Adelaide University

School of Economics

Working Papers

ISSN 2203-6024

Adaptive Bayesian Shrinkage of High-Dimensional Panel VARs

Zhiruo Zhang
Adelaide University

Firmin Doko Tchatoka
Adelaide University

Qazi Haque
Adelaide University

Working Paper No. 05-2026
May 2026

Copyright © The Author(s)

PRV 14404 / CRICOS 04249J

Adaptive Bayesian Shrinkage of High-Dimensional Panel VARs*

Zhiruo Zhang† Firmin Doko Tchatoka‡

Qazi Haque§

School of Economics, College of Business and Law,

Adelaide University, Australia

Abstract

We develop a Bayesian framework that combines adaptive shrinkage with variable selection to address over-parameterisation and sparsity in high-dimensional panel vector autoregressions (PVARs). The proposed approach employs Laplace-based spike-and-slab priors to enable flexible modelling of dynamic cross-sectional interdependencies and unit-specific heterogeneity. Monte Carlo evidence shows that the method delivers improvements in estimation accuracy and forecasting performance relative to existing regularisation approaches. We illustrate its empirical relevance in two applications. The first investigates financial contagion in euro area sovereign bond markets, while the second examines international forecasting performance in a multi-country macroeconomic panel. The results highlight the benefits of adaptive, component-specific shrinkage for capturing heterogeneous spillover structures in complex panel 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. Firmin Doko Tchatoka acknowledges funding from Discovery Grants DP200101498 and DP210103094. Qazi Haque acknowledges funding from Discovery Grant DP240100970. Additional computational resources were provided by the Phoenix High Performance Computing (HPC) service at the University of Adelaide.

†School of Economics, College of Business and Law, Adelaide University. 10 Pulteney Street, Adelaide SA 5005, Australia. Email: zhiruo.zhang@adelaide.edu.au

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

§School of Economics, College of Business and Law, Adelaide University. 10 Pulteney Street, Adelaide SA 5005, Australia. Email: qazi.haque@adelaide.edu.au

1 Introduction

Advances in data availability and statistical methodology have facilitated the increasing use of high-dimensional datasets in macroeconometric analysis.[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). This flexibility makes PVARs particularly useful for studying spillovers, contagion, and heterogeneous adjustment processes in interconnected economies. However, the empirical implementation of PVAR models is often constrained by a fundamental challenge: the dimensionality of the parameter space increases rapidly with the number of panel units, variables, and lags. As a result, unrestricted PVAR specifications are prone to overfitting, unstable inference, and weak forecasting performance. Addressing this high-dimensionality problem is therefore central to the reliable estimation and practical usefulness of panel VAR models.[2]

This paper proposes a Bayesian framework that combines adaptive shrinkage with explicit variable selection to address the problems of over-parameterisation and sparsity in high-dimensional panel systems. Since the seminal contribution of Park and Casella (2008), the Bayesian Lasso has become a standard tool for regularisation in high-dimensional settings. Casella et al. (2010) extend this framework by providing a Bayesian interpretation of the fused Lasso through the penalisation of successive coefficient differences. More recently, Ročková and George (2018) introduce the spike-and-slab Lasso (SSL), which integrates shrinkage and model selection within a unified Bayesian framework. By assigning each coefficient to either a tightly concentrated "spike'' component or a more diffuse "slab'' component, the SSL provides an effective mechanism for distinguishing economically meaningful signals from negligible effects. Building on these developments, we adapt the SSL methodology to the panel VAR context, enabling flexible estimation of high-dimensional macroeconomic systems characterised by heterogeneous dynamics and rich cross-unit interactions.

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

[2] Early approaches rely on shrinkage priors such as the Minnesota prior of Doan et al. (1984) and Litterman (1986). Extensions based on hierarchical structures are proposed by Jarociński (2010), while more recent contributions incorporate heterogeneous interdependencies across panel units (Canova and Ciccarelli, 2009, 2013; Koop and Korobilis, 2016, 2019).

We extend this framework by developing the *Panel Spike-and-Slab Lasso* (PSSL) approach. The proposed method partitions the parameter space into blocks associated with individual panel units and assigns each block to either a spike or slab component governed by Laplace distributions, following Ročková and George (2018) and Bai et al. (2022). This blockwise prior structure allows the model to simultaneously account for dynamic interdependencies across units and cross-sectional heterogeneity in own-lag dynamics. As a result, the estimator can adaptively distinguish between relevant and irrelevant interactions, promoting parsimony while preserving economically meaningful propagation channels.

Embedding the spike-and-slab structure within the PVAR framework enhances the model's ability to perform selective shrinkage. Through extensive Monte Carlo simulations spanning low-, moderate-, and high-dimensional environments with varying degrees of sparsity, we show that the proposed PSSL approach delivers substantial improvements in estimation accuracy and forecasting performance relative to several established alternatives. These include the Stochastic Search Specification Selection (S4) procedure of Koop and Korobilis (2016), which builds on the Stochastic Search Variable Selection framework of George et al. (2008), as well as frequentist Lasso, Bayesian Lasso, Bayesian fused Lasso, and global VAR benchmarks.

These gains reflect fundamental differences in prior design and shrinkage behaviour. Although both PSSL and S4 rely on spike-and-slab priors, S4 employs mixtures of normal distributions, whereas PSSL utilises Laplace-based priors that induce stronger shrinkage for coefficients close to zero while allowing economically meaningful effects to remain relatively unrestricted. Moreover, the joint estimation of the scale parameters governing the spike and slab components enables the degree of regularisation to adapt endogenously to the sparsity structure of the data. Relative to continuous shrinkage approaches such as frequentist and Bayesian Lasso, the PSSL framework allows explicit selection of dynamic interdependencies and provides posterior inclusion probabilities that facilitate structural interpretation. Compared with Bayesian fused Lasso, it is less restrictive in settings characterised by heterogeneous cross-sectional dynamics. It also offers a more direct mechanism for detecting sparse bilateral spillovers than factor-based models or global VAR specifications, which typically summarise dependence through latent common components or pre-specified weighted averages.

We illustrate the empirical relevance of the proposed approach using two macro-financial applications. The first revisits the analysis of financial contagion in euro area sovereign bond markets studied by Koop and Korobilis (2016). This application examines how shocks to sovereign bond yields propagate across countries and how the strength of transmission evolves over time. In contrast to approaches that rely on pre-determined shrinkage structures, our method allows the pattern of dynamic interdependencies and cross-country heterogeneity to be identified directly from the data. The results reveal a pronounced asymmetry in financial transmission within the euro area, with shocks originating in financially integrated core economies spreading widely to peripheral countries, while reverse transmission is comparatively limited. These findings point to a hierarchical network of spillovers in which core sovereign markets play a central role in shaping regional financial dynamics.

In a second application, we examine international macroeconomic forecasting using the multi-country dataset of Camehl (2023), which comprises twelve economies — ten euro area countries together with the United Kingdom and the United States. The objective is to forecast key macroeconomic variables, including consumer prices, industrial production, the real effective exchange rate, and unemployment, while accounting for potentially complex cross-country spillovers and heterogeneous dynamic linkages. The proposed model delivers accurate and stable forecasts across horizons and country groups, outperforming stochastic search selection procedures, Lasso-type shrinkage methods, and global VAR benchmarks. By employing adaptive, coefficient-specific regularisation, the PSSL approach mitigates the risk of large forecast errors while preserving economically meaningful heterogeneity. This feature is particularly valuable in international forecasting environments characterised by structural uncertainty and potential model misspecification.

Our contribution relates to a broader literature on shrinkage and sparsity in panel VAR estimation, which can be broadly classified into Bayesian and frequentist strands. Bayesian approaches — including factor-based models (Canova and Ciccarelli, 2013; Koop and Korobilis, 2019; Huber et al., 2023) and specification-selection methods (Koop and Korobilis, 2016; Korobilis, 2016) — offer flexible ways to incorporate shrinkage while imposing structural restrictions. However, factor-based methods may have limited ability to enforce true sparsity, as coefficients are often only weakly shrunk toward zero. In par-

allel, an extensive literature on frequentist penalisation methods builds on ℓ_1-regularised estimation (Tibshirani, 1996; Zhao and Yu, 2006; Zou, 2006; Yuan and Lin, 2006; Huang et al., 2008). Extensions to VAR settings include Bayesian Lasso approaches that account for temporal dependence and group-level sparsity (Song and Bickel, 2011; Nicholson et al., 2020). These contributions highlight the importance of regularisation in high-dimensional time-series models but typically rely on global shrinkage schemes that may not fully capture heterogeneous interdependencies across panel units.

The remainder of the paper is organised as follows. Section 2 introduces the baseline PVAR framework and presents the Panel Spike-and-Slab Lasso estimator. Section 3 reports Monte Carlo evidence on estimation and forecasting performance. Sections 4 and 5 provide empirical applications to financial contagion and international macroeconomic forecasting, respectively. Section 6 concludes.

2 Empirical Framework

2.1 General Specification

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 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. This specification allows for both within-unit dynamics ($i = j$, so unit i's own lags influence its future) and cross-unit dynamics ($i \neq j$, so lags of other units affect unit i). The error vector ε_{it} has unit-specific covariance matrix $\Sigma_{ii} \in \mathbb{R}^{G \times G}$.

