($h = 6$), and long horizon ($h = 12$), normalized with respect to the GAS method. Since all values are reported relative to GAS, numbers below one indicate better forecasting performance compared to GAS.

At short forecast horizons ($h = 1, 3$), both GAS and GVS exhibit competitive forecast performance across most DGPs and panel sizes. In DGPs 1 and 2, corresponding to homogeneous and heterogeneous structures respectively, GVS remains competitive and in several cases outperforms GAS as the panel dimension increases. This suggests that, from a forecasting perspective, the stronger shrinkage imposed by GVS becomes more effective when more cross-sectional information is available, even in the presence of heterogeneity. In contrast, in DGP 3, which assumes cross-sectional homogeneity but involves a dense coefficient matrix, GVS tends to underperform as dimensionality increases, reflecting the limitations of over-shrinkage in settings with strong interdependencies across units.

At medium and long forecast horizons ($h = 6, 12$), the proposed methods—particularly GAS—continue to deliver stable and reliable forecast accuracy across a range of settings. While GAS does not always outperform all alternative methods, it remains robust under varying conditions. In sparse systems like DGP 1, it maintains strong accuracy, especially for small and medium panel sizes. In more complex or dense environments—such as DGP 3—alternative

[9]Additional results on the bias of the estimators are provided in Appendix D.1, and selection accuracy for $N = 3$ and $N = 5$ is reported in Appendix D.2 and D.3.

Table 3.3: Relative forecasting performance with respect to GAS for different methods across DGPs, at short horizons ($h = 1$ and $h = 3$)

DGP	**Horizon**	**Proposed Methods**		**Other Bayesian Methods**			**Lasso**
		GAS	**GVS**	**BL**	**BFL**	**S4**	$\lambda, c, \alpha, \Sigma$
DGP1							
$N = 3$	$h = 1$	1.00	1.05	1.12	1.22	1.31	1.16
	$h = 3$	1.00	1.13	1.06	1.09	1.15	1.33
$N = 5$	$h = 1$	1.00	1.05	1.22	1.22	1.42	1.03
	$h = 3$	1.00	1.04	1.13	1.05	1.18	1.32
$N = 10$	$h = 1$	1.00	0.95	1.43	1.22	1.68	1.00
	$h = 3$	1.00	1.02	1.11	1.32	1.36	1.21
DGP2							
$N = 3$	$h = 1$	1.00	1.03	1.05	1.04	0.99	1.02
	$h = 3$	1.00	0.97	1.00	1.02	1.11	1.01
$N = 5$	$h = 1$	1.00	1.08	1.14	1.17	1.26	1.09
	$h = 3$	1.00	1.10	1.06	1.17	1.22	1.15
$N = 10$	$h = 1$	1.00	1.01	1.16	1.14	1.57	1.05
	$h = 3$	1.00	0.95	1.10	1.31	1.17	0.99
DGP3							
$N = 3$	$h = 1$	1.00	0.99	1.00	0.96	1.21	0.87
	$h = 3$	1.00	0.84	1.04	0.97	1.02	0.90
$N = 5$	$h = 1$	1.00	1.05	1.08	1.08	1.54	1.01
	$h = 3$	1.00	1.20	1.00	1.05	1.13	1.04
$N = 10$	$h = 1$	1.00	1.16	1.54	1.38	1.58	0.99
	$h = 3$	1.00	1.21	1.21	1.09	1.17	1.06

Note: The table reports mean squared forecast errors (MSFE) at 1-step and 3-step ahead horizons across different models, normalized to the benchmark model (GAS). A value below one indicates better forecasting performance relative to the benchmark. Methods include Bayesian Lasso (BL), Bayesian Fused Lasso (BFL), S4, and frequentist Lasso. The Lasso setting $\lambda, c, \alpha, \Sigma$ refers to DI and CSH penalization also with penalization in Σ.

methods occasionally perform better, benefiting from more aggressive early shrinkage that helps capture the dynamics in densely interconnected systems.

Notably, in DGP 2, which introduces cross-sectional heterogeneity, GAS performs reliably at small dimensions, but as the panel size increases, GVS tends to outperform. This improvement reflects the ability of GVS to more effectively pool information between units in larger panels, where its stronger shrinkage can help stabilize forecasts by reducing estimation noise, even in the presence of heterogeneity.

S4 performs competitively in several scenarios, particularly in high-dimensional settings and at short to medium forecast horizons. However, its accuracy can be sensitive to prior scaling, especially as model complexity increases. Similarly, BL and BFL exhibit strong forecast performance in low-dimensional or weakly sparse environments, but their relative advantage

Table 3.4: Aggregated MSFE at Medium and Long Horizons ($h = 6$ and $h = 12$) Across DGPs and Dimensions (Normalized to GAS)

DGP	Horizon	Proposed Methods		Other Bayesian Methods			Lasso
		GAS	GVS	BL	BFL	S4	$\lambda, c, \alpha, \Sigma$
DGP1							
$N = 3$	$h = 6$	1.00	1.04	1.11	0.91	1.06	1.17
	$h = 12$	1.00	0.86	0.98	0.85	0.95	0.99
$N = 5$	$h = 6$	1.00	1.07	1.14	0.99	1.10	1.23
	$h = 12$	1.00	1.11	0.87	0.91	0.97	1.04
$N = 10$	$h = 6$	1.00	0.87	1.02	1.29	1.01	1.17
	$h = 12$	1.00	1.19	1.26	1.18	1.31	1.02
DGP2							
$N = 3$	$h = 6$	1.00	1.08	1.10	1.02	1.17	1.19
	$h = 12$	1.00	1.07	1.10	1.17	1.07	1.10
$N = 5$	$h = 6$	1.00	1.09	1.02	1.04	0.97	0.94
	$h = 12$	1.00	0.76	0.95	0.79	0.89	0.83
$N = 10$	$h = 6$	1.00	0.91	0.96	0.94	0.95	0.97
	$h = 12$	1.00	0.88	1.10	0.97	0.87	1.03
DGP3							
$N = 3$	$h = 6$	1.00	1.08	0.93	1.06	0.99	0.91
	$h = 12$	1.00	1.04	0.73	0.87	0.85	0.78
$N = 5$	$h = 6$	1.00	1.20	1.11	1.11	1.17	1.02
	$h = 12$	1.00	1.42	1.07	1.18	1.22	0.76
$N = 10$	$h = 6$	1.00	1.10	0.97	0.99	1.02	1.06
	$h = 12$	1.00	0.91	1.03	0.86	1.06	1.10

Note: MSFE values at forecast horizons $h = 6$ and $h = 12$, normalized to the base model (GAS). Values below 1 indicate improved forecast performance relative to GAS. Models include: Bayesian Lasso (BL), Bayesian Fused Lasso (BFL), S4, and frequentist Lasso with shrinkage on DI and/or error structure.

tends to diminish as forecast horizons lengthen or panel dimensionality increases.

Taken together, these results highlight the robustness and adaptability of the proposed PSSL framework. GAS, in particular, delivers a well-balanced trade-off between estimation accuracy, variable selection, and forecasting performance. It consistently performs well in sparse and semi-sparse environments and maintains reasonable accuracy even as the underlying structure becomes dense. GVS is also competitive, particularly in larger panels where more cross-sectional information supports stronger shrinkage. Because shrinkage toward a common value imposes strong restrictions, it requires sufficient cross-sectional information to distinguish true homogeneity from variation. As a result, its flexibility may be limited in settings with moderate heterogeneity. While S4 shows resilience in high-dimensional settings, its reliance on fixed prior scaling can reduce performance in more complex environments. Overall, these findings reinforce the value of incorporating structured and adaptive regularization when estimating

high-dimensional panel VAR models.

3.2.3 Data-based Simulation Results (DGP4)

The first three simulations (DGP1–DGP3) demonstrate that the proposed PSSL approach enhances both estimation and forecasting accuracy across a range of sparse and dense settings. We now turn to DGP4, a pseudo-realistic simulation designed to mimic the structural characteristics of real-world macroeconomic data.

This simulation uses data from Camehl (2023), which includes 20 countries[10] and four core macroeconomic indicators: the Harmonized Index of Consumer Prices (HICP), Industrial Production Growth (IP), the Real Effective Exchange Rate (REER), and the Unemployment Rate (UN). In our data-based simulation, we construct panel subsets using 12 countries, comprising 10 euro area countries commonly used in panel VAR research[11], specifically five core countries (Austria, Belgium, Finland, France, and the Netherlands) and five periphery countries (Spain, Greece, Ireland, Italy, and Portugal), along with two externally influential economies: the United Kingdom and the United States.

To generate the pseudo-real data-generating process (DGP4), we estimate a VAR using the full sample via ordinary least squares (OLS) to obtain the coefficient matrix A^{OLS}. We then use A^{OLS} to generate fitted values $\hat{Y}$, which serve as the simulated dependent variables. This approach preserves the underlying empirical structure of the data and simulates the conditions a researcher would face in this applied setting. Figure 3.2 illustrates the data construction and simulation process.

Table 3.5 reports aggregate-level forecast performance, showing relative mean squared forecast errors (MSFE) averaged across countries and variables. Tables 3.6 and 3.7 present country-level MSFE results. Across all comparisons, GAS is used for normalization.

At the aggregate level, the proposed Panel Spike-and-Slab Lasso (PSSL) framework delivers consistently strong forecasting performance, particularly at short and medium horizons ($h = 1, 3$). From the comparison with other methods, three key points stand out. First, compared with non-adaptive shrinkage, the adaptive shrinkage of PSSL outperforms it. Because the fixed penalty in non-adaptive shrinkage might overshrink by pushing nonzero coefficients too close to zero in the first step. However, as the forecast horizon increases, this gap becomes smaller. Second, PSSL outperforms the S4 method. The gain comes from PSSL's ability to adjust shrinkage based on the data, rather than relying on S4's preset shrinkage level from a fixed

[10] Austria, Belgium, Germany, Denmark, Spain, Finland, France, Greece, Ireland, Italy, Luxembourg, Netherlands, Norway, Poland, Portugal, Sweden, Slovenia, Slovakia, the United Kingdom, and the United States.

[11] See, for example, Koop and Korobilis (2016) and Korobilis (2016).

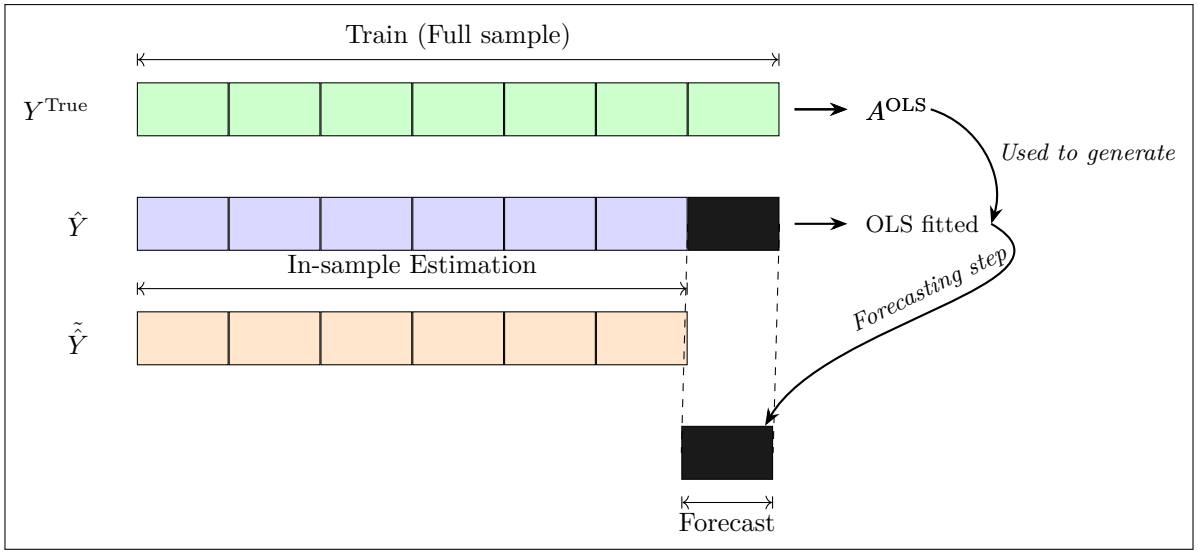


Figure 3.2: Real-World DGP Setting

Table 3.5: Aggregated MSFE Across All Countries and Variables at Forecast Horizons $h = 1, 3, 6, 12$

Horizon	**PSSL**		**Other Bayesian Methods**			**Lasso**
	GAS	**GVS**	**BL**	**BFL**	**S4**	$\lambda, c, \alpha, \Sigma$
$h = 1$	1.00	0.92	1.12	1.11	1.16	1.07
$h = 3$	1.00	0.99	1.12	1.12	1.20	1.35
$h = 6$	1.00	1.00	1.00	1.00	1.06	1.13
$h = 12$	1.00	1.00	1.03	1.03	1.04	1.05

Note: This table reports the average Mean Squared Forecast Error (MSFE) for different methods across multiple forecast horizons, aggregated over both countries and forecast variables. All MSFE values are normalized relative to the GAS benchmark (GAS = 1). The methods include PSSL with GVS, Bayesian shrinkage alternatives (Bayesian Lasso (BL), Bayesian Fused Lasso (BFL), Stochastic Search Variable Selection (S4)), and a frequentist Lasso with tuning over λ, c, α, and Σ.

normal prior. Third, in the comparison between GVS and GAS, GVS has better results in this setting, due to the increasing number of countries and variables. With more cross-sectional information, GVS can apply stronger shrinkage to group-specific values, which proves more effective than the mild shrinkage on group averages used by GAS.