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.[3]

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

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

[3]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 and instead allow for contemporaneous dependence flexibly through a fully unrestricted error covariance matrix.

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 Panel Spike-and-Slab Lasso

To flexibly accommodate both both dynamic inter-dependencies (DI) across units and cross-sectional heterogeneity (CSH) in own-lag dynamics, 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.

2.2.1 Dynamic Interdependency (DI) Prior

To model potential DI across units, 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:[4]

[4]Double-exponential is also known as Laplace.

$$\pi(\mathbf{A}_p^{ij} \mid \gamma_{p,ij}^{DI}) = (1 - \gamma_{p,ij}^{DI})\, \Phi_0 \left(\mathbf{A}_p^{ij} \mid \lambda_0\right) + \gamma_{p,ij}^{DI}\, \Phi_1 \left(\mathbf{A}_p^{ij} \mid \lambda_1\right), \tag{2.5}$$

where Φ_0 and Φ_1 are double-exponential (DE) densities with different scales. For each block of off-diagonal coefficients $\left\{\mathbf{A}_p^{ij}\right\}$ with $i \neq j$, we introduce a latent binary indicator $\gamma_{p,ij}^{DI} \in (0, 1)$. This indicator determines 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), and follows a Bernoulli distribution with prior:

$$\gamma_{p,ij}^{DI} \sim \text{Bernoulli}(\theta^{DI}). \tag{2.6}$$

This prior structure can be written hierarchically as:

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

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

and 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.

2.2.2 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.1.1.

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:

$$\begin{aligned}\pi(\mathbf{A}, \Sigma) = \pi(\Sigma) \prod_{(i,j)\in\mathcal{R}} \prod_{p=1}^{P} \left[(1-\gamma_{p,ij}^{DI})\Phi_0(\mathbf{A}_p^{ij} \mid \lambda_0) + \gamma_{p,ij}^{DI}\Phi_1(\mathbf{A}_p^{ij} \mid \lambda_1)\right] \\ \times \prod_{i=1}^{N} \prod_{p=1}^{P} \left[(1-\gamma_{p,ii}^{CSH})\Phi_0(\mathbf{A}_p^{ii} - \bar{\mathbf{A}}_p \mid \eta_0) + \gamma_{p,ii}^{CSH}\Phi_1(\mathbf{A}_p^{ii} - \bar{\mathbf{A}}_p \mid \eta_1)\right],\end{aligned} \tag{2.16}$$

where $\mathcal{R} := \{(i,j) : i,j = 1,\ldots,N,\ i \neq j\}$ denotes the set of off-diagonal DI blocks, and the second product is taken over the main-diagonal CSH blocks $\{\mathbf{A}_p^{ii} : i = 1,\ldots,N\}$. 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.

2.2.3 Choice of the Penalty

Figure A.4.1 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:

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

where $r = N(N-1)$ and $c = N(N-1)/2$ represent the total number of dynamic in-

terdependency (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 $\tau^2_{0,ij}$ and $\kappa^2_{0,ij}$ 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 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 A.3–A.4.

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:

$$\underset{\mathbf{A}}{\operatorname{argmin}} \left\{ \frac{1}{T} \operatorname{tr}\left[(Y_t - Z_t\mathbf{A})'\Omega(Y_t - Z_t\mathbf{A})\right] + \operatorname{pen}(\Lambda, \mathbf{A}, \bar{\mathbf{A}}) \right\}, \tag{2.17}$$

where $\Omega = \Sigma^{-1}$ is the precision matrix and $\operatorname{tr}(\cdot)$ denotes the trace. The coefficient vector $\mathbf{A} = (\operatorname{vec}(A_1)', \operatorname{vec}(A_2)', \ldots, \operatorname{vec}(A_P)')'$ 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 $\operatorname{pen}(\Lambda, \mathbf{A}, \bar{\mathbf{A}})$ 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.

Let $\gamma_{p,ij}^{DI} \sim \text{Bernoulli}(\theta_{ij}^{DI})$, $\theta_{ij}^{DI} \in (0,1)$, and $\pi(\mathbf{A}_p^{ij} \mid \theta_{ij}^{DI})$ be the spike-and-slab Laplace mixture prior given θ_{ij}^{DI}:

$$\pi(\mathbf{A}_p^{ij} \mid \theta_{ij}^{DI}) = \theta_{ij}^{DI}\, \Phi_1(\mathbf{A}_p^{ij}) + (1 - \theta_{ij}^{DI})\, \Phi_0(\mathbf{A}_p^{ij}), \tag{2.18}$$

where $\Phi_0(\cdot)$ and $\Phi_1(\cdot)$ are the spike and slab densities given by

$$\Phi_0(a) = \tfrac{\lambda_0}{2} \exp(-\lambda_0 \|a\|_1),\ \Phi_1(a) = \tfrac{\lambda_1}{2} \exp(-\lambda_1 \|a\|_1),\ 0 < \lambda_1 < \lambda_0. \tag{2.19}$$

Here $\Phi_0(\cdot)$ and $\Phi_1(\cdot)$ are specified as block-level Laplace kernels depending on the scalar block norm $\|\mathbf{A}_p^{ij}\|_1$. Thus the normalizing factor is taken to be $\lambda_k/2$, following the scalar PSSL form, rather than the product-Laplace normalizing constant for the vectorized block.

Definition 1 (PSSL penalty under DI restriction) *The posterior inclusion probability associated with the conditional spike-and-slab Laplace mixture prior is*

$$\mathbb{P}_{\gamma_{p,ij}^{DI}=1}(\mathbf{A}_p^{ij}) := \frac{\theta_{ij}^{DI}\,\Phi_1(\mathbf{A}_p^{ij})}{\theta_{ij}^{DI}\,\Phi_1(\mathbf{A}_p^{ij}) + (1-\theta_{ij}^{DI})\,\Phi_0(\mathbf{A}_p^{ij})}. \tag{2.20}$$

The panel spike-and-slab lasso (PSSL) penalty under the DI restriction is

$$\text{pen}_{DI}(\mathbf{A} \mid \theta^{DI}) = \log\left(\frac{\pi(\mathbf{A} \mid \theta^{DI})}{\pi(\mathbf{0} \mid \theta^{DI})}\right), \tag{2.21}$$

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

The inverse of the posterior inclusion probability in (2.20) is:

$$\big[\mathbb{P}_{\gamma_{p,ij}^{DI}=1}(\mathbf{A}_p^{ij})\big]^{-1} = \frac{\theta_{ij}^{DI}\,\Phi_1(\mathbf{A}_p^{ij}) + (1-\theta_{ij}^{DI})\,\Phi_0(\mathbf{A}_p^{ij})}{\theta_{ij}^{DI}\,\Phi_1(\mathbf{A}_p^{ij})} = 1 + \frac{1-\theta_{ij}^{DI}}{\theta_{ij}^{DI}} \cdot \frac{\Phi_0(\mathbf{A}_p^{ij})}{\Phi_1(\mathbf{A}_p^{ij})}.$$

Using the expressions of $\Phi_0(\mathbf{A}_p^{ij})$ and $\Phi_1(\mathbf{A}_p^{ij})$ from (2.19), it is straightforward to see that

$$\frac{\Phi_0(\mathbf{A}_p^{ij})}{\Phi_1(\mathbf{A}_p^{ij})} = \frac{\lambda_0}{\lambda_1}\,\exp\big(-(\lambda_0-\lambda_1)\,\|\mathbf{A}_p^{ij}\|_1\big).$$

Thus, we obtain

$$\mathbb{P}_{\gamma_{p,ij}^{DI}=1}(\mathbf{A}_p^{ij}) = \left[1 + \frac{\lambda_0}{\lambda_1}\left(\frac{1-\theta_{ij}^{DI}}{\theta_{ij}^{DI}}\right)\exp\big(-(\lambda_0-\lambda_1)\,\|\mathbf{A}_p^{ij}\|_1\big)\right]^{-1}. \tag{2.22}$$

Expression (2.22) reflects two key forms of adaptivity. First, as $\|\mathbf{A}_p^{ij}\|_1$ 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.

Building on Definition 1, the DI prior factorizes over off-diagonal blocks $\mathbf{A}_p^{ij}$, so the log ratio in (2.21) separates additively by block. The next proposition provides the blockwise decomposition, isolating an explicit ℓ_1-shrinkage term from a data-driven adjustment governed by the posterior inclusion probability (2.20).

Proposition 1 (Blockwise decomposition of the DI penalty) *Under Definition 1, for any off-diagonal block* $\mathbf{A}_p^{ij}$*,*

$$\mathrm{pen}_{DI}(\mathbf{A}_p^{ij} \mid \theta_{ij}^{DI}) = -\lambda_1 \|\mathbf{A}_p^{ij}\|_1 + \log\left(\frac{\mathbb{P}_{\gamma_{p,ij}^{DI}=1}(\mathbf{0})}{\mathbb{P}_{\gamma_{p,ij}^{DI}=1}(\mathbf{A}_p^{ij})}\right), \tag{2.23}$$

where $\mathbb{P}_{\gamma_{p,ij}^{DI}=1}(\cdot)$ *is the posterior inclusion probability defined in* (2.20)*. The first term is the baseline* ℓ_1*-shrinkage induced by the slab; the second is an adaptive correction governed by the posterior inclusion probability.*

Proposition 1 expresses the penalty in terms of the posterior inclusion probability $\mathbb{P}_{\gamma_{p,ij}^{DI}=1}(\mathbf{A}_p^{ij})$ introduced in (2.20). The following corollary uses this decomposition to characterize the local shrinkage rate of the penalty as a weighted average of λ_0 and λ_1, with weight equal to the posterior inclusion probability.

Corollary 1 (Adaptive shrinkage) *For any off-diagonal block* $\mathbf{A}_p^{ij}$ *with* $\|\mathbf{A}_p^{ij}\|_1 > 0$*,*

$$\frac{\partial\, \mathrm{pen}_{DI}(\mathbf{A} \mid \theta^{DI})}{\partial\, \|\mathbf{A}_p^{ij}\|_1} = -\,\lambda^*_{\theta_{ij}^{DI}}(\mathbf{A}_p^{ij}), \tag{2.24}$$

where the effective shrinkage rate is the posterior-inclusion-weighted average

$$\lambda^*_{\theta_{ij}^{DI}}(\mathbf{A}_p^{ij}) = \lambda_1\, w(\mathbf{A}_p^{ij}) + \lambda_0\,\big(1 - w(\mathbf{A}_p^{ij})\big), \qquad w(\mathbf{A}_p^{ij}) := \mathbb{P}_{\gamma_{p,ij}^{DI}=1}(\mathbf{A}_p^{ij}). \tag{2.25}$$

Corollary 1 formalizes the adaptive nature of the DI penalty: the effective shrinkage rate $\lambda^*_{\theta_{ij}^{DI}}(\mathbf{A}_p^{ij})$ smoothly transitions from λ_0 (heavy shrinkage, when the posterior inclusion probability is close to zero) to λ_1 (light shrinkage, when it is close to one). The transition weight is determined endogenously by the coefficient magnitude through Proposition 1.[5]

[Figure 1 about here]

Figure 1a shows the shrinkage pattern implied by Equation (2.24). The gradient begins steep near zero, indicating aggressive shrinkage for small coefficients, and flattens as $\|\mathbf{A}_p^{ij}\|_1$ increases, thereby selectively preserving significant dependencies.

[5]Proofs of Proposition 1, and Corollary 1 are provided in Appendix A.2.1, and Appendix A.2.2, respectively.

2.3.2 CSH Penalty and Posterior Inclusion Probability

The cross-sectional homogeneity (CSH) penalty 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 as the DI penalty, we formalize this as follows.

Let $\gamma_{p,ii}^{CSH} \sim \text{Bernoulli}(\theta_{ii}^{CSH})$, $\theta_{ii}^{CSH} \in (0,1)$, and $\pi(\Delta\mathbf{A}_p^{ii} \mid \theta_{ii}^{CSH})$ be the conditional spike-and-slab Laplace mixture prior given θ_{ii}^{CSH}:

$$\pi(\Delta\mathbf{A}_p^{ii} \mid \theta_{ii}^{CSH}) = \theta_{ii}^{CSH}\, \Phi_1(\Delta\mathbf{A}_p^{ii}) + (1 - \theta_{ii}^{CSH})\, \Phi_0(\Delta\mathbf{A}_p^{ii}), \tag{2.26}$$

where the spike and slab densities are

$$\Phi_0(\Delta\mathbf{A}_p^{ii}) = \tfrac{\eta_0}{2} \exp(-\eta_0 \|\Delta\mathbf{A}_p^{ii}\|_1),\ \Phi_1(\Delta\mathbf{A}_p^{ii}) = \tfrac{\eta_1}{2} \exp(-\eta_1 \|\Delta\mathbf{A}_p^{ii}\|_1),\ 0 < \eta_1 < \eta_0. \tag{2.27}$$

Here $\Phi_0(\cdot)$ and $\Phi_1(\cdot)$ are specified as block-level Laplace kernels depending on the scalar deviation norm $\|\Delta\mathbf{A}_p^{ii}\|_1$. Again, the normalizing factor is $\eta_k/2$, following the scalar PSSL form, rather than the product-Laplace normalizing constant for the vectorized block.

Definition 2 (PSSL penalty under CSH restriction) *The posterior inclusion probability associated with the conditional spike-and-slab Laplace mixture prior is*

$$\mathbb{P}_{\gamma_{p,ii}^{CSH}=1}(\Delta\mathbf{A}_p^{ii}) := \frac{\theta_{ii}^{CSH}\, \Phi_1(\Delta\mathbf{A}_p^{ii})}{\theta_{ii}^{CSH}\, \Phi_1(\Delta\mathbf{A}_p^{ii}) + (1 - \theta_{ii}^{CSH})\, \Phi_0(\Delta\mathbf{A}_p^{ii})}. \tag{2.28}$$

The panel spike-and-slab lasso penalty under the CSH restriction is

$$\text{pen}_{CSH}(\Delta\mathbf{A} \mid \theta^{CSH}) = \log\left(\frac{\pi(\Delta\mathbf{A} \mid \theta^{CSH})}{\pi(\mathbf{0} \mid \theta^{CSH})}\right), \tag{2.29}$$

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

The inverse of the posterior inclusion probability in (2.28) is:

$$\left[\mathbb{P}_{\gamma_{p,ii}^{CSH}=1}(\Delta\mathbf{A}_p^{ii})\right]^{-1} = \frac{\theta_{ii}^{CSH}\, \Phi_1(\Delta\mathbf{A}_p^{ii}) + (1 - \theta_{ii}^{CSH})\, \Phi_0(\Delta\mathbf{A}_p^{ii})}{\theta_{ii}^{CSH}\, \Phi_1(\Delta\mathbf{A}_p^{ii})} = 1 + \frac{1 - \theta_{ii}^{CSH}}{\theta_{ii}^{CSH}} \cdot \frac{\Phi_0(\Delta\mathbf{A}_p^{ii})}{\Phi_1(\Delta\mathbf{A}_p^{ii})}.$$

Substituting the Laplace densities expression from (2.27), the ratio simplifies to

$$\frac{\Phi_0(\Delta\mathbf{A}_p^{ii})}{\Phi_1(\Delta\mathbf{A}_p^{ii})} = \frac{\eta_0}{\eta_1}\,\exp\big(-(\eta_0-\eta_1)\,\|\Delta\mathbf{A}_p^{ii}\|_1\big).$$

Thus, we obtain

$$\mathbb{P}_{\gamma_{p,ii}^{CSH}=1}(\Delta\mathbf{A}_p^{ii}) = \left[1+\frac{\eta_0}{\eta_1}\left(\frac{1-\theta_{ii}^{CSH}}{\theta_{ii}^{CSH}}\right)\exp\big(-(\eta_0-\eta_1)\,\|\Delta\mathbf{A}_p^{ii}\|_1\big)\right]^{-1}. \tag{2.30}$$

This probability behaves similarly to its DI counterpart. For units with strong deviations from the group trend (large $\|\Delta\mathbf{A}_p^{ii}\|_1$), the inclusion probability is high, reducing shrinkage. Conversely, smaller deviations are aggressively shrunk to zero.

By Definition 2, the CSH prior factorizes over diagonal-block deviations $\{\Delta\mathbf{A}_p^{ii}\}$, so the log-ratio in (2.29) separates additively by block. The next proposition provides the blockwise decomposition, isolating an explicit ℓ_1-shrinkage term from a data-driven adjustment governed by the posterior inclusion probability (2.28).

Proposition 2 (Blockwise decomposition of the CSH penalty) *Under Definition 2, for any diagonal-block deviation* $\Delta\mathbf{A}_p^{ii}$,

$$\mathrm{pen}_{CSH}(\Delta\mathbf{A}_p^{ii} \mid \theta_{ii}^{CSH}) = -\eta_1\,\|\Delta\mathbf{A}_p^{ii}\|_1 + \log\left(\frac{\mathbb{P}_{\gamma_{p,ii}^{CSH}=1}(\mathbf{0})}{\mathbb{P}_{\gamma_{p,ii}^{CSH}=1}(\Delta\mathbf{A}_p^{ii})}\right), \tag{2.31}$$

where $\mathbb{P}_{\gamma_{p,ii}^{CSH}=1}(\cdot)$ *is the posterior inclusion probability defined in* (2.28)*. The first term is the baseline* ℓ_1*-shrinkage induced by the slab; the second is an adaptive correction governed by the posterior inclusion probability.*

Because the CSH prior is conditionally independent across blocks, the total CSH penalty decomposes as

$$\mathrm{pen}_{CSH}(\Delta\mathbf{A} \mid \theta^{CSH}) = -\eta_1\,\|\Delta\mathbf{A}\|_1 + \sum_{i=1}^{R}\sum_{p=1}^{P}\log\left(\frac{\mathbb{P}_{\gamma_{p,ii}^{CSH}=1}(\mathbf{0})}{\mathbb{P}_{\gamma_{p,ii}^{CSH}=1}(\Delta\mathbf{A}_p^{ii})}\right), \tag{2.32}$$

where $\|\Delta\mathbf{A}\|_1 = \sum_{i=1}^{R}\sum_{p=1}^{P}\|\Delta\mathbf{A}_p^{ii}\|_1$. Equation (2.31) expresses the penalty in terms of the posterior inclusion probability $\mathbb{P}_{\gamma_{p,ii}^{CSH}=1}(\Delta\mathbf{A}_p^{ii})$ introduced in (2.28). The following corollary uses this decomposition to characterize the local shrinkage rate of the penalty as

a weighted average of η_0 and η_1 with weight equal to the posterior inclusion probability.[6]