Disaggregated country-level results broadly reinforce the aggregate findings. At the 1-step horizon ($h = 1$), simpler shrinkage methods (BL, BFL, or the frequentist Lasso) match or outperform PSSL in several countries, including Finland, Portugal, the United Kingdom, and the United States. One possible explanation is that, apart from Finland, these economies are either geographically peripheral or external to the euro area and therefore less subject to the type of cross-sectional homogeneity that PSSL can capture. In contrast, in more closely integrated euro-area economies such as Belgium, France, Greece, and the Netherlands, the simpler methods yield higher MSFE values than PSSL, suggesting that models combining both

Table 3.6: Country-Level Relative MSFE at $h = 1$ and $h = 3$ (Normalized to GAS)

Country	***Panel A:*** $h = 1$					
	PSSL		**Bayesian Methods**			**Lasso**
	GAS	**GVS**	**BL**	**BFL**	**S4**	$\lambda, c, \alpha, \Sigma$
Austria	1.00	0.95	1.19	1.19	1.28	1.01
Belgium	1.00	1.08	1.21	1.15	1.06	1.10
Finland	1.00	0.99	0.94	0.93	0.90	0.68
France	1.00	1.13	1.15	1.15	1.17	1.88
Greece	1.00	1.08	1.38	1.36	1.51	1.26
Ireland	1.00	0.80	1.13	1.13	1.15	1.22
Italy	1.00	1.05	0.94	0.93	1.03	1.36
Netherlands	1.00	1.03	1.37	1.37	1.39	1.61
Portugal	1.00	0.91	0.78	0.78	0.79	0.97
Spain	1.00	0.90	1.66	1.60	1.70	1.85
United Kingdom	1.00	0.95	1.01	1.00	0.99	0.92
United States	1.00	0.97	1.01	1.01	1.06	0.79
Country	***Panel B:*** $h = 3$					
	PSSL		**Bayesian Methods**			**Lasso**
	GAS	**GVS**	**BL**	**BFL**	**S4**	$\lambda, c, \alpha, \Sigma$
Austria	1.00	0.98	0.99	0.99	1.06	1.15
Belgium	1.00	1.04	1.43	1.41	1.37	1.33
Finland	1.00	1.00	0.89	0.89	0.96	1.07
France	1.00	1.04	1.28	1.28	1.36	1.43
Greece	1.00	0.97	1.40	1.41	1.42	1.35
Ireland	1.00	0.97	1.14	1.14	1.13	1.16
Italy	1.00	0.98	0.97	0.98	1.01	1.13
Netherlands	1.00	0.98	1.23	1.23	1.26	1.12
Portugal	1.00	0.93	0.87	0.87	0.93	1.13
Spain	1.00	0.89	1.09	1.09	1.05	1.01
United Kingdom	1.00	0.93	0.88	0.88	0.89	0.84
United States	1.00	1.08	1.11	1.11	1.15	1.15

Note: The table reports country-level relative Mean Squared Forecast Errors (MSFE) at $h = 1$ (Panel A) and $h = 3$ (Panel B). Entries are computed as $\text{MSFE}_{\text{model}}/\text{MSFE}_{\text{GAS}}$, where GAS is a benchmark forecast model. Methods include: PSSL with GVS, Bayesian shrinkage models (BL, BFL, S4), and a frequentist Lasso with shrinkage over hyperparameters λ, c, α, and Σ.

CSH and DI provide more useful information.

By the 3-step horizon ($h = 3$), the performance of the simpler methods tends to deteriorate in many countries, with increases in MSFE for Belgium, France, and the United States. One

Table 3.7: Country-Level Relative MSFE at $h = 6$ and $h = 12$ (Normalized to GAS)

Country	***Panel A:*** $h = 6$					
	PSSL		**Bayesian Methods**			**Lasso**
	GAS	**GVS**	**BL**	**BFL**	**S4**	$\lambda, c, \alpha, \Sigma$
Austria	1.00	1.09	1.40	1.41	1.44	1.28
Belgium	1.00	1.06	1.12	1.13	1.16	1.15
Finland	1.00	0.99	0.91	0.92	0.98	1.05
France	1.00	0.95	0.69	0.69	0.80	1.11
Greece	1.00	0.98	0.92	0.92	0.97	1.10
Ireland	1.00	0.99	0.89	0.90	0.98	1.07
Italy	1.00	1.00	0.99	0.99	1.01	1.05
Netherlands	1.00	1.06	1.25	1.25	1.22	1.26
Portugal	1.00	0.97	0.90	0.91	0.95	1.02
Spain	1.00	1.01	0.96	0.97	1.02	1.06
United Kingdom	1.00	1.05	1.15	1.15	1.17	1.17
United States	1.00	1.02	1.16	1.16	1.10	1.04
Country	***Panel B:*** $h = 12$					
	PSSL		**Bayesian Methods**			**Lasso**
	GAS	**GVS**	**BL**	**BFL**	**S4**	$\lambda, c, \alpha, \Sigma$
Austria	1.00	1.00	1.02	1.02	1.01	1.00
Belgium	1.00	1.02	1.13	1.13	1.10	1.06
Finland	1.00	1.00	0.93	0.93	0.98	1.02
France	1.00	1.00	0.98	0.98	1.00	1.03
Greece	1.00	1.01	1.18	1.18	1.09	1.00
Ireland	1.00	1.00	0.95	0.95	0.99	1.03
Italy	1.00	1.01	1.10	1.10	1.08	1.07
Netherlands	1.00	1.00	1.11	1.11	1.06	1.02
Portugal	1.00	0.99	0.93	0.93	0.97	1.01
Spain	1.00	1.00	1.02	1.02	1.03	1.03
United Kingdom	1.00	0.99	0.86	0.87	0.94	1.00
United States	1.00	1.01	1.24	1.23	1.10	0.96

Note: The table reports country-level relative Mean Squared Forecast Errors (MSFE) at $h = 6$ (Panel A) and $h = 12$ (Panel B). Entries are computed as $\text{MSFE}_{\text{model}}/\text{MSFE}_{\text{GAS}}$, where GAS is a benchmark forecast model. Methods include: PSSL with GVS, Bayesian shrinkage models (BL, BFL, S4), and a frequentist Lasso with shrinkage over hyperparameters λ, c, α, and Σ.

possible reason is that non-adaptive shrinkage methods impose stronger, uniform penalization that fits dominant DI well but may ignore cross-sectional components, leading to less accurate multi-step forecasts. PSSL, by contrast, maintains comparatively stable accuracy across

countries, even when they are not the single best-performing method in every case. This result suggests that incorporating both CSH and DI in estimation can help maintain forecast precision beyond the very short run by capturing broader in-sample fluctuations.

At medium and long horizons ($h = 6$ and $h = 12$), differences across methods become smaller. PSSL remains the best or joint-best performer in most countries, indicating that the benefits of combining DI and CSH extend beyond the short run to longer-horizon forecasts. In some countries, other approaches, most often BL or BFL, perform equally well. The S4 method also delivers good results in certain countries. One possible explanation is that, at these horizons, forecast performance relies more on persistent components that are relatively stable across countries, which PSSL, Bayesian shrinkage methods, and S4 can all capture effectively.

Overall, these country-level patterns indicate that while simpler shrinkage and S4 methods can perform well for certain economies, PSSL delivers strong forecast results in most countries and exhibits greater stability as the forecast horizon increase, maintaining accuracy across a wide range of country–horizon combinations.

4 Empirical Applications

We illustrate the group-specific variable selection performance and the forecasting performance of the proposed *PSSL* framework through two separate empirical applications. The first application revisits and extends the dataset from Koop and Korobilis (2016) to examine financial contagion in the Eurozone sovereign debt market, with a focus on the model's ability to identify relevant coefficients through posterior inclusion probabilities. The second application assesses the forecasting performance of the *PSSL* framework in a macroeconomic spillover context. It builds on the dataset from Camehl (2023), previously used in the data-based simulation section, and considers a panel of 10 European countries and two key macroeconomic indicators.

4.1 Eurozone Financial Contagion

The euro area sovereign debt crisis and its aftermath offer a compelling context for examining spillover effects within the euro area, given the heightened volatility of sovereign spreads since 2008 and the asymmetric responses of member countries during periods of financial stress. Building on the empirical frameworks of Koop and Korobilis (2016) and Korobilis (2016), we investigate cross-country interlinkages and the dynamics of financial contagion in the euro area. We use an extended monthly dataset spanning January 2001 to December 2019, thereby covering not only the pre-crisis and sovereign debt crisis periods but also the subsequent post-crisis years. The analysis focus on three key variables: (i) 10-year government bond yields, (ii) total industrial

production, and (iii) average bid-ask spreads as a proxy for market liquidity. As in previous studies, all variables are expressed as spreads relative to Germany, which is commonly treated as the benchmark risk-free country within the euro area. This normalization facilitates cross-country comparisons of sovereign credit risk, economic activity, and liquidity. The euro area is divided into core countries (Austria, Belgium, Finland, France, and the Netherlands) and periphery countries (Greece, Ireland, Italy, Portugal, and Spain). To ensure stationarity, bond yields and bid-ask spreads are transformed using first differences, while industrial production is expressed as year-over-year log changes.

4.1.1 Model Selection and Structural Restrictions

To determine the most appropriate model structure for forecasting and structural interpretation, we compare five PVAR specifications using their marginal likelihoods. The first is a PSSL model that imposes both dynamic interdependencies (DI) and cross-sectional heterogeneity (CSH) restrictions, implemented through the GVS and GAS approaches. The second allows for DI only while permitting full cross-sectional heterogeneity. The third imposes CSH restrictions while allowing for unrestricted dynamic interactions. The fourth is a fully unrestricted PVAR model without any shrinkage. Finally, we consider the S4 model of Koop and Korobilis (2016), which incorporates both DI and CSH using a fixed, calibrated shrinkage based on a normal prior.

Table 4.1: Model Comparison Using Marginal Likelihood (Gelfand–Dey Method)

Model	**DI + CSH**	**DI**	**CSH**	**Unrestricted**	**S4**
GVS	−41.01	−38.46	−49.48	−43.65	−42.41
GAS	−50.64		−50.76		

Note: The table reports the marginal log-likelihood values computed using the Gelfand–Dey method. Larger values indicate better model fit. The columns represent alternative restrictions: imposing both DI and CSH ("Both"), DI restriction only ("DI only"), CSH restriction only ("CSH only"), unrestricted specification ("Unrestricted"), and the S4 method from Koop and Korobilis (2016) ("S4").