Corollary 2 (Adaptive shrinkage in the CSH penalty) *For any diagonal-block deviation $\Delta\mathbf{A}_p^{ii}$ with $\|\Delta\mathbf{A}_p^{ii}\|_1 > 0$,*

$$\frac{\partial\, \mathrm{pen}_{CSH}(\Delta\mathbf{A} \mid \theta^{CSH})}{\partial\, \|\Delta\mathbf{A}_p^{ii}\|_1} = -\,\eta^*_{\theta_{ii}^{CSH}}(\Delta\mathbf{A}_p^{ii}), \tag{2.33}$$

where the effective shrinkage rate is the posterior-inclusion-weighted average

$$\eta^*_{\theta_{ii}^{CSH}}(\Delta\mathbf{A}_p^{ii}) = \eta_1\, w(\Delta\mathbf{A}_p^{ii}) + \eta_0\,\big(1 - w(\Delta\mathbf{A}_p^{ii})\big), \qquad w(\Delta\mathbf{A}_p^{ii}) := \mathbb{P}_{\gamma_{p,ii}^{CSH}=1}(\Delta\mathbf{A}_p^{ii}). \tag{2.34}$$

As shown in Figure 1b, 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 experiments designed to evaluate the ability of the proposed PSSL method to recover structural features of panel VAR systems. The data-generating processes are constructed to isolate the roles of dynamic interdependencies (DI), cross-sectional heterogeneity (CSH), and their interaction. This design enables a direct assessment of how well the estimator identifies the relevant propagation structure and how alternative restrictions affect estimation accuracy and forecasting performance.

Each DGP is evaluated under low, moderate, and high-dimensional settings, defined by progressively increasing the sparsity of the coefficient matrix $\mathbf{A}$. We vary the cross-sectional dimension with $N = 3$, 5, and 10, while holding the time dimension fixed at $T = 100$. To capture different levels of within-unit richness, we consider $G = 2$ and $G = 4$, corresponding to limited and more informative variable sets, respectively. The VAR specification includes a single lag ($P = 1$). Simulation results are reported for six configurations, covering different combinations of N and G, thereby allowing us to examine how the performance of the model scales with dimensionality.

[6]Proofs of Proposition 2, and Corollary 2 are provided in Appendix A.2.3, and Appendix A.2.4, respectively.

3.1 Experiment Design

Figure 2 illustrates the specific autoregressive coefficient patterns used in each DGP, offering a visual summary of their structural designs.

[Figure 2 about here]

DGP1 (Both DI and CSH Restrictions): DGP1 is constructed to assess the ability of the estimator to recover both dynamic interdependencies (DI) and cross-sectional heterogeneity (CSH). The cross-sectional units follow a hierarchical structure in which unit 1 influences all other units but is not affected by them, unit 2 transmits shocks only to units lower in the hierarchy, and so on, while units from position 4 onward are isolated and do not affect others. This design is reflected in the coefficient matrix through nonzero lower-triangular off-diagonal blocks for units 1–3, with all remaining off-diagonal blocks set to zero. At the same time, the diagonal blocks contain a mixture of common and unit-specific dynamics, generating cross-sectional heterogeneity that must be inferred from the data. All nonzero coefficients are drawn independently from the uniform distribution on $[-0.9, 0.9]$, allowing for variation in both the magnitude and direction of interactions across units.

DGP2 (DI Restriction Only): DGP2 extends the baseline design of DGP1 by removing cross-sectional homogeneity. Specifically, all diagonal coefficient blocks are set to distinct values, i.e., $A_{1,1} \neq A_{2,2} \neq \cdots \neq A_{N,N}$, so that only dynamic interdependencies (DI) need to be identified. In this setting, the true value of the CSH shrinkage indicator is $\gamma^{\text{CSH}} = 1$, implying the absence of homogeneity restrictions across units. A well-performing estimator should therefore recover posterior estimates $\hat{\gamma}^{\text{CSH}}$ close to one. The off-diagonal structure is kept unchanged from DGP1, thereby preserving the hierarchical pattern of shock transmission across units.

DGP3 (CSH Restriction Only): DGP3 is characterised by a dense coefficient matrix with weak but pervasive dynamic interdependencies, such that most off-diagonal blocks are nonzero but small in magnitude. This design reduces the relevance of DI restriction search, as spillovers are present across units by construction and should therefore not be strongly shrunk. In contrast, the diagonal blocks replicate the structure introduced in DGP1, with some units sharing common own-dynamics while others exhibit unit-specific persistence. This configuration motivates a search for cross-sectional heterogeneity (CSH)

restrictions. Consistent with this design, the true value of the DI shrinkage indicator is $\gamma^{\mathrm{DI}} = 1$, implying that accurate estimation should yield posterior estimates $\hat{\gamma}^{\mathrm{DI}}$ close to one and hence only limited regularization of the off-diagonal coefficients.

To benchmark the performance of the PSSL approach, we compare it with several well-established alternatives. These include the Stochastic Search Variable Selection (S4) method of Koop and Korobilis (2016), the Bayesian Lasso (BL), and the Bayesian Fused Lasso (BFL). Together, these approaches represent two prominent shrinkage paradigms in Bayesian sparse modelling: mixture-normal priors (S4) and Laplace-based priors (BL and BFL).[7]

We further compare the proposed estimator with the frequentist penalised panel VAR of Camehl (2023), as well as with factor-based benchmarks that model cross-sectional dependence through cross-sectional shrinkage (Canova and Ciccarelli, 2009). In addition, we include the Global VAR (GVAR) model of Dees et al. (2007), which captures international spillovers using weighted foreign variables constructed from cross-country linkages.

As discussed by Koop and Korobilis (2016), the S4 approach employs a mixture of normal distributions with a calibrated shrinkage parameter, typically inducing relatively mild and smooth regularisation. This often leads to so-called *soft sparsity*, whereby both noise and weak signals are retained due to the light tails of the normal distribution. By contrast, the Laplace priors underlying BL and BFL impose stronger shrinkage around zero, thereby promoting more aggressive sparsity while still allowing large coefficients to escape penalisation through their heavier tails. The BFL further extends this framework by incorporating a fused penalty that encourages similarity across neighbouring coefficients, promoting local smoothness or block-wise structure.[8]

[7] While restrictions on static interdependencies (SI) are important for structural analysis and have been emphasised by Canova and Ciccarelli (2009) and Koop and Korobilis (2016), we do not impose them here. Our focus is on forecasting rather than structural inference, consistent with the forecasting oriented framework of Korobilis (2016).

[8] The fused component adds a penalty on successive coefficient differences, $\sum |\beta_j - \beta_{j-1}|$, in addition to the standard Lasso penalty on coefficient magnitudes. In contrast, the Bayesian Lasso penalises only the absolute values of the coefficients, encouraging sparsity but not smoothness across adjacent effects.

Table 1: Estimation Accuracy - Relative RMSE ($G = 4$)

	Spike & Slab		**Regularization**					**Frequentist**
	GVS	**GAS**	**BL**	**BFL**	**CC**	**GVAR**	**S4**	**Lasso** $(\lambda, c, \alpha, \Sigma)$
A. DGP 1								
$N = 5$	1.00	1.01	8.54	2.86	95.46	104.25	4.92	15.89
$N = 10$	1.00	1.09	6.62	3.30	60.5	94.84	3.53	10.60
B. DGP 2								
$N = 5$	1.00	1.02	8.20	2.82	90.19	111.53	4.72	15.32
$N = 10$	1.00	1.08	6.67	3.32	61.39	71.35	3.52	10.67
C. DGP 3								
$N = 5$	1.00	1.07	1.63	0.97	31.3	36.62	0.99	10.25
$N = 10$	1.00	1.03	1.87	0.99	21.98	25.6	0.99	8.98

Note: Average estimation errors across three DGPs (G = 4) and two sample sizes (N = 5, 10), normalized to GVS. Values below 1.00 indicate better performance. GVS and GAS are Bayesian selection methods; BL and BFL are Bayesian shrinkage methods; CC denotes factor-based models; GVAR is a global VAR with weighted averages; S4 is a stochastic selection procedure; Lasso uses frequentist penalization.

3.2 Estimation Accuracy

We evaluate estimation accuracy using the *root mean squared error* (RMSE) of the autoregressive coefficient matrix $\mathbf{A}$. In our framework, RMSE is defined as

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

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

Table 1 reports estimation accuracy for the four-variable specification ($G = 4$) across the three data-generating processes (DGPs 1–3) and two panel dimensions ($N = 5$ and $N = 10$). The results show that the proposed group-value shrinkage estimator (GVS) consistently delivers the highest estimation accuracy across most configurations. By exploiting block-specific shrinkage, GVS effectively captures heterogeneous dynamic linkages across panel units and outperforms alternative methods — including Bayesian Lasso (BL), Bayesian fused Lasso (BFL), stochastic search selection (S4), and frequentist Lasso — in both sparse (DGPs 1–2) and relatively dense (DGP 3) environments.

The performance gains are particularly pronounced in sparse settings characterised by strong dynamic interdependency restrictions. In DGPs 1 and 2, adaptive shrinkage plays a crucial role in distinguishing economically relevant spillovers from negligible interactions. Relative to non-adaptive shrinkage approaches such as BL and BFL, the proposed estimator more effectively eliminates irrelevant coefficients, leading to substantially lower RMSE values. In contrast, under DGP 3 — where the restriction structure is driven primarily by cross-sectional homogeneity rather than sparsity — the relative advantage of adaptive shrinkage is reduced. In this case, the estimation accuracy of S4 and BFL becomes comparable to that of the PSSL-based estimators, reflecting the limited scope for sparsity-driven gains when most dynamic interdependencies are weak but pervasive.

Additional results reported in Appendix Table A.5.1 indicate that when the group structure is simpler ($G = 2$), the group-average shrinkage estimator (GAS) performs particularly well, especially in sparse designs with larger panel dimensions. This finding highlights the complementary nature of the two proposed estimators: GAS is better suited to environments with limited group heterogeneity, while GVS becomes increasingly advantageous as the richness of cross-sectional structure rises. Importantly, both adaptive estimators exhibit clear improvements over non-adaptive alternatives as dimensionality increases.

By contrast, the factor-based CC model exhibits substantially weaker estimation accuracy. Because it summarises cross-sectional dependence through a small number of latent common components rather than explicitly modelling sparse bilateral interdependencies, it is less well aligned with the heterogeneous propagation structures embedded in the DGPs. Similarly, the large RMSE values observed for the GVAR benchmark reflect a structural mismatch between the model specification and the underlying data-generating processes. In particular, the GVAR framework captures spillovers using pre-specified weighted averages of foreign variables. When the true transmission mechanism involves sparse and heterogeneous bilateral linkages, this aggregation leads to specification error that accumulates across equations, resulting in markedly poorer estimation performance.

3.3 Forecasting Accuracy

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 $T_{\text{test}} = 12$, implying $T_{\text{train}} = T - 12$ 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:

$$\text{MSFE}_l^i(h) = \frac{1}{T_{\text{test}} - h + 1} \sum_{t=T_{\text{train}}}^{T-h} \left(\hat{Y}_{l,t+h}^i - Y_{l,t+h}^i \right)^2, \tag{3.2}$$

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 observed 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 $h_{\max} = 12$.

Table 2 reports relative MSFEs for the four variable specification ($G = 4$) across the three DGPs and two panel dimensions ($N = 5$ and $N = 10$), normalised to the GVS benchmark. Values below one indicate superior performance relative to GVS and vice versa. For the two-group specification ($G = 2$), the corresponding results are reported in Appendix Table A.5.2 and are normalised to GAS, because GAS provides the more relevant adaptive benchmark when the group structure is less rich.[9]

Two main findings emerge from the forecasting exercise. First, the advantages of adaptive shrinkage become increasingly pronounced as the dimensionality of the panel rises. In smaller systems ($N = 5$), differences across estimators are relatively modest. In this setting, non-adaptive shrinkage approaches such as BL and BFL, as well as stochastic search variable selection (S4), remain competitive and, in the case of DGP 3, perform better. However, when the dimension increases to $N = 10$, the proposed GVS estimator delivers systematically lower MSFEs in the sparse designs (DGPs 1–2), while performing comparably to competing methods in the denser environment (DGP 3).

This pattern closely mirrors the estimation results and highlights the key advantage of adaptive shrinkage: its ability to detect and retain the relevant bilateral transmission channels while aggressively shrinking irrelevant coefficients toward zero when dynamic interdependencies are sparse. By contrast, when the primary restrictions arise from cross-sectional homogeneity rather than sparsity, as in DGP 3, the benefits of shrinkage toward

[9]The GVAR benchmark is excluded from the forecasting comparison because its structural misspecification in the simulated environments—already evident in the estimation accuracy results—leads to poor predictive performance. In particular, the use of pre-specified weighted averages to summarise cross-sectional dependence fails to capture the sparse and heterogeneous bilateral spillovers embedded in the DGPs, resulting in rapidly increasing forecast errors, particularly at longer horizons.

Table 2: Average Relative RMSFE Across Forecast Horizons for DGP1–DGP3, $N \in \{5, 10\}$, and $G = 4$

	PSSL		Other Bayesian Methods			Factor	Lasso
Horizon	GVS	GAS	BL	BFL	S4	CC	$(\lambda, c, \alpha, \Sigma)$
			Panel A: $N = 5$				
DGP1							
$h = 1$	1.00	1.08	1.00	0.99	1.03	1.75	1.04
$h = 3$	1.00	1.12	0.99	0.99	1.01	1.31	1.20
$h = 6$	1.00	1.14	1.01	1.01	1.02	1.03	1.00
$h = 12$	1.00	1.20	1.00	0.99	1.00	1.04	0.93
DGP2							
$h = 1$	1.00	1.08	0.99	0.98	1.01	1.96	1.04
$h = 3$	1.00	1.11	0.97	0.97	0.99	1.44	1.20
$h = 6$	1.00	1.13	0.99	0.99	1.00	1.09	1.00
$h = 12$	1.00	1.15	0.99	0.99	1.00	0.88	0.93
DGP3							
$h = 1$	1.00	1.17	0.91	0.91	0.92	4.28	2.29
$h = 3$	1.00	1.23	0.87	0.87	0.87	2.25	1.53
$h = 6$	1.00	1.36	0.81	0.81	0.81	1.68	1.37
$h = 12$	1.00	4.55	0.69	0.69	0.69	1.17	12.81
			Panel B: $N = 10$				
DGP1							
$h = 1$	1.00	1.01	1.11	1.11	1.09	2.80	0.95
$h = 3$	1.00	1.01	1.08	1.08	1.08	2.04	1.14
$h = 6$	1.00	1.01	1.06	1.06	1.06	1.25	1.00
$h = 12$	1.00	1.01	1.06	1.07	1.05	0.83	0.79
DGP2							
$h = 1$	1.00	1.01	1.13	1.13	1.10	3.11	0.97
$h = 3$	1.00	1.01	1.10	1.10	1.09	1.94	1.24
$h = 6$	1.00	1.02	1.07	1.07	1.07	1.17	1.12
$h = 12$	1.00	1.01	1.03	1.03	1.03	0.78	0.84
DGP3							
$h = 1$	1.00	1.11	0.99	0.99	1.04	4.25	2.51
$h = 3$	1.00	1.24	0.99	0.99	1.04	2.50	1.80
$h = 6$	1.00	1.42	0.99	0.99	1.03	1.95	1.60
$h = 12$	1.00	1.82	1.00	1.00	1.02	1.58	1.43

Notes: The table reports average relative root mean squared forecast errors (RMSFE) across countries and variables. Each entry is computed as $\text{RMSFE}_m/\text{RMSFE}_{\text{GVS}}$, so values below one indicate an improvement relative to the GVS benchmark. GAS denotes an alternative PSSL specification; BL, BFL, and S4 denote alternative Bayesian shrinkage methods; CC is a factor-based model; GVAR is a global VAR with weighted foreign variables; and Lasso is a penalized regression model with tuning parameters $(\lambda, c, \alpha, \Sigma)$.

zero are naturally more limited. As a result, predictive performance across estimators becomes more similar in such settings.

Examining the behaviour of individual estimators helps clarify the sources of these differences in predictive performance. Relative to BL, BFL, and frequentist Lasso, the proposed estimators benefit from shrinkage that adapts jointly to dynamic interdependency (DI) and cross-sectional homogeneity (CSH) structures. In particular, Lasso performs relatively poorly at short and medium horizons, as its global penalisation tends to overshrink coefficients and attenuate economically meaningful transmission mechanisms. By contrast, the Laplace-based spike-and-slab structure underlying the PSSL framework allows weak interactions to be strongly penalised while preserving persistent signals, thereby improving short-horizon forecast accuracy. The factor-based CC model exhibits poor performance at shorter horizons but becomes more competitive at longer horizons. This reflects the fact that common-component structures are less effective when short-run forecasts depend on heterogeneous bilateral linkages, yet can capture persistent co-movements that dominate at longer horizons.

Second, the richness of the group structure determines which adaptive estimator performs best. When group heterogeneity is substantial ($G = 4$), GVS delivers superior forecasting performance, consistent with its ability to accommodate complex cross-group dynamics. By contrast, when group information is more limited ($G = 2$), GAS performs better in the sparse DGPs. These findings highlight the complementary roles of the two estimators: each is well suited to a different degree of group complexity, and both provide clear improvements over global shrinkage methods as panel dimensionality increases.[10]

3.4 Data-based Simulation Results (DGP4)

The first three simulations (DGP1–DGP3) show that the proposed PSSL approach improves both estimation and forecasting accuracy across a variety of sparse and dense settings. We now turn to DGP4, a pseudo-realistic experiment designed to reflect the structural features of real-world macroeconomic data. This simulation draws on the dataset of Camehl (2023) and focuses on 12 countries: ten euro area economies commonly used in panel VAR research[11]—five core (Austria, Belgium, Finland, France, and the Netherlands) and five periphery (Spain, Greece, Ireland, Italy, and Portugal)—together with two externally influential economies, the United Kingdom and the United States. For each

[10]The $G = 4$ results are normalised to GVS, while the $G = 2$ results are normalised to GAS, as each estimator serves as the natural adaptive benchmark in the setting where it performs best.