Table 4.1 shows that the model imposing only dynamic interdependency (DI) restrictions achieves the highest marginal likelihood, confirming that dynamic spillovers are the main drivers of model fit in this context. In contrast, models that impose cross-sectional homogeneity (i.e., exclude cross-sectional heterogeneity, CSH) perform poorly—worse even than the fully unrestricted specification—underscoring the limitations of imposing strong homogeneity assumptions across euro area countries.

The GAS model, which incorporates both DI and CSH via group-average shrinkage, offers

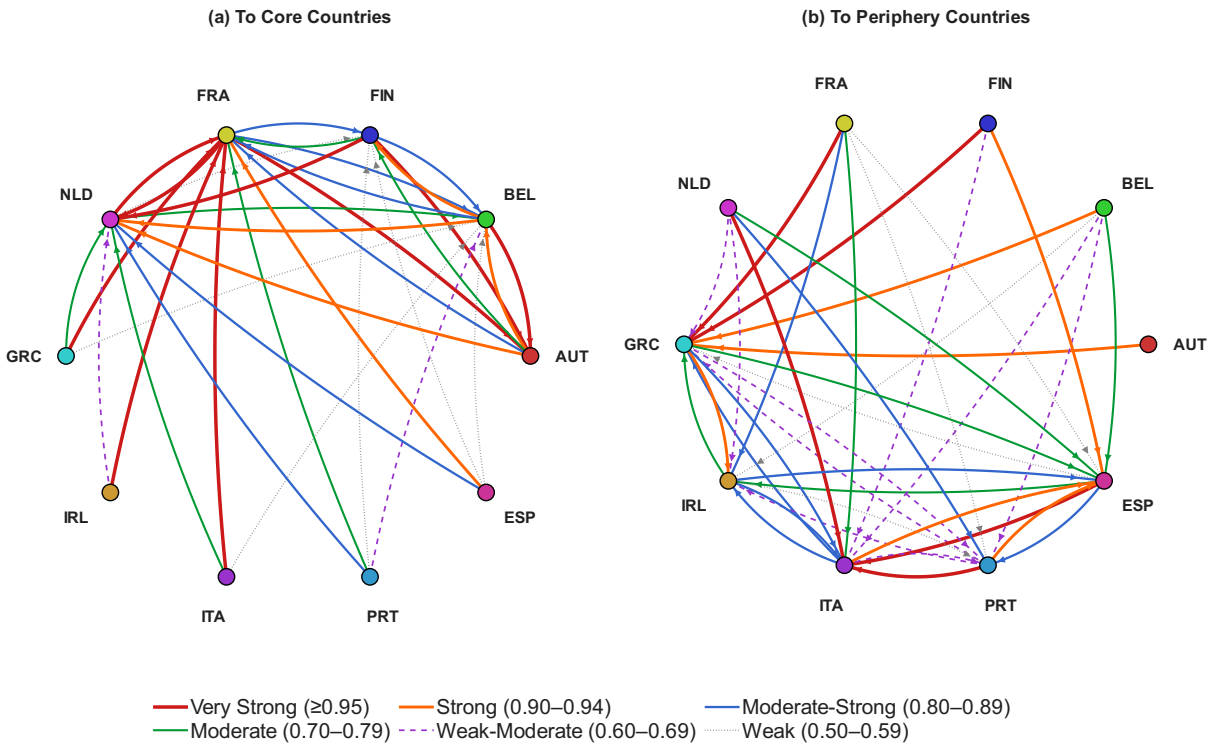


Figure 4.1: Posterior Dynamic Interdependencies Across the Euro Area

Note: This figure visualizes the posterior dynamic interdependencies (DI) between Euro Area countries. The left panel shows linkages where the receiving country belongs to the core group (Austria, Belgium, Finland, France, Netherlands), while the right panel displays linkages targeting the periphery group (Greece, Ireland, Italy, Portugal, Spain). Arrow styles and colors indicate the strength of posterior evidence, with thicker and darker lines representing stronger spillover effects.

only a marginal improvement over the CSH-only specification. This outcome suggests that shrinking toward a group mean may obscure important within-group variation. While some degree of homogeneity may exist, it likely occurs at the variable level rather than uniformly across group averages. As a result, the diffuse shrinkage employed by GAS fails to adequately account for the structural heterogeneity present in the data.

By contrast, GVS and S4 apply shrinkage toward group-specific values rather than group means, allowing them to preserve meaningful within-group variation while still benefiting from regularization. GVS, in particular, delivers more targeted and effective shrinkage, consistently outperforming GAS in models that impose both DI and CSH restrictions. S4 also outperforms the unrestricted model, positioning itself between GVS and GAS in terms of overall performance. These results highlight the value of regularization strategies that strike a balance between parsimony and the preservation of variable-level heterogeneity, rather than relying on overly smooth group-level shrinkage that may obscure important structural differences.

4.1.2 Dynamic Interdependencies Across the Euro Area

Posterior inclusion probabilities for dynamic interdependencies (DIs) are summarized in Tables F.1 and F.2, and visualized in Figure 4.1[12]. These results highlight clear directional asymmetries in spillover structures. Strong evidence emerges for spillovers from core to periphery economies—for instance, France to Greece (0.96), Belgium to Greece (0.94), and Finland to Greece (0.95). In contrast, reverse linkages such as Greece to Austria (0.29) or Ireland to Austria (0.31) show little evidence of significant DI.

These patterns suggest that macro-financial shocks largely originate from core economies and propagate outward. This may reflect institutional factors such as deeper financial markets, more stable fiscal policy frameworks, and greater systemic relevance of core countries. Within the core, interdependencies are dense and highly significant—for example, Belgium to Austria (0.96) or Netherlands to France (0.96)—reflecting the high integration of sovereign bond and liquidity markets.

In contrast, DI patterns among periphery countries are weaker and more variable. Some strong bilateral linkages exist—such as Italy to Spain (0.92) or Portugal to Italy (0.95)—but most lie in the moderate to weak range. These results mirror structural differences among peripheral economies, which vary widely in terms of debt sustainability, institutional quality, and exposure to external shocks. Nevertheless, some exceptions emerge—such as Greece and Ireland transmitting shocks to France (0.96 and 0.95, respectively)—likely reflecting contagion during the sovereign crisis period and France's high exposure to peripheral debt.

Together, these findings underscore a euro area structure characterized by dominant core-to-periphery spillovers, moderate within-core linkages, and weak periphery-to-core transmission, consistent with a more asymmetric and hierarchical network of interdependencies.

4.2 Forecasting the international spillovers

In the second application, we evaluate the forecasting performance of our PSSL model from an international spillover perspective. Building on the 10 countries used in the financial contagion analysis, we additionally include the United Kingdom and the United States, resulting in a panel that captures both euro area core and periphery dynamics, as well as external influences. This extended panel follows the structure used in Camehl (2023), which we also employed in the data-based simulation study. Here, however, we used real-world data over an extended horizon from January 2000 to December 2019, which ended just before the COVID-19 pandemic. We

[12]For the interpretation of the strength of evidence (from weak to strong) based on posterior probabilities, see Table E.1 in Appendix E.

Table 4.2: Average Relative Root of MSFE Across Countries and Variables for Forecast Horizons $h = 1, 2, 3, 6$

Horizon	**PSSL**		**Other Bayesian Methods**			**Lasso**
	GVS	**GAS**	**BL**	**BFL**	**S4**	$\lambda, c, \alpha, \Sigma$
$h = 1$	1	1.25	3.10	3.04	2.31	2.48
$h = 3$	1	1.10	2.09	2.08	1.79	1.52
$h = 6$	1	1.05	1.58	1.57	1.43	1.24
$h = 12$	1	0.99	2.62	2.60	1.81	1.29

Note: This table reports the average relative Root Mean Squared Forecast Errors (rMSFE) across models, forecast horizons, countries, and variables. Each entry represents $\text{rMSFE}_{\text{model}}/\text{rMSFE}_{\text{benchmark}}$), where the benchmark is GVS model. Models include GAS (Alternative PSSL), several Bayesian shrinkage priors (BL, BFL, S4), and a frequentist variant (HgavLL). Averages are taken across all countries and forecast variables.

include all four macroeconomic variables considered in Camehl (2023): seasonally adjusted log-differences of the consumer price index (CPI) and industrial production (IP), the log-differenced and demeaned real effective exchange rate (REER), and the first difference of the unemployment rate.[13]

We estimate all models using data up to December 2018 and generate a single 12-step-ahead iterated forecast covering January to December 2019. Table 4.2 reports the aggregate log-relative Mean Squared Forecast Errors across all countries and variables at horizons $h = 1, 3, 6, 12$, normalized to GVS (as GVS gives the best performance in the model selection). Note that GVS is included in the table with a value of 1 at all horizons, and all other values are reported relative to it; values greater than 1 indicate worse performance than GVS, while values less than 1 reflect improvements. For example, a value of 1.25 from GAS at horizon $h = 1$ corresponds to a 25% higher rMSFE relative to GVS. The GVS model consistently achieves strong forecast performance across all horizons. While GAS performs slightly better than GVS at long horizons, its advantage is not consistent across settings. Other Bayesian alternatives, such as the Bayesian Lasso (BL), Bayesian Fused Lasso (BFL), and S4—perform reasonably well, but none consistently outperform the GVS benchmark. These empirical findings are consistent with our simulation results, where GVS tends to outperform GAS as the dimension of the panel increases. In settings with a larger number of variables within groups, such as macroeconomic panels with rich information within the group and cross-sectional, the GVS benefits from its stronger shrinkage toward common structures, resulting in improved forecast stability and accuracy.

[13]The CPI series is seasonally adjusted using the X-13ARIMA-SEATS filter in `R`. The IP series is seasonally adjusted at source. Both series are sourced from the IMF.

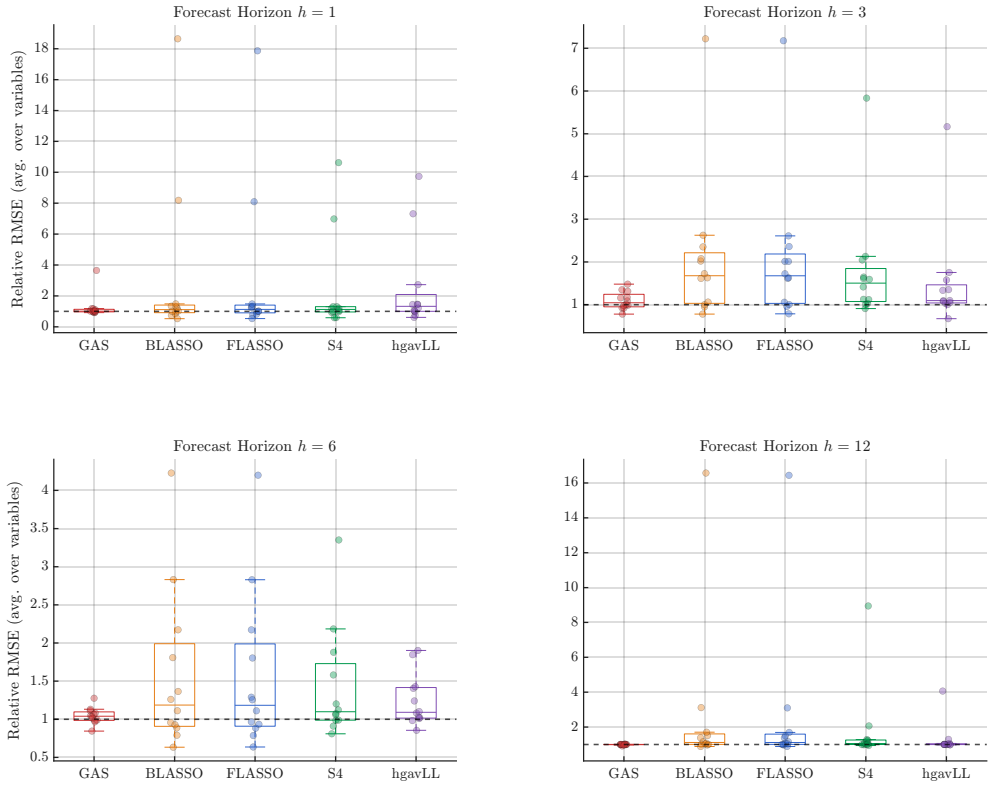


Figure 4.2: Country-Level Distribution of Relative rMSFE for Forecast Horizons $h = 1, 3, 6, 12$.