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

Table 3: Aggregated Forecast Performance Across Countries and Variables

Horizon	**PSSL**		**Other Bayesian Methods**					**Lasso**
	GVS	**GAS**	**BL**	**BFL**	**S4**	**CC**	**GVAR**	$\lambda, c, \alpha, \Sigma$
$h = 1$	1.00	1.08	1.21	1.20	1.25	1.05	5.40	1.16
$h = 3$	1.00	1.01	1.13	1.13	1.21	1.25	2.74	1.37
$h = 6$	1.00	1.00	0.99	1.00	1.05	1.08	2.45	1.12
$h = 12$	1.00	1.00	1.03	1.03	1.04	1.05	5.03	1.05

Note: This table reports the average relative Mean Squared Forecast Errors (MSFE) across models, forecast horizons, countries, and variables. Each entry represents $\text{MSFE}_{\text{model}}/\text{MSFE}_{\text{benchmark}}$, where the benchmark is the GVS model. Models include GAS (an alternative PSSL specification), Bayesian shrinkage priors (BL, BFL, S4), factor-based models (CC), global VAR with weighted averages (GVAR), and a frequentist penalized regression approach (Lasso with hyperparameters $(\lambda, c, \alpha, \Sigma)$). Averages are taken across all countries and forecast variables.

country, we include four macroeconomic indicators: the Consumer Price Index (CPI), Industrial Production Growth (IP), the Real Effective Exchange Rate (REER), and the Unemployment Rate (UN).

To generate the pseudo-real data-generating process (DGP4), we first 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 illustrates the data construction and simulation process.

[Figure 3 about here]

Table 3 reports aggregate forecast performance, with relative MSFE averaged across countries and variables. These averages summarize the overall predictive accuracy of each method across countries. Figure 4 complements this by presenting country-level MSFE results [12], while the detailed numerical values are provided in Appendix Tables A.5.3–A.5.6. The country-level results allow us to assess whether the aggregate findings are consistent across individual countries and to evaluate the extent of cross-country heterogeneity in forecast performance. In all cases, results are normalized to the GVS benchmark.

[12]GVAR is excluded from the figure because its large forecast errors in several countries would distort the scale; the corresponding results are nevertheless reported in the appendix.

At the aggregate level, the proposed GVS estimator delivers consistently strong forecasting performance, particularly at short and medium horizons ($h = 1$ and $h = 3$). Its relative advantage over competing approaches reflects the adaptive nature of its shrinkage mechanism. Compared with non-adaptive shrinkage specifications, GVS produces lower forecast errors because fixed penalty structures tend to overshrink economically meaningful coefficients toward zero. This difference becomes less pronounced at longer horizons, where persistent common dynamics play a greater role in determining predictive accuracy. The results also indicate that GVS systematically outperforms the stochastic search selection procedure (S4), highlighting the benefits of Laplace-based shrinkage that adjusts endogenously to the information content of the data rather than relying on fixed mixture structures.

[Figure 4 about here]

Country-level results broadly reinforce the aggregate findings while highlighting substantial cross-country dispersion in the forecasting performance of alternative estimators. At the one-step horizon ($h = 1$), the boxplots reveal considerable variability in relative MSFEs across countries for most competing methods. In many cases the distributions are wide and include several large deviations from the benchmark, indicating that these estimators perform well in some countries but poorly in others.

As the forecast horizon increases to $h = 3$, this pattern largely persists. Forecast errors for competing approaches remain dispersed across countries, suggesting that their performance varies substantially across the panel.

At longer horizons ($h = 6$ and $h = 12$), the dispersion in relative MSFEs across methods becomes smaller, although noticeable differences across estimators remain. This pattern suggests that persistent common dynamics and global cyclical forces play a larger role in shaping longer-run forecasts, reducing the influence of modelling detailed cross-country transmission mechanisms.

Overall, the country-level evidence indicates that while alternative shrinkage methods can perform competitively in individual cases, their forecasting accuracy varies considerably across countries and horizons.

4 Eurozone Financial Contagion

We examine the role of group-specific variable selection in the *PSSL* framework by revisiting and extending the dataset of Koop and Korobilis (2016) to analyse financial contagion in euro area sovereign bond markets. The relevance of cross country linkages is assessed using posterior inclusion probabilities (PIPs).

The euro area sovereign debt crisis and its aftermath provide a natural setting for studying spillover effects within the monetary union, given the heightened volatility in sovereign spreads since 2008 and the asymmetric responses of member countries during episodes of financial stress. Building on the empirical frameworks of Koop and Korobilis (2016) and Korobilis (2016), we investigate the structure of cross-country interdependencies and the dynamics of financial contagion. Our analysis uses an extended monthly dataset covering the period January 2001 to December 2019, thereby encompassing the pre-crisis phase, the sovereign debt crisis, and the subsequent post-crisis recovery.

The empirical specification focuses on three key variables: (i) 10-year government bond yields, (ii) total industrial production, and (iii) average bid–ask spreads, which serve as a proxy for market liquidity. Consistent with the existing literature, all variables are expressed as spreads relative to Germany, commonly treated as the benchmark risk-free economy within the euro area. This normalisation facilitates cross-country comparisons of sovereign credit risk, real activity, and market liquidity conditions. The sample is partitioned 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 measured in year-over-year log changes.

4.1 Model Selection and Structural Restrictions

To identify the specification most appropriate for forecasting and structural interpretation, we compare alternative PVAR models using marginal likelihood criteria. Specifically, we consider five specifications. The first is the PSSL model imposing both dynamic interdependencies (DI) and cross-sectional heterogeneity (CSH) restrictions, implemented under both the GVS and GAS shrinkage schemes. The second allows for DI restrictions only, while permitting full cross-sectional heterogeneity. The third im-

poses CSH restrictions while allowing unrestricted dynamic interactions. The fourth is a fully unrestricted PVAR without shrinkage. Finally, we include the S4 model of Koop and Korobilis (2016), which incorporates both DI and CSH through fixed, pre-calibrated normal-mixture shrinkage.

Table 4: 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 reports marginal log-likelihood values computed using the Gelfand–Dey method. The specification imposing only DI restrictions achieves the highest marginal likelihood, indicating that dynamic spillovers play a central role in explaining variation in euro area sovereign bond markets. By contrast, models that impose cross-sectional homogeneity (i.e., exclude CSH) perform poorly—indeed worse than the fully unrestricted specification—highlighting the limitations of strong homogeneity assumptions across countries.

The GAS specification, which introduces both DI and CSH through group-average shrinkage, yields only limited improvements relative to the CSH-only model. This suggests that shrinkage toward group means may mask economically relevant within-group variation. While some degree of homogeneity may be present, the evidence points to heterogeneity arising primarily at the variable level rather than uniformly across group averages. Consequently, the relatively diffuse regularisation implied by GAS appears insufficient to capture the structural differences embedded in the data.

In contrast, GVS and S4 shrink coefficients toward group-specific targets rather than common group means, allowing these approaches to preserve meaningful within-group heterogeneity while still benefiting from regularisation. Among these alternatives, GVS delivers more precise and targeted shrinkage, consistently outperforming GAS when both DI and CSH restrictions are imposed. The S4 model also improves upon the unrestricted

benchmark, placing its performance between that of GVS and GAS. Overall, these results underscore the importance of regularisation strategies that balance parsimony with the preservation of variable-level heterogeneity, rather than relying on overly smooth group-level shrinkage that may obscure key structural features.

[Figure 5 about here]

4.2 Dynamic Interdependencies Across the Euro Area

Posterior inclusion probabilities for dynamic interdependencies (DIs) are summarized in Tables A.7.1 and A.7.2, and visualized in Figure 5[13]. 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

[13]For the interpretation of the strength of evidence (from weak to strong) based on posterior probabilities, see Table A.6.1 in Appendix A.6.

Table 5: Average Relative Root of MSFE Across Countries and Variables

Horizon	**PSSL**		**Other Bayesian Methods**					**Lasso**
	GVS	**GAS**	**BL**	**BFL**	**S4**	**CC**	**GVAR**	$\lambda, c, \alpha, \Sigma$
$h = 1$	1	1.02	1.11	1.11	1.24	0.99	2.19	0.94
$h = 3$	1	1.00	1.04	1.04	1.11	1.06	2.99	1.02
$h = 6$	1	1.00	0.98	0.98	0.98	1.04	2.49	1.01
$h = 12$	1	1.00	0.99	0.99	0.98	1.01	2.50	1.01

Note: This table reports the average relative Mean Squared Forecast Errors (MSFE) across models, forecast horizons, countries, and variables. Each entry represents $\text{MSFE}_{\text{model}}/\text{MSFE}_{\text{benchmark}}$, where the benchmark is the GVS model. Models include GAS (an alternative PSSL specification), Bayesian shrinkage priors (BL, BFL, S4), factor-based models (CC), global VAR with weighted averages (GVAR), and a frequentist penalized regression approach (Lasso with hyperparameters $(\lambda, c, \alpha, \Sigma)$). Averages are taken across all countries and forecast variables.

transmission, consistent with a more asymmetric and hierarchical network of interdependencies.