Note: This figure displays the distribution of relative mean squared forecast errors (MSFE) across five forecasting models and multiple countries. Each subplot corresponds to a forecast horizon $h = 1, 3, 6, 12$. For each model, the boxplot summarizes the distribution of rMSFE / Benchmark values across all countries. **Each dot represents a single country's relative rMSFE** at that horizon. A value below one indicates better performance than the benchmark model; a value above one indicates worse performance. The dashed line at one denotes equal performance to the benchmark.

Tables F.3 and F.4 report country-level relative rMSFE values across forecast horizons, and Figure 4.2 provides a visual summary of these results. The boxplots display the distribution of rMSFE / $\text{rMSFE}_{\text{GVS}}$ values for each competing method at horizons $h = 1, 3, 6, 12$. The horizontal dashed line at one represents the benchmark GVS performance. Values above one indicate worse forecast accuracy relative to GVS, while values below one indicate better performance. Given this benchmark structure, the figure shows that although GVS is not explicitly plotted, as all MSFEs are normalized to it, which remain a strong benchmark across all horizons. GAS exhibits competitive performance, particularly at shorter horizons, with a relatively tight distribution around the benchmark. In contrast, methods such as BLASSO, FLASSO, and Frequentist Lasso display greater dispersion and more frequent positive outliers, indicating occasional but substantial declines in forecast accuracy compared to GVS. S4 also shows greater variability across both horizons and countries. In other words, across all horizons and countries,

GVS demonstrates a better forecast stability.

GAS shows competitive performance, particularly at shorter horizons (e.g., $h = 1$), outperforming GVS in countries like Finland, France, and Greece. However, the performance of GAS is less consistent. For instance, GAS performs significantly worse than GVS in countries such as Belgium, especially at short horizons. Other methods such as the Bayesian Lasso (BL), Bayesian Fused Lasso (BFL), and S4 show mixed results. While they occasionally outperform GVS in specific countries or horizons, they generally yield higher MSFEs and display wider cross-country variability. Specifically, these methods produce large MSFE values in countries like Belgium and Spain. The frequentist Lasso method also lacks consistency and shows few outliers in forecast accuracy across countries.

Overall, these results underscore the strength of the GVS approach. Its adaptive shrinkage structure flexibly accommodates both sparse and dense country-specific dynamics. Even when prior assumptions do not perfectly align with the underlying data, GVS remains stable and competitive. Its structure and stronger shrinkage highlight its stable and reliable forecast performance across diverse countries and time horizons. While GAS also remains competitive at short horizons, GVS offers more stability and competitiveness, especially in larger panels where richer cross-sectional information supports stronger pooling. This property is particularly valuable in This property is especially valuable in international macroeconomic forecasting, where cross-country interdependencies are heterogeneous and often uncertain. From a policymaking or practitioner's perspective, this robustness is crucial: rather than performing well only in selected countries, GVS avoids major forecasting failures and provides consistently reliable global signals.

5 Conclusion

This paper proposes a novel Bayesian framework for estimating high-dimensional panel VARs that combines flexibility, interpretability, and strong empirical performance. The central contribution is the development of the Panel Spike-and-Slab Lasso (PSSL)—a hierarchical prior structure that facilitates variable selection at the unit-variable level while preserving dynamic interdependencies across units. Unlike global shrinkage priors such as the Bayesian Lasso, which apply uniform penalization across all coefficients, PSSL introduces componentwise, data-driven shrinkage using spike-and-slab priors. This adaptive structure enables sharper discrimination between relevant and irrelevant predictors, allowing the model to uncover both sparse and dense patterns of interaction.

The proposed method also improves upon recent approaches such as the Stochastic Search

Structure Selection (S4) prior. While S4 captures both dynamic interdependencies (DI) and cross-sectional heterogeneity (CSH) through group-specific shrinkage, it relies on Gaussian mixture priors with fixed variances and shared inclusion probabilities within groups. This soft pooling mechanism can limit important within-group heterogeneity, particularly when individual countries or units exhibit divergent dynamics. In contrast, PSSL imposes a fully hierarchical spike-and-slab structure with unit-specific inclusion probabilities, yielding sharper sparsity and more interpretable posterior inclusion measures. As a result, PSSL better captures heterogeneity in cross-sectional dynamics and delivers improved estimation and forecasting performance, especially in panels with moderate sparsity and uneven spillover patterns.

Simulation results confirm that the method delivers substantial improvements in estimation accuracy, particularly in settings where only a subset of potential interdependencies is active. The adaptive shrinkage mechanism allows PSSL to remain robust even when the degree of sparsity varies across units or over time, avoiding overfitting while preserving signal.

Two empirical applications illustrate the method's practical advantages. In modeling sovereign bond markets in the euro area, PSSL uncovers meaningful patterns of financial contagion—highlighting dense and persistent linkages among core countries, weaker connectivity among peripheral economies, and dynamic feedback effects during periods of crisis. In a macroeconomic forecasting context, using a 12-country panel of inflation and output, PSSL delivers consistent and stable forecast performance across countries and horizons. It frequently outperforms benchmark methods and, importantly, avoids large forecast errors that can undermine aggregate insights in international macro panels.

In sum, this paper presents a flexible and empirically grounded approach to modeling high-dimensional panel systems. By combining hierarchical variable selection, adaptive shrinkage, and interpretable coefficient inclusion, the PSSL framework offers a valuable tool for analyzing heterogeneous interdependencies in complex macroeconomic and financial environments. Its design is broadly applicable to settings where spillovers, heterogeneity, and dynamic interactions are central—ranging from regional business cycles to global financial networks—and provides researchers and policymakers with a robust, transparent, and scalable method for understanding interconnected systems.

References

Bai, Ray, Gemma E Moran, Joseph L Antonelli, Yong Chen, and Mary R Boland (2022) "Spike-and-slab group lassos for grouped regression and sparse generalized additive models," *Journal of the American Statistical Association*, 117 (537), 184–197.

Bai, Ray, Veronika Ročková, and Edward I George (2021) "Spike-and-slab meets LASSO: A review of the spike-and-slab LASSO," *Handbook of Bayesian variable selection*, 81–108.

Camehl, Annika (2023) "Penalized estimation of panel vector autoregressive models: A panel LASSO approach," *International Journal of Forecasting*, 39 (3), 1185–1204.

Canova, Fabio and Matteo Ciccarelli (2004) "Forecasting and turning point predictions in a Bayesian panel VAR model," *Journal of Econometrics*, 120 (2), 327–359.

——— (2009) "Estimating multicountry VAR models," *International economic review*, 50 (3), 929–959.

——— (2013) "Panel Vector Autoregressive Models: A Survey," in *VAR models in macroeconomics–new developments and applications: Essays in honor of Christopher A. Sims*, 205–246: Emerald Group Publishing Limited.

Casella, George, Malay Ghosh, Jeff Gill, and Minjung Kyung (2010) "Penalized regression, standard errors, and Bayesian lassos."

Dees, Stephane, Filippo di Mauro, M Hashem Pesaran, and L Vanessa Smith (2007) "Exploring the international linkages of the euro area: a global VAR analysis," *Journal of applied econometrics*, 22 (1), 1–38.

Doan, Thomas, Robert Litterman, and Christopher Sims (1984) "Forecasting and conditional projection using realistic prior distributions," *Econometric reviews*, 3 (1), 1–100.

Feldkircher, Martin and Florian Huber (2016) "The international transmission of US shocks—evidence from Bayesian global vector autoregressions," *European Economic Review*, 81, 167–188.

George, Edward I, Dongchu Sun, and Shawn Ni (2008) "Bayesian stochastic search for VAR model restrictions," *Journal of Econometrics*, 142 (1), 553–580.

Huang, Jian, Shuangge Ma, and Cun-Hui Zhang (2008) "Adaptive Lasso for sparse high-dimensional regression models," *Statistica Sinica*, 1603–1618.

Huber, Florian, Tamás Krisztin, and Michael Pfarrhofer (2023) "A Bayesian panel vector autoregression to analyze the impact of climate shocks on high-income economies," *The Annals of Applied Statistics*, 17 (2), 1543–1573.

Jarociński, Marek (2010) "Responses to monetary policy shocks in the east and the west of Europe: a comparison," *Journal of Applied Econometrics*, 25 (5), 833–868.

Koop, Gary and Dimitris Korobilis (2016) "Model uncertainty in panel vector autoregressive models," *European Economic Review*, 81, 115–131.

——— (2019) "Forecasting with high-dimensional panel VARs," *Oxford Bulletin of Economics and Statistics*, 81 (5), 937–959.

Korobilis, Dimitris (2016) "Prior selection for panel vector autoregressions," *Computational Statistics & Data Analysis*, 101, 110–120.

Kose, M Ayhan, Christopher Otrok, and Charles H Whiteman (2003) "International business cycles: World, region, and country-specific factors," *american economic review*, 93 (4), 1216–1239.

Litterman, Robert B (1986) "Forecasting with Bayesian vector autoregressions—five years of experience," *Journal of Business & Economic Statistics*, 4 (1), 25–38.

Nicholson, William B, Ines Wilms, Jacob Bien, and David S Matteson (2020) "High dimensional forecasting via interpretable vector autoregression," *The Journal of Machine Learning Research*, 21 (1), 6690–6741.

Park, Trevor and George Casella (2008) "The bayesian lasso," *Journal of the American Statistical Association*, 103 (482), 681–686.

Ročková, Veronika and Edward I George (2018) "The spike-and-slab lasso," *Journal of the American Statistical Association*, 113 (521), 431–444.

Song, Song and Peter J Bickel (2011) "Large vector auto regressions," *arXiv preprint arXiv:1106.3915*.

Tibshirani, Robert (1996) "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58 (1), 267–288.

Yuan, Ming and Yi Lin (2006) "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68 (1), 49–67.

Zhao, Peng and Bin Yu (2006) "On model selection consistency of Lasso," *The Journal of Machine Learning Research*, 7, 2541–2563.

Zou, Hui (2006) "The adaptive lasso and its oracle properties," *Journal of the American statistical association*, 101 (476), 1418–1429.

Appendices

A Model Setup and Additional Derivation

This appendix provides further details on the model specification used in the paper, as well as additional derivations not covered in the main text. We begin by summarizing the general model setup and notation. We then outline the key steps in deriving the estimation procedure and its associated formulas.

A.1 Panel VAR Setup

We define that the panel VAR model consists of N units, which might be countries, industry sectors, companies, etc. A standard VAR model for each unit i includes endogenous variables G, lags P, and over time periods T. We assume balanced panels, which implies that the G variables are identical for each entity and are defined in the same T time periods.