5 Forecasting international spillovers

We next evaluate the forecasting performance of the proposed *PSSL* framework using the international dataset of Camehl (2023), which also underlies the data-based simulation design. This application examines the extent to which adaptive shrinkage improves the modelling of cross-country spillovers and enhances predictive accuracy relative to standard benchmarks.

[Figure 6 about here]

The panel comprises 12 economies—the 10 euro area countries considered in the contagion analysis, together with the United Kingdom and the United States—thereby capturing both intra-euro area dynamics and external macroeconomic influences. Following Camehl (2023), we employ monthly data from January 2000 to December 2019 on four variables: seasonally adjusted log-differences of consumer prices (CPI) and industrial production (IP), the log-differenced and demeaned real effective exchange rate (REER), and the first difference of the unemployment rate.[14] Models are estimated over

[14]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 obtained from the IMF.

the period January 2000 to December 2014, while the subsequent 60 months (January 2015 to December 2019) are reserved for forecast evaluation.

Table 5 reports average relative mean squared forecast errors (MSFE) at horizons $h = 1, 3, 6, 12$, normalised to the Group Variable Selection (GVS) specification. By construction, the benchmark takes the value one, with larger values indicating weaker predictive performance. Across forecast horizons, the GVS specification delivers consistently strong results. The alternative PSSL specification (GAS) remains competitive and occasionally matches or slightly improves upon GVS at longer horizons, although these gains are not systematic. Other Bayesian shrinkage approaches — including BL, BFL, and S4 — generate reasonable forecasts but generally underperform relative to GVS. Performance differences are most pronounced at short horizons ($h = 1$ and $h = 3$), whereas predictive accuracy across methods converges at medium and longer horizons. This pattern suggests that adaptive shrinkage is particularly valuable for short-to-medium-run forecasting, where the accurate identification of dynamic bilateral spillovers plays a central role.

While aggregate MSFEs provide a convenient summary of overall predictive performance, they can conceal substantial cross-country heterogeneity. To evaluate the stability of forecasting accuracy across individual economies, Figure 6 reports the country-level distribution of relative MSFEs, with detailed numerical results presented in Appendix Tables A.7.3–A.7.6.[15]

The distributional evidence broadly confirms the aggregate findings. Both GVS and GAS exhibit relatively tight MSFE distributions centred around unity across forecast horizons, indicating stable and robust predictive performance across countries. In contrast, BL, BFL, and S4 display wider dispersion, suggesting greater sensitivity to country-specific dynamics even when their average forecasting accuracy is comparable to that of the benchmark. This pattern reflects the limitations of global or weakly adaptive shrinkage schemes, which may either excessively attenuate economically relevant spillovers or fail to eliminate negligible interactions. As the forecast horizon increases, dispersion across methods gradually narrows, consistent with the growing importance of persistent common dynamics relative to short-to-medium-run bilateral transmission mechanisms.

[15]The GVAR specification is excluded from the boxplots because its large forecast errors in several countries would distort the scale of the figures; the corresponding results are nevertheless reported in the appendix.

A distinct horizon-dependent performance profile also emerges for the frequentist Lasso. The Lasso performs relatively well at the one-step horizon ($h = 1$), where uniform penalisation can help stabilise short-run forecasts by mitigating estimation noise. Its relative predictive accuracy declines at longer horizons ($h = 3$, $h = 6$, and $h = 12$), however, as global shrinkage attenuates economically meaningful transmission mechanisms and weakens the persistence of dynamic spillovers. Finally, for CC, the low-dimensional common-component structure can capture broad co-movements, but is less effective when forecasting depends on heterogeneous country-to-country spillovers.

Overall, the empirical evidence underscores the benefits of adaptive, coefficient-level shrinkage for modelling international spillovers. By flexibly balancing sparse bilateral linkages with broader common dynamics, the PSSL specifications deliver stable and reliable forecasts across countries and horizons.

6 Conclusion

This paper develops a Bayesian framework for estimating high-dimensional panel VAR models that combines flexibility, interpretability, and strong empirical performance. The main contribution is the Panel Spike-and-Slab Lasso (PSSL), a hierarchical prior structure that enables group selection at the unit–variable level while preserving dynamic interdependencies across units. Unlike global shrinkage priors such as the Bayesian Lasso, which impose uniform penalisation across coefficients, PSSL introduces component-specific, data-driven shrinkage through spike-and-slab priors. This adaptive mechanism improves the separation of relevant and irrelevant predictors and allows the model to capture both sparse and dense interaction patterns.

The proposed approach also advances recent developments such as the Stochastic Search Specification Selection (S4) prior. Although S4 accommodates dynamic interdependencies (DI) and cross-sectional heterogeneity (CSH) through group-based shrinkage, it relies on Gaussian-mixture priors with fixed variances and common inclusion probabilities within groups. Such soft pooling can obscure economically meaningful within-group differences when units exhibit divergent dynamics. By contrast, the fully hierarchical spike-and-slab specification of PSSL allows inclusion probabilities to vary across units and variables, yielding sharper sparsity and more informative posterior inclusion measures.

This feature enhances the model's ability to represent heterogeneous cross-sectional dynamics and contributes to improved estimation and forecasting performance, particularly in panels characterised by uneven spillover structures.

Monte Carlo evidence shows that the proposed framework delivers notable gains in estimation accuracy, especially in environments where only a subset of potential interdependencies is active. The adaptive shrinkage mechanism allows PSSL to remain robust to changes in the degree and distribution of sparsity, mitigating overfitting while preserving economically relevant signals.

Two empirical applications illustrate the practical value of the method. In the analysis of euro area sovereign bond markets, PSSL uncovers a hierarchical structure of financial linkages characterised by dense interactions among core economies and more limited transmission within the periphery. In an international macroeconomic forecasting exercise using a 12-country panel, the model delivers stable and competitive predictive performance across horizons, often outperforming standard shrinkage and factor-based benchmarks while avoiding large forecast errors.

Overall, the PSSL framework provides a flexible and empirically grounded approach to modelling high-dimensional panel systems. By combining hierarchical variable selection with adaptive shrinkage and interpretable posterior inclusion measures, the method offers a useful tool for analysing heterogeneous interdependencies in macroeconomic and financial environments. Its structure is readily applicable to settings in which spillovers and cross-sectional heterogeneity play a central role, including regional business cycles, international policy transmission, and global financial networks.

7 Disclosure statement

The authors declare that there are no conflicts of interest.

8 Data Availability Statement

Replication data and code will be made available upon acceptance or upon reasonable request.

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 (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.

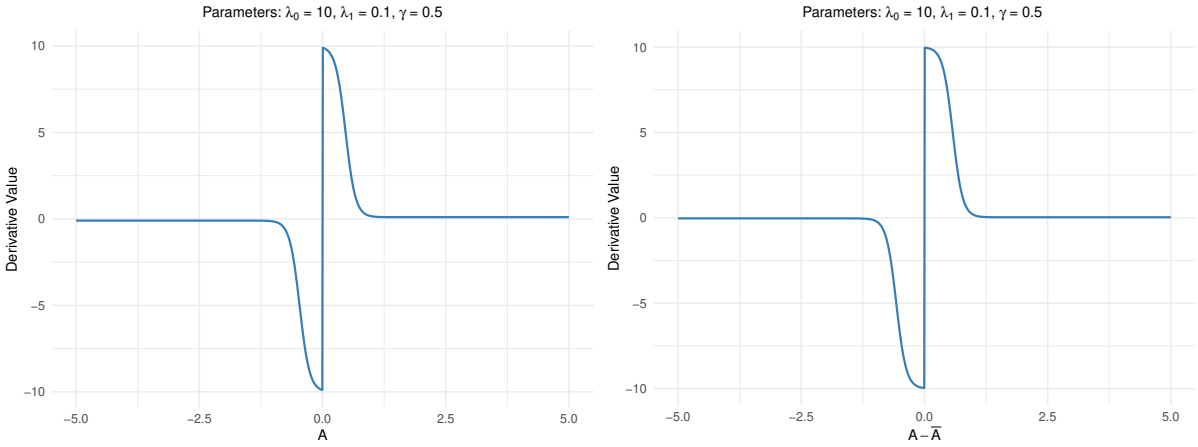
Song, Song and Peter J Bickel (2011) "Large vector auto regressions," *arXiv preprint arXiv:1106.3915*.

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.