Panel VAR model for each unit i (with $i = 1, \cdots, N$), is defined as

$$\begin{aligned} y_{i,t} &= \sum_{j=1}^{N} \sum_{p=1}^{P} A_p^{ij} y_{j,t-p} + \varepsilon_t \\ &= A_1^{i1} y_{1,t-1} + \cdots + A_P^{i1} y_{1,t-P} \\ &+ A_1^{i2} y_{2,t-1} + \cdots + A_P^{i2} y_{2,t-P} \\ &+ \cdots \\ &+ A_1^{iN} y_{N,t-1} + \cdots + A_P^{iN} y_{N,t-P} + \varepsilon_{i,t}, \end{aligned}$$

with

$$y_{i,t} = \underbrace{\begin{pmatrix} y_{1,t}^i \\ y_{2,t}^i \\ \vdots \\ y_{G,t}^i \end{pmatrix}}_{G \times 1} \qquad A_p^{ij,t} = \underbrace{\begin{pmatrix} \alpha_p^{ij,11,t} & \alpha_p^{ij,12,t} & \cdots & \alpha_p^{ij,1G,t} \\ \alpha_p^{ij,21,t} & \alpha_p^{ij,22,t} & \cdots & \alpha_p^{ij,2G,t} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_p^{ij,G1,t} & \alpha_p^{ij,G2,t} & \cdots & \alpha_p^{ij,GG,t} \end{pmatrix}}_{G \times G} \qquad \epsilon_{i,t} = \underbrace{\begin{pmatrix} \epsilon_{1,t}^i \\ \epsilon_{2,t}^i \\ \vdots \\ \epsilon_{G,t}^i \end{pmatrix}}_{G \times 1}.$$

y_{it} is a $G \times 1$ vector representing G endogenous variables of unit i at time t, while $y_{ij,t}$ denotes the endogenous variable from unit j to unit i. $A_{ij,t}^p$ is a $G \times G$ coefficient matrix that captures the response of unit i to the p^{th} lag of unit j at time t. For $\epsilon_{i,t}$, it follows that $\epsilon_{i,t} \sim \mathcal{N}(0, \Sigma_{ii,t})$,

where $\Sigma_{ii,t}$ is defined as:

$$\Sigma_{ii,t} = \mathbb{E}(\epsilon_{i,t}\epsilon'_{i,t}) = \mathbb{E}\begin{pmatrix}\epsilon^{i}_{1,t}\\ \epsilon^{i}_{2,t}\\ \vdots\\ \epsilon^{i}_{G,t}\end{pmatrix}\begin{pmatrix}\epsilon^{i'}_{1,t} & \epsilon^{i'}_{2,t} & \cdots & \epsilon^{i'}_{G,t}\end{pmatrix} = \underbrace{\begin{pmatrix}\sigma^{ii}_{11,t} & \sigma^{ii}_{12,t} & \cdots & \sigma^{ii}_{1G,t}\\ \sigma^{ii}_{21,t} & \sigma^{ii}_{22,t} & \cdots & \sigma^{ii}_{2G,t}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma^{ii}_{G1,t} & \sigma^{ii}_{G2,t} & \cdots & \sigma^{ii}_{GG,t}\end{pmatrix}}_{G\times G}.$$

And for

$$y_{it} = A_{i,1}Y_{t-1} + \cdots + A_{i,p}Y_{t-p} + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathcal{N}(0, \Sigma_{ii}), \tag{A.1}$$

$$Y_{i,t} = \underbrace{\begin{pmatrix}y_{1,t}\\ y_{2,t}\\ \vdots\\ y_{N,t}\end{pmatrix}}_{NG\times 1} \qquad A_{p,i} = \underbrace{\begin{pmatrix}\alpha_p^{11,11,t} & \cdots & \alpha_p^{11,1G,t} & \cdots & \alpha_p^{1N,11,t} & \cdots & \alpha_p^{1N,1G,t}\\ \alpha_p^{11,21,t} & \cdots & \alpha_p^{11,2G,t} & \cdots & \alpha_p^{1N,21,t} & \cdots & \alpha_p^{1N,2G,t}\\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots\\ \alpha_p^{11,G1,t} & \cdots & \alpha_p^{11,GG,t} & \cdots & \alpha_p^{1N,G1,t} & \cdots & \alpha_p^{1N,GG,t}\end{pmatrix}}_{G\times NG}.$$

A.2 Combination Check for CSH Restriction

In order to check all possible combination of CSH, we adopt the restriction selection matrices introduced by Koop and Korobilis (2016) to determine $\Gamma_{i,j}$:

$$\Gamma = \prod_{i=1}^{N-1}\prod_{j=i+1}^{N}\Gamma_{i,j}. \tag{A.2}$$

$\Gamma_{i,j}$ is the selection combination matrix for $\gamma_{ij} \in \{0,1\}$ and $(1-\gamma_{ij})$ between countries i and j. In more detail, each matrix Γ_{ij} has a dimension of $\left[N^2 \times N^2\right]$. The diagonal element located at $\mathbf{A}_p^{ii}$ is set to a value of $\gamma_{p,ij}^{CSH}$, while the off-diagonal element corresponding to $\mathbf{A}_p^{jj}$ is assigned a value of $\left(1-\gamma_{p,ij}^{CSH}\right)$. If every pair of countries i and j is heterogeneous, meaning $\gamma_{ij}^{CSH}=1$, then Γ_{ij} will be an identity matrix. As an example, consider a PVAR(1) model involving two variables (G=2) from three countries (N=3). According to the definition provided in equation (2.4), the coefficient matrix is shown below:

$$A = \begin{bmatrix}A^1\\ A^2\\ A^3\end{bmatrix} = \begin{bmatrix}A^{11} & A^{12} & A^{13}\\ A^{21} & A^{22} & A^{23}\\ A^{31} & A^{32} & A^{33}\end{bmatrix}. \tag{A.3}$$

To achieve shrinkage of the homogeneous coefficients using GAS (shrink towards the group average $\bar{\mathbf{A}}$), we have $\mathbf{A}-\bar{\mathbf{A}}$. If $\gamma_{ij}=0$, this indicates cross-sectional homogeneity, which implies

that $\mathbf{A} - \bar{\mathbf{A}} = 0$ or equivalently, $\mathbf{A} = \bar{\mathbf{A}}$. The matrix $\bar{\mathbf{A}}$ is given by:

$$\bar{\mathbf{A}} = \begin{bmatrix} \bar{\mathbf{A}} & 0 & 0 \\ 0 & \bar{\mathbf{A}} & 0 \\ 0 & 0 & \bar{\mathbf{A}} \end{bmatrix}. \tag{A.4}$$

In this example, the dimensions of the restriction matrix of Γ_{ij} are shown in equation (A.5) are $[9 \times 9]$. To verify cross-sectional homogeneity between country one and country three, where $\mathrm{A}^{11} = \mathrm{A}^{33}$, we perform the following steps: If homogeneity is confirmed, indicated by $\gamma_{13}^{CSH} = 0$, the element located in the first row and first column of equation (A.5), A^{11}, is replaced by $\bar{\mathbf{A}}$. This implies $\mathrm{A}^{11} - \bar{\mathbf{A}} = 0$, effectively setting A^{11} to the average value $\bar{\mathbf{A}}$. Simultaneously, the element in the ninth row and first column equation (A.5), which is $1 - \gamma_{13}^{CSH} = 1$ and corresponds to the position of $\mathrm{A}^{33} - \bar{\mathbf{A}}$ in the vectorized matrix $\mathbf{A}$, is also replaced by $\bar{\mathbf{A}}$. This procedure ensures that A^{11} and A^{33} are equal to $\bar{\mathbf{A}}$, achieving the intended shrinkage. If $\gamma_{13}^{CSH} = 1$, it denotes cross-sectional heterogeneity, suggesting different behaviors between the two countries.

$$\Gamma_{1,3}A = \begin{bmatrix} \gamma_{13}^{CSH} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-\gamma_{13}^{CSH} \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} A^{11} - \bar{\mathbf{A}} \\ A^{12} \\ A^{13} \\ A^{21} \\ A^{22} - \bar{\mathbf{A}} \\ A^{23} \\ A^{31} \\ A^{32} \\ A^{33} - \bar{\mathbf{A}} \end{bmatrix} = \begin{bmatrix} A^{33} - \bar{\mathbf{A}} \\ A^{12} \\ A^{13} \\ A^{21} \\ A^{22} - \bar{\mathbf{A}} \\ A^{23} \\ A^{31} \\ A^{32} \\ A^{33} - \bar{\mathbf{A}} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{A}} \\ A^{12} \\ A^{13} \\ A^{21} \\ A^{22} \\ A^{23} \\ A^{31} \\ A^{32} \\ \bar{\mathbf{A}}. \end{bmatrix} \tag{A.5}$$

Using the same logic to verify cross-sectional homogeneity, if we intend to shrink A^{11} toward A^{33} with the same value, we can directly replace A^{11} with A^{33} in the vectorized matrix A, rather than replacing both A^{11} and A^{33} with $\bar{\mathbf{A}}$.

A.3 Additional Derivation

A.3.1 Derivation of DI Penalty 2.21

For $\text{pen}_{DI}\left(\mathbf{A}_p^{ij}\right) = \log\left(\frac{\pi(\mathbf{A}_p^{ij}|\gamma_{ij}^{DI})}{\pi(\mathbf{0}|\gamma)}\right) = -\lambda_1 \mid \mathbf{A}_p^{ij} \mid + \log\left(\frac{\mathbb{P}_{\gamma^{DI}}(\mathbf{0})}{\mathbb{P}_{\gamma_{ij}^{DI}}(\mathbf{A}_p^{ij})}\right)$, which can be simplified through the following steps:

$$
\begin{aligned}
\log\left(\frac{\pi(\mathbf{A}_p^{ij}|\gamma_{ij}^{DI})}{\pi(\mathbf{0}|\gamma)}\right) &= \log\left(\frac{\gamma_{ij}^{DI}\Phi_1\left(\mathbf{A}_p^{ij}\right) + (1-\gamma_{ij}^{DI})\Phi_0\left(\mathbf{A}_p^{ij}\right)}{\gamma_{ij}^{DI}\Phi_1(\mathbf{0}) + (1-\gamma_{ij}^{DI})\Phi_0(\mathbf{0})}\right) \\
&= \log\left(\frac{\gamma_{ij}^{DI}\frac{\lambda_1}{2}\exp(-\lambda_1|\mathbf{A}_p^{ij}|) + (1-\gamma_{ij}^{DI})\frac{\lambda_0}{2}\exp(-\lambda_0|\mathbf{A}_p^{ij}|)}{\gamma_{ij}^{DI}\Phi_1(\mathbf{0}) + (1-\gamma_{ij}^{DI})\Phi_0(\mathbf{0})}\right) \\
&= \log\left(\frac{\gamma_{ij}^{DI}\Phi_1\left(\mathbf{A}_p^{ij}\right) + (1-\gamma_{ij}^{DI})\Phi_0\left(\mathbf{A}_p^{ij}\right)}{\gamma_{ij}^{DI}\Phi_1(\mathbf{0}) + (1-\gamma_{ij}^{DI})\Phi_0(\mathbf{0})}\frac{\gamma_{ij}\frac{\lambda_1}{2}}{\gamma_{ij}\frac{\lambda_1}{2}\exp(-\lambda_1|\mathbf{A}_p^{ij}|)}\exp(-\lambda_1|\mathbf{A}_p^{ij}|)\right) \\
&= \log\left(\frac{\gamma_{ij}^{DI}\Phi_1\left(\mathbf{A}_p^{ij}\right) + (1-\gamma_{ij}^{DI})\Phi_0\left(\mathbf{A}_p^{ij}\right)}{\gamma_{ij}^{DI}\Phi_1(\mathbf{0}) + (1-\gamma_{ij}^{DI})\Phi_0(\mathbf{0})}\frac{\gamma_{ij}\Phi_1(\mathbf{0})}{\gamma_{ij}\Phi_1\left(\mathbf{A}_p^{ij}\right)}\exp(-\lambda_1|\mathbf{A}_p^{ij}|)\right) \\
&= \log\left(\frac{\gamma_{ij}^{DI}\Phi_1(\mathbf{0})\left(\gamma_{ij}^{DI}\Phi_1\left(\mathbf{A}_p^{ij}\right) + (1-\gamma_{ij}^{DI})\Phi_0\left(\mathbf{A}_p^{ij}\right)\right)}{\left(\gamma_{ij}^{DI}\Phi_1(\mathbf{0}) + (1-\gamma_{ij}^{DI})\Phi_0(\mathbf{0})\right)\gamma_{ij}\Phi_1\left(\mathbf{A}_p^{ij}\right)}\exp(-\lambda_1|\mathbf{A}_p^{ij}|)\right) \\
&= -\lambda_1|\mathbf{A}_p^{ij}| + \log\left(\frac{\gamma_{ij}^{DI}\Phi_1(\mathbf{0})\left(\gamma_{ij}^{DI}\Phi_1\left(\mathbf{A}_p^{ij}\right) + (1-\gamma_{ij}^{DI})\Phi_0\left(\mathbf{A}_p^{ij}\right)\right)}{\left(\gamma_{ij}^{DI}\Phi_1(\mathbf{0}) + (1-\gamma_{ij}^{DI})\Phi_0(\mathbf{0})\right)\gamma_{ij}\Phi_1\left(\mathbf{A}_p^{ij}\right)}\right) \\
&= -\lambda_1|\mathbf{A}_p^{ij}| + \log\left(\frac{\gamma_{ij}^{DI}\Phi_1(\mathbf{0})}{\left(\gamma_{ij}^{DI}\Phi_1(\mathbf{0}) + (1-\gamma_{ij}^{DI})\Phi_0(\mathbf{0})\right)}\frac{\left(\gamma_{ij}^{DI}\Phi_1\left(\mathbf{A}_p^{ij}\right) + (1-\gamma_{ij}^{DI})\Phi_0\left(\mathbf{A}_p^{ij}\right)\right)}{\gamma_{ij}\Phi_1\left(\mathbf{A}_p^{ij}\right)}\right) \\
&= -\lambda_1|\mathbf{A}_p^{ij}| + \log\left(\frac{\mathbb{P}_{\gamma^{DI}}(\mathbf{0})}{\mathbb{P}_{\gamma_{ij}^{DI}}(\mathbf{A}_p^{ij})}\right)
\end{aligned}
$$

A.3.2 Conditional Probability Function

Derivation of $\mathbb{P}_{\gamma_{ij}^{DI}}\left(\mathbf{A}_p^{ij}\right)$

Given $\mathbb{P}_{\gamma_{ij}^{DI}}\left(\mathbf{A}_p^{ij}\right) = \mathbb{P}_{\gamma_{ij}^{DI}}\left(\mathbf{A}_p^{ij}\right)^{-1} = \left(\frac{(1-\gamma_{ij}^{DI})\Phi_0(\mathbf{A}_p^{ij}) + \gamma_{ij}^{DI}\Phi_1(\mathbf{A}_p^{ij})}{\gamma_{ij}^{DI}\Phi_1(\mathbf{A}_p^{ij})}\right)^{-1}$, we have

$$
\begin{aligned}
\left(\mathbb{P}_{\gamma_{ij}^{DI}}\left(\mathbf{A}_p^{ij}\right)\right)^{-1} &= \frac{(1-\gamma_{ij}^{DI})\frac{\lambda_0}{2}\exp\left(-\lambda_0|\mathbf{A}_p^{ij}|\right) + \gamma_{ij}^{DI}\frac{\lambda_1}{2}\exp\left(-\lambda_1|\mathbf{A}_p^{ij}|\right)}{\gamma_{ij}^{DI}\frac{\lambda_1}{2}\exp\left(-\lambda_1|\mathbf{A}_p^{ij}|\right)} \\
&= 1 + \frac{(1-\gamma_{ij}^{DI})\frac{\lambda_0}{2}\exp\left(-\lambda_0|\mathbf{A}_p^{ij}|\right)}{\gamma_{ij}^{DI}\frac{\lambda_1}{2}\exp\left(-\lambda_1|\mathbf{A}_p^{ij}|\right)} \\
&= 1 + \left(\frac{1-\gamma_{ij}^{DI}}{\gamma_{ij}^{DI}}\right)\frac{\lambda_0}{\lambda_1}\exp\left(-|\mathbf{A}_p^{ij}|\left(\lambda_0-\lambda_1\right)\right). \qquad \text{(A.6)}
\end{aligned}
$$

Derivation of $\mathbb{P}_{\gamma_{ij}^{CSH}}\left(\Delta\mathbf{A}_p^{ij}\right)$

Given $\mathbb{P}_{\gamma_{ij}^{CSH}}\left(\Delta\mathbf{A}_p^{ij}\right) = \left(\mathbb{P}_{\gamma_{ij}^{CSH}}\left(\Delta\mathbf{A}_p^{ij}\right)\right)^{-1} = \left(\frac{(1-\gamma_{ij}^{CSH})\Phi_0\left(\Delta\mathbf{A}_p^{ij}\right)+\gamma_{ij}^{CSH}\Phi_1\left(\Delta\mathbf{A}_p^{ij}\right)}{\gamma_{ij}^{CSH}\Phi_1\left(\Delta\mathbf{A}_p^{ij}\right)}\right)^{-1}$, we have

$$\begin{aligned}
\left(\mathbb{P}_{\gamma_{ij}^{CSH}}\left(\Delta\mathbf{A}_p^{ij}\right)\right)^{-1} &= \frac{(1-\gamma_{ij}^{CSH})\frac{\eta_0}{2}\exp\left(-\eta_0|\Delta\mathbf{A}_p^{ij}|\right) + \gamma_{ij}^{CSH}\frac{\eta_1}{2}\exp\left(-\eta_1|\Delta\mathbf{A}_p^{ij}|\right)}{\gamma_{ij}^{CSH}\frac{\eta_1}{2}\exp\left(-\eta_1|\Delta\mathbf{A}_p^{ij}|\right)} \\
&= 1 + \frac{(1-\gamma_{ij}^{CSH})\frac{\eta_0}{2}\exp\left(-\eta_0|\Delta\mathbf{A}_p^{ij}|\right)}{\gamma_{ij}^{CSH}\frac{\eta_1}{2}\exp\left(-\eta_1|\Delta\mathbf{A}_p^{ij}|\right)} \\
&= 1 + \left(\frac{1-\gamma_{ij}^{CSH}}{\gamma_{ij}^{CSH}}\right)\frac{\eta_0}{\eta_1}\exp\left(-|\Delta\mathbf{A}_p^{ij}|\left(\eta_0-\eta_1\right)\right). \qquad \text{(A.7)}
\end{aligned}$$

A.4 Proof of DI and CSH Penalty

Proof of DI Penalty Equation 2.24. We begin by writing the penalty as

$$\text{pen}_{DI}\left(\mathbf{A} \mid \gamma^{DI}\right) = \log\Big[\pi(\mathbf{A} \mid \gamma^{DI})\Big] - \underbrace{\log\Big[\pi(\mathbf{0} \mid \gamma^{DI})\Big]}_{\text{constant w.r.t } |\mathbf{A}_p^{ij}|}.$$

The derivative can be written as follows:

$$\begin{aligned}
\frac{\partial \text{pen}_{DI}(\mathbf{A}|\gamma^{\mathbf{DI}})}{\partial|\mathbf{A}_p^{ij}|} &= \frac{\partial}{\partial|\mathbf{A}_p^{ij}|}\log\Big[\pi(\mathbf{A}_p^{ij})\Big] = \frac{1}{\pi(\mathbf{A}_p^{ij})}\frac{\partial}{\partial|\mathbf{A}_p^{ij}|}\pi(\mathbf{A}_p^{ij}) \\
&= \frac{1}{\pi(\mathbf{A}_p^{ij})}\left(\gamma^{DI}\frac{\lambda_1}{2}\left(-\lambda_1\right)\exp\Big[-\lambda_1|\mathbf{A}_p^{ij}|\Big] + (1-\gamma^{DI})\frac{\lambda_0}{2}\left(-\lambda_0\right)\exp\Big[-\lambda_0\,|\mathbf{A}_p^{ij}|\Big]\right) \\
&= -\frac{\left(\gamma^{DI}\frac{\lambda_1^2}{2}\exp\left[-\lambda_1|\mathbf{A}_p^{ij}|\right] + (1-\gamma^{DI})\frac{\lambda_0^2}{2}\exp\left[-\lambda_0|\mathbf{A}_p^{ij}|\right]\right)}{\left(\gamma^{DI}\frac{\lambda_1}{2}\exp\left[-\lambda_1|\mathbf{A}_p^{ij}|\right] + (1-\gamma^{DI})\frac{\lambda_0}{2}\exp\left[-\lambda_0|\mathbf{A}_p^{ij}|\right]\right)} \\
&= -\left(\lambda_1 w(\mathbf{A}_p^{ij}) + \lambda_0\left(1-w(\mathbf{A}_p^{ij})\right)\right) \\
&= -\lambda^*_{\gamma^{DI}}(\mathbf{A}_p^{ij}). \qquad \text{(A.8)}
\end{aligned}$$

Proof of CSH Penalty Equation 2.27. Penalty can be written as:

$$\text{pen}_{CSH}\left(\Delta\mathbf{A} \mid \gamma^{CSH}\right) = \log\Big[\pi(\Delta\mathbf{A} \mid \gamma^{CSH})\Big] - \underbrace{\log\Big[\pi(\mathbf{0} \mid \gamma^{CSH})\Big]}_{\text{constant w.r.t } |\mathbf{A}_p^{ij}|}.$$

$$
\begin{aligned}
\frac{\partial \text{pen}_{CSH}\left(\Delta\mathbf{A} \mid \gamma^{CSH}\right)}{\partial|\Delta\mathbf{A}_p^{ij}|} &= \frac{\partial}{\partial|\Delta\mathbf{A}_p^{ij}|}\log\left[\pi(\Delta\mathbf{A}_p^{ij})\right] = \frac{1}{\pi(\Delta\mathbf{A}_p^{ij})}\frac{\partial}{\partial|\Delta\mathbf{A}_p^{ij}|}\pi(\Delta\mathbf{A}_p^{ij}) \\
&= \frac{1}{\pi(\Delta\mathbf{A}_p^{ij})}\left(-\frac{\gamma^{CSH}\eta_1^2}{2}\exp\left(-\eta_1|\Delta\mathbf{A}_p^{ij}|\right) - \frac{(1-\gamma^{CSH})\eta_0^2}{2}\exp\left(-\eta_0|\Delta\mathbf{A}_p^{ij}|\right)\right) \\
&= -\frac{\left(\gamma^{CSH}\frac{\eta_1^2}{2}\exp\left[-\eta_1|\Delta\mathbf{A}_p^{ij}|\right] + (1-\gamma^{CSH})\frac{\eta_0^2}{2}\exp\left[-\eta_0|\Delta\mathbf{A}_p^{ij}|\right]\right)}{\left(\gamma^{CSH}\frac{\eta_1}{2}\exp\left[-\eta_1|\Delta\mathbf{A}_p^{ij}|\right] + (1-\gamma^{CSH})\frac{\eta_0}{2}\exp\left[-\eta_0|\Delta\mathbf{A}_p^{ij}|\right]\right)} \\
&= -\left(\eta_1 w(\Delta\mathbf{A}_p^{ij}) + \eta_0\left(1-w(\Delta\mathbf{A}_p^{ij})\right)\right) \\
&= -\eta_\gamma^*(\mathbf{A}_p^{ij}). \qquad \text{(A.9)}
\end{aligned}
$$

B GVS and GAS Full Conditional Posteriors

Given with $\lambda_0, \lambda_1, \eta_0, \eta_1$, the joint posterior probability density function of $\mathbf{A}, \tau_0, \tau_1, \kappa_0, \kappa_1, \Sigma$ given $\mathbf{Y}_t, \mathbf{Z}_t$ is

$$
\begin{aligned}
\pi\left(\mathbf{A}, \Sigma \mid \mathbf{Y}_t, \mathbf{Z}_t\right) &\propto \pi(\mathbf{Y}_t \mid \Sigma)\pi(\Sigma)\prod_{i\neq j}^{R}\prod_{p=1}^{P}\pi\left(\mathbf{A}_p^{ij}|\gamma_{p,ij}^{DI}\right)\prod_{i=j}^{C}\prod_{p=1}^{P}\pi\left(\mathbf{A}_p^{ij}|\gamma_{p,ij}^{CSH}\right) \\
&= \pi(\mathbf{Y}_t \mid \Sigma)\pi(\Sigma)\Bigg(\prod_{i\neq j}^{R}\prod_{p=1}^{P}\left[\gamma^{DI}\Phi_1\left(\mathbf{A}_p^{ij}|\lambda_1\right) + (1-\gamma^{DI})\Phi_0\left(\mathbf{A}_p^{ij}|\lambda_0\right)\right] \qquad \text{(B.1)} \\
&\times \prod_{i=j}^{C}\prod_{p=1}^{P}\left[\gamma^{CSH}\Phi_1\left(\left(\mathbf{A}_p^{ij}-\bar{\mathrm{A}}\right)|\eta_1\right) + (1-\gamma^{CSH})\Phi_0\left(\left(\mathbf{A}_p^{ij}-\bar{\mathrm{A}}\right)|\eta_0\right)\right]\Bigg)
\end{aligned}
$$

Then we can generate from the posterior distribution using the following full conditional posteriors,

$$
\begin{aligned}
\mathbf{A}_p^{ij} \mid \text{rest} &\sim \mathcal{N}\left(\Gamma\mu_A, V_A\right) \\
\left(\frac{1}{\tau_{0,ij}}\right)^2 \mid \text{rest} &\sim \text{IGauss}\left(\sqrt{\frac{\Sigma_{ij}\lambda_0^2}{\left(\mathbf{A}_p^{ij}\right)^2}}, \lambda_0^2\right) \\
\left(\frac{1}{\tau_{1,ij}}\right)^2 \mid \text{rest} &\sim \text{IGauss}\left(\sqrt{\frac{\Sigma_{ij}\lambda_1^2}{\left(\mathbf{A}_p^{ij}\right)^2}}, \lambda_1^2\right) \\
\left(\frac{1}{\kappa_{0,ii}}\right)^2 \mid \text{rest} &\sim \text{IGauss}\left(\sqrt{\frac{\Sigma_{ii}\eta_0^2}{\left(\mathbf{A}_p^{ii}-\bar{\mathbf{A}}\right)^2}}, \eta_0^2\right) \\
\left(\frac{1}{\kappa_{1,ii}}\right)^2 \mid \text{rest} &\sim \text{IGauss}\left(\sqrt{\frac{\Sigma_{ii}\eta_1^2}{\left(\mathbf{A}_p^{ii}-\bar{\mathbf{A}}\right)^2}}, \eta_1^2\right) \\
\Sigma^{-1} &\sim \mathcal{W}(S, v),
\end{aligned}
$$

whereas

- n is total regression coefficients defining as $p \times NG \times NG$.
- $S = \text{SSE} + I$
- $\nu = T$ is the sampler size (or posterior degrees of freedom))

Estimation of μ_A and V_A: We define

$$V_A = \left(\left(D'D \right)^{-1} + \Sigma \otimes Z_t' Z_t \right)^{-1}, \tag{B.2}$$

and μ_A as

$$\mu_A = V_A \left(\left(\Sigma^{-1} \otimes Z_t' Z_t \right) \mathbf{A}_{OLS} \right). \tag{B.3}$$

D is a diagonal matrix with $D = \text{diag}(h_{11}, \cdots, h_{NN})$, and h_{ij} is defined as follows:

$$h_{ij} = \begin{cases} \tau_{0,ij}^2 & \text{if } \gamma_{ij}^{DI} = 0 \\ \tau_{1,ij}^2 & \text{if } \gamma_{ij}^{DI} = 1 \end{cases}$$

and

$$h_{ii} = \begin{cases} \kappa_{0,ii}^2 & \text{if } \gamma_{ii}^{CSH} = 0 \\ \kappa_{1,ii}^2 & \text{if } \gamma_{ii}^{CSH} = 1 \end{cases}.$$

Estimation of $\lambda_0, \lambda_1, \eta_0, \eta_1$:

- **Monte Carlo EM:**

$$\lambda_0^{(k)} = \sqrt{\frac{2r}{\sum_{p=1}^{P} \sum_{i \neq j}^{R} \mathbb{E}_{\lambda_0^{(k-1)}} \left[\tau_{0,ij}^2 | \tilde{Y}_{ij} \right]}},$$

and

$$\eta_0^{(k)} = \sqrt{\frac{2c}{\sum_{p=1}^{P} \sum_{i=j}^{C} \mathbb{E}_{\eta_0^{(k-1)}} \left[\kappa_{0,ij}^2 | \tilde{Y}_{ij} \right]}},$$

- **Gamma priors on λ^2**

$$\lambda_0^2 \mid \text{rest} \sim \text{Gamma} \left(n + r, \frac{1}{2} \sum_{i \neq j}^{R} \tau_{0,ij}^2 + \delta \right),$$

$$\eta_0^2 \mid \text{rest} \sim \text{Gamma} \left(C + r, \frac{1}{2} \sum_{i \neq j}^{C} \kappa_{0,ij}^2 + \delta \right).$$

Estimation of $\gamma^{DI}, \gamma^{CSH}$: $\gamma^{DI}, \gamma^{CSH}$ are update from Bernoulli distribution:

$$\gamma_{ij}^{DI} \sim \text{Bernoulli}\left(\theta_{ij}^{DI}\right)$$
$$\gamma_{ij}^{CSH} \sim \text{Bernoulli}\left(\theta_{ij}^{CSH}\right)$$

C Prior Hyperparameters: MCMC Algorithm

The prior hyperparameters of the model include the initial value of λ_0, λ_1, η_0, η_1. These initial values are slightly different between the Monte Carlo simulations and the empirical application considering the differences in PVAR size and data structure. However, because the penalization parameters are iteratively updated via the EM algorithm, the choice of initial values has limited influence on the final results.

In the simulation exercise, we set $\lambda_0, \eta_0 \in [1, 10, 50]$, and fix $\lambda_1^2 = \eta_1^2 = 1$. For data-based simulation and also the empirical applications, to better capture the characteristics of real-world data, the initial values are set following the formula suggested by Park and Casella (2008):

$$\lambda_0^0 = r \cdot \frac{\sqrt{\hat{\Sigma}^{\text{OLS}}}}{\sum_{p=1}^{P} \sum_{i \neq j}^{R} \left| \hat{A} ij^{\text{OLS},(p)} \right|}, \quad \lambda_1 = \zeta_{DI},$$

$$\eta_0^0 = r \cdot \frac{\sqrt{\hat{\Sigma}^{\text{OLS}}}}{\sum p = 1^{P} \sum_{i=j}^{C} \left| \hat{A}_{ij}^{\text{OLS},(p)} \right|}, \quad \eta_1 = \zeta_{CSH}.$$

Here, $\hat{A}^{\text{OLS}}$ and $\hat{\Sigma}^{\text{OLS}}$ denote the OLS estimators of the coefficient matrices and the residual covariance matrix, respectively. The scaling ensures that the spike penalties λ_0 and η_0 reflect the magnitude and sparsity of the data.

The slab penalties λ_1 and η_1 are initialized to be smaller than the corresponding spike penalties, typically satisfying:

$$\lambda_1 = \delta_\lambda \cdot \lambda_0, \quad \eta_1 = \delta_\eta \cdot \eta_0, \quad \text{where } \delta_\lambda, \delta_\eta \in [10^{-1}, 10^{-4}]. \tag{C.1}$$

This choice reflects the intuition that the slab component should allow for large, non-zero coefficient with relatively mild shrinkage, while the spike component enforces strong shrinkage on irrelevant coefficients. In practice, the scaling factor δ_η and δ_η can be chosen based on the data structure, typically in the range of 0.001 to 0.1.

Once initialized, the slab component can be updated in one of two ways: (i) they can be maintained as fixed proportions of the updating spike penalties (i.e., updated at each iteration as $\lambda_1^{(k)} = \delta_\lambda \cdot \lambda_0^{(k)}$ and $\eta_1^{(k)} = \delta_\eta \cdot \eta_0^{(k)}$) or (ii) they can be updated from the fully data-driven

EM step:

$$\lambda_1^{(k)} = \sqrt{\frac{2r}{\sum_{p=1}^{P}\sum_{i\neq j}^{R}\mathbb{E}_{\lambda_1^{(k-1)}}\left[\tau_{1,ij}^2 \mid \tilde{Y}_{ij}\right]}}, \quad \eta_1^{(k)} = \sqrt{\frac{2c}{\sum_{p=1}^{P}\sum_{i=j}^{C}\mathbb{E}_{\eta_1^{(k-1)}}\left[\kappa_{1,ij}^2 \mid \tilde{Y}_{ij}\right]}}.$$

D Additional Simulation Results

This appendix provides supplementary results to complement the main findings of the paper. We first present additional simulation results that assess the finite-sample properties of the proposed estimators. Specifically, we report bias across increasing sample sizes to provide evidence of the consistency of the estimators. In addition, we evaluate variable selection performance in terms of sensitivity, specificity, and accuracy for sample sizes $N = 3$ and $N = 5$. Second, we present additional forecasting results based on data-based simulations for settings for three and five countries.

D.1 Bias of Estimators

Table D.1 reports the bias of the estimators across DGP1–DGP3, for increasing sample sizes ($N = 3$, $N = 5$, $N = 10$) while holding T fixed. Across all settings, the proposed GAS and GVS estimators show near-zero bias, and this bias remains stable or slightly increases as the dimensionality (N) increases. These results provide empirical support for the finite-sample accuracy and consistency of our estimator as the panel dimension increases.

Table D.1: Bias for Different Estimators

DGP	**Our Methods**		**Bayesian Methods**			**Lasso**
	GAS	**GVS**	**BL**	**BFL**	**S4**	$\boldsymbol{\lambda, c, \alpha, \Sigma}$
DGP1						
$N = 3$	0.00	0.00	0.02	0.02	0.03	0.02
$N = 5$	0.00	0.00	0.11	0.10	0.03	0.01
$N = 10$	0.00	0.01	0.09	0.09	0.03	0.01
DGP2						
$N = 3$	0.00	0.04	0.04	0.04	0.03	0.02
$N = 5$	0.00	0.00	0.09	0.10	0.03	0.01
$N = 10$	0.00	0.00	0.07	0.06	0.03	0.01
DGP3						
$N = 3$	0.00	0.00	0.00	0.00	0.00	0.03
$N = 5$	0.00	0.00	0.03	0.03	0.01	0.00
$N = 10$	0.00	0.02	0.00	0.01	0.00	0.00

Note: Reported bias values by DGP and method. Lower bias values indicate more accurate estimation. Methods include GAS, GVS (proposed), Bayesian Lasso (BL), Bayesian Fused Lasso (BFL), S4, and Lasso with DI and CSH penalization (frequentist).

D.2 Variable Selection Performance

Tables D.2 and D.3 summarize the sensitivity (SEN), specificity (SP), and accuracy (ACC) of different methods under $N = 3$ and $N = 5$, respectively. The proposed methods (GAS and GVS) generally achieve higher overall accuracy across different DGPs compared to both Bayesian and frequentist benchmarks. This highlights the strong selection performance of our approach in small-sample settings.

Table D.2: Sensitivity (SEN), Specificity (SP), and Accuracy (ACC) for Different Estimators Across DGPs: N = 3

	DGP1			**DGP2**			**DGP3**		
Method	**SE**	**SP**	**AC**	**SE**	**SP**	**AC**	**SE**	**SP**	**AC**
Proposed Methods									
GAS	86.85%	51.28%	71.04%	75.03%	47.78%	61.40%	56.56%	0.00%	56.56%
GVS	88.85%	50.94%	72.00%	82.25%	49.17%	65.71%	56.42%	0.00%	56.42%
Bayesian Methods									
BL	88.20%	49.34%	70.93%	80.83%	49.19%	65.01%	56.94%	0.00%	56.94%
BFL	88.25%	49.44%	71.00%	80.89%	50.31%	65.60%	57.40%	0.00%	57.40%
S4	88.72%	48.28%	70.75%	82.00%	50.50%	66.25%	59.24%	0.00%	59.24%
Lasso									
$\lambda, c, \alpha, \Sigma$	52.87%	31.31%	43.29%	45.08%	38.81%	41.94%	43.74%	0.00%	43.74%

Note: The table presents Sensitivity (SEN), Specificity (SP), and Accuracy (ACC) for different estimation methods across DGPs, using N = 3. The best accuracy values for each DGP are highlighted in bold. The Lasso setting $\lambda, c, \alpha, \Sigma$ refers to DI and CSH penalization also with a penalization in Σ.

Table D.3: Sensitivity (SEN), Specificity (SP), and Accuracy (ACC) for Different Estimators Across DGPs: N = 5

	DGP1			**DGP2**			**DGP3**		
Method	**SE**	**SP**	**AC**	**SE**	**SP**	**AC**	**SE**	**SP**	**AC**
Proposed Methods									
GAS	90.89%	49.86%	69.14%	88.09%	49.49%	66.47%	55.82%	0.00%	55.82%
GVS	90.63%	49.44%	68.80%	88.26%	48.62%	66.06%	56.85%	0.00%	56.85%
Bayesian Methods									
BL	85.39%	47.49%	65.30%	82.89%	31.32%	54.01%	55.85%	0.00%	55.85%
BFL	85.82%	47.58%	65.55%	83.34%	31.27%	54.18%	55.91%	0.00%	55.91%
S4	90.20%	49.12%	68.43%	87.45%	41.42%	61.67%	57.38%	0.00%	57.38%
Lasso									
$\lambda, c, \alpha, \Sigma$	43.61%	25.58%	34.05%	39.15%	29.96%	34.00%	46.56%	0.00%	46.56%

Note: The table presents Sensitivity (SEN), Specificity (SP), and Accuracy (ACC) for different estimation methods across four DGPs. The best specificity and accuracy values for each DGP are highlighted in bold. The Lasso setting $\lambda, c, \alpha, \Sigma$ refers to DI and CSH penalization also with a penalization in Σ.

E Posterior Probability and Strength of Evidence

Table E.1 summarizes the interpretation of posterior probabilities in terms of the qualitative strength of evidence.

Table E.1: Posterior Probability and Strength of Evidence

Posterior Probability	**Evidence Strength**
$\geq 99\%$	Very strong
95–98%	Strong
85–94%	Moderate
75–84%	Weak to moderate
65–74%	Weak
50–64%	Minimal
$\leq 50\%$	No evidence

F Additional Application Results

This appendix provides additional results from the empirical analysis. First, it reports the posterior probabilities of dynamic interdependencies across all directed country pairs in the sample, offering a probabilistic structure behind the network graphs presented in the main text. Second, it presents country-level relative Root of MSFEs, which offer further detail of the boxplot summary shown in Figure 4.2.

Table F.2 displays the posterior probabilities for the dynamic interdependencies directed toward core countries (Austria, Belgium, Finland, France, and the Netherlands), while Table F.1 reports the corresponding probabilities for dynamic interdependencies toward periphery countries (Greece, Ireland, Italy, Portugal, and Spain). Each entry reports the posterior probability that exists from the sending country (row) to the receiving country (column), denoted by $p(A^{ij} \neq 0)$. Values closer to one indicate stronger evidence for the existence of DI.

Table F.3 and Table F.4 report the country-level forecasting performance of each model relative to the benchmark. The benchmark model is the GVS, selected based on marginal log-likelihood from the model selection exercise. Positive values indicate that GVS outperforms the alternative model, while negative values suggest that GVS performs worse.

Table F.1: Posterior Probabilities of Dynamic Interdependencies To Periphery Countries

No	From	To	Prob.	No	From	To	Prob.
1	AUT	GRC	0.90	24	IRL	ITA	0.87
2	BEL	GRC	0.94	25	NLD	ITA	0.95
3	FIN	GRC	0.95	26	PRT	ITA	0.95
4	FRA	GRC	0.96	27	ESP	ITA	0.96
5	IRL	GRC	0.79	28	AUT	PRT	0.41
6	ITA	GRC	0.89	29	BEL	PRT	0.65
7	NLD	GRC	0.63	30	FIN	PRT	0.48
8	PRT	GRC	0.64	31	FRA	PRT	0.53
9	ESP	GRC	0.57	32	GRC	PRT	0.62
10	AUT	IRL	0.37	33	IRL	PRT	0.55
11	BEL	IRL	0.54	34	ITA	PRT	0.69
12	FIN	IRL	0.43	35	NLD	PRT	0.84
13	FRA	IRL	0.86	36	ESP	PRT	0.89
14	GRC	IRL	0.91	37	AUT	ESP	0.50
15	ITA	IRL	0.82	38	BEL	ESP	0.78
16	NLD	IRL	0.67	39	FIN	ESP	0.93
17	PRT	IRL	0.63	40	FRA	ESP	0.55
18	ESP	IRL	0.78	41	GRC	ESP	0.70
19	AUT	ITA	0.38	42	IRL	ESP	0.81
20	BEL	ITA	0.60	43	ITA	ESP	0.92
21	FIN	ITA	0.69	44	NLD	ESP	0.73
22	FRA	ITA	0.79	45	PRT	ESP	0.94
23	GRC	ITA	0.88				

Notes: The table reports posterior probabilities, $p(A^{ij} \neq 0)$, which represent the likelihood of dynamic interdependencies between countries. Probabilities closer to 1 indicate stronger interdependencies. The Core countries include Austria, Belgium, Finland, France, and the Netherlands, while the Periphery countries include Greece, Ireland, Italy, Portugal, and Spain.

Table F.2: Posterior Probabilities of Dynamic Interdependencies To Core Countries

No	From	To	Prob.	No	From	To	Prob.
1	BEL	AUT	0.96	24	ITA	FIN	0.36
2	FIN	AUT	0.96	25	NLD	FIN	0.56
3	FRA	AUT	0.96	26	PRT	FIN	0.58
4	GRC	AUT	0.29	27	ESP	FIN	0.54
5	IRL	AUT	0.30	28	AUT	FRA	0.81
6	ITA	AUT	0.30	29	BEL	FRA	0.85
7	NLD	AUT	0.45	30	FIN	FRA	0.78
8	PRT	AUT	0.36	31	GRC	FRA	0.96
9	ESP	AUT	0.37	32	IRL	FRA	0.95
10	AUT	BEL	0.94	33	ITA	FRA	0.96
11	FIN	BEL	0.85	34	NLD	FRA	0.96
12	FRA	BEL	0.89	35	PRT	FRA	0.72
13	GRC	BEL	0.53	36	ESP	FRA	0.93
14	IRL	BEL	0.50	37	AUT	NLD	0.91
15	ITA	BEL	0.55	38	BEL	NLD	0.90
16	NLD	BEL	0.72	39	FIN	NLD	0.95
17	PRT	BEL	0.62	40	FRA	NLD	0.96
18	ESP	BEL	0.55	41	GRC	NLD	0.77
19	AUT	FIN	0.79	42	IRL	NLD	0.67
20	BEL	FIN	0.93	43	ITA	NLD	0.78
21	FRA	FIN	0.89	44	PRT	NLD	0.87
22	GRC	FIN	0.34	45	ESP	NLD	0.89
23	IRL	FIN	0.30				

Notes: The table reports posterior probabilities, $p(A^{ij} \neq 0)$, which represent the likelihood of dynamic interdependencies between countries. Probabilities closer to 1 indicate stronger interdependencies. The Core countries include Austria, Belgium, Finland, France, and the Netherlands, while the Periphery countries include Greece, Ireland, Italy, Portugal, and Spain.

Table F.3: Country-Level Relative rMSFE at $h = 1$ and $h = 3$ (Normalized to GVS)

Country	PSSL		Bayesian Methods			Lasso
	GVS	**GAS**	**BL**	**BFL**	**S4**	$\lambda, c, \alpha, \Sigma$
Panel A: $h = 1$						
Austria	1	0.98	1.33	1.33	1.30	1.04
Belgium	1	3.65	18.64	17.87	10.63	9.74
Finland	1	0.92	8.19	8.10	6.98	7.32
France	1	0.95	0.98	0.99	0.93	1.45
Greece	1	0.94	0.74	0.77	1.20	0.61
Ireland	1	1.18	0.85	0.82	0.60	0.97
Italy	1	1.01	0.53	0.53	0.59	0.74
Netherlands	1	1.01	1.20	1.23	0.99	2.73
Portugal	1	1.16	1.29	1.29	1.20	1.45
Spain	1	1.10	1.48	1.48	1.30	1.44
United Kingdom	1	1.03	1.00	1.00	0.99	1.03
United States	1	1.01	1.02	1.02	1.05	1.21
Panel B: $h = 3$						
Austria	1	0.99	0.95	0.96	1.03	1.00
Belgium	1	1.09	1.00	1.00	1.13	1.09
Finland	1	1.35	2.02	2.01	1.65	1.34
France	1	1.17	2.35	2.36	2.13	1.35
Greece	1	0.92	0.78	0.79	0.91	1.03
Ireland	1	0.92	2.08	2.01	1.12	1.08
Italy	1	1.00	1.06	1.06	1.00	1.05
Netherlands	1	0.98	2.63	2.61	2.05	0.67
Portugal	1	1.31	1.61	1.62	1.59	1.75
Spain	1	1.48	7.22	7.18	5.83	5.17
United Kingdom	1	0.78	1.63	1.63	1.61	1.58
United States	1	1.16	1.73	1.72	1.42	1.10

Note: The table reports country-level relative root Mean Squared Forecast Errors (rMSFE) at $h = 6$ (Panel A) and $h = 12$ (Panel B). Entries are computed as $\text{rMSFE}_{\text{model}}/\text{rMSFE}_{\text{GVS}}$, where GVS is a benchmark forecast model. Methods include: PSSL with GAS, Bayesian shrinkage models (BL, BFL, S4), and a frequentist Lasso with shrinkage over hyperparameters λ, c, α, and Σ.

Table F.4: Country-Level Relative rMSFE at $h = 6$ and $h = 12$ (Normalized to GVS)

Country	PSSL		Bayesian Methods			Lasso
	GVS	**GAS**	**BL**	**BFL**	**S4**	$\lambda, c, \alpha, \Sigma$
Panel A: $h = 6$						
Austria	1	1.05	1.11	1.11	1.12	1.08
Belgium	1	1.28	1.36	1.29	1.07	1.85
Finland	1	1.08	1.26	1.26	1.20	1.10
France	1	1.11	2.17	2.17	1.88	1.41
Greece	1	1.07	0.89	0.88	0.99	1.01
Ireland	1	0.85	4.23	4.20	3.35	1.90
Italy	1	1.03	0.96	0.97	1.06	1.02
Netherlands	1	0.98	0.93	0.93	0.99	0.98
Portugal	1	1.13	0.63	0.64	0.91	1.03
Spain	1	1.01	0.79	0.79	0.81	0.86
United Kingdom	1	0.99	1.81	1.80	1.58	1.24
United States	1	0.97	2.83	2.83	2.18	1.43
Panel B: $h = 12$						
Country	**PSSL**		**Bayesian Methods**			**Lasso**
	GVS	**GAS**	**BL**	**BFL**	**S4**	$\lambda, c, \alpha, \Sigma$
Austria	1	1.00	0.97	0.97	0.98	0.99
Belgium	1	1.00	1.51	1.50	1.23	1.00
Finland	1	0.99	1.01	1.01	0.99	0.99
France	1	1.01	1.38	1.38	1.20	1.06
Greece	1	0.99	1.16	1.16	1.07	1.01
Ireland	1	0.92	16.57	16.44	8.95	4.06
Italy	1	1.03	1.71	1.69	1.28	1.00
Netherlands	1	1.00	1.03	1.03	1.01	1.00
Portugal	1	1.00	1.07	1.07	1.04	1.02
Spain	1	1.00	0.89	0.89	0.94	0.98
United Kingdom	1	1.00	1.01	1.01	1.01	1.00
United States	1	0.97	3.12	3.10	2.07	1.29

Note: The table reports country-level relative root Mean Squared Forecast Errors (rMSFE) at $h = 6$ (Panel A) and $h = 12$ (Panel B). Entries are computed as $\text{rMSFE}_{\text{model}}/\text{rMSFE}_{\text{GVS}}$, where GVS is a benchmark forecast model. Methods include: PSSL with GAS, Bayesian shrinkage models (BL, BFL, S4), and a frequentist Lasso with shrinkage over hyperparameters λ, c, α, and Σ.