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

Figure 1: Plot of the derivative of the penalty functions

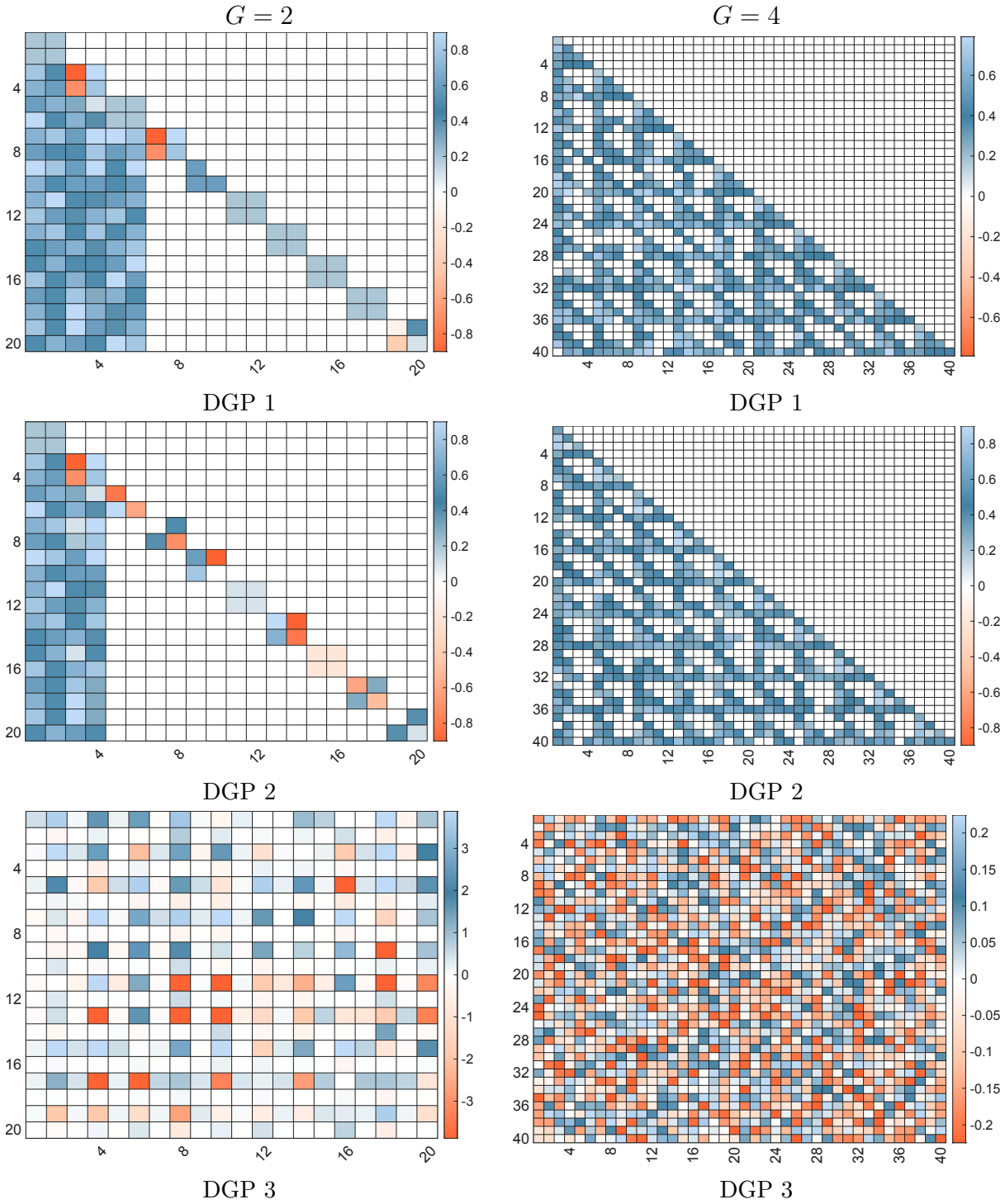


Figure 2: Autoregressive Coefficients of the Simulated DGPs. Each row corresponds to a different DGP (from top to bottom: DGP 1, DGP 2, DGP 3). The left column corresponds to $G = 2$, and the right column to $G = 4$. Zero coefficients are shown in white, while non-zero coefficients are shaded from pink to blue, with darker shades indicating higher values. Darker blue indicates larger positive values, and darker orange indicates larger negative values.

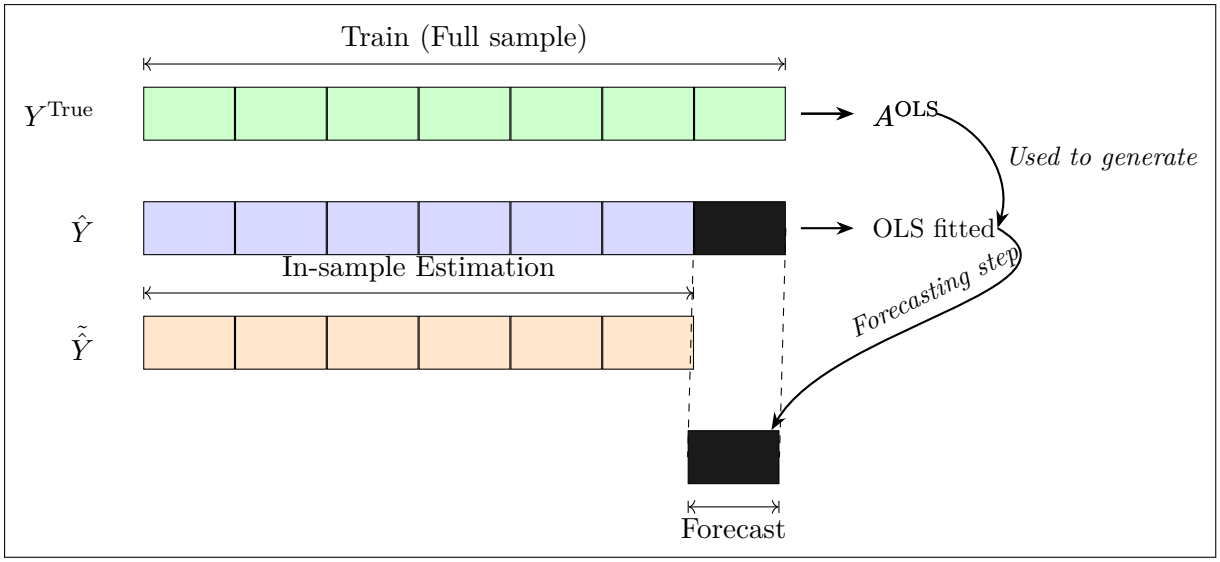


Figure 3: Real-World DGP Setting

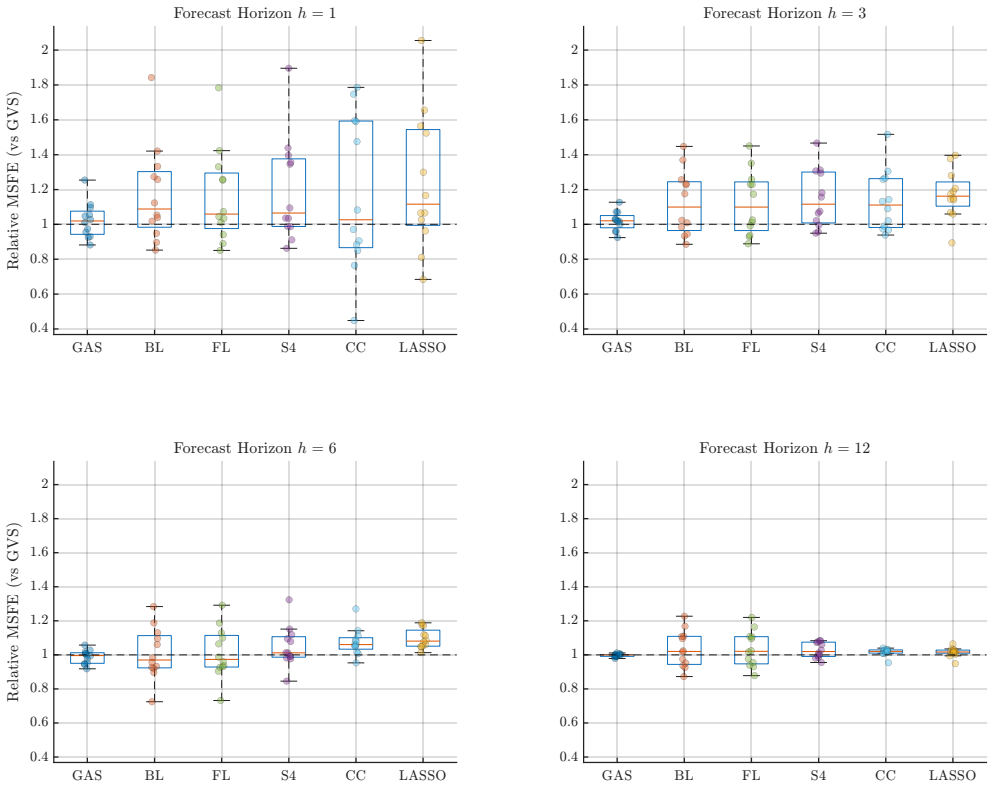


Figure 4: *Country-level distribution of relative MSFEs at horizons* $h = 1, 3, 6, 12$. MSFEs are reported relative to the benchmark (GVS). Boxplots show the cross-country distribution at horizons $h = 1, 3, 6, 12$; dots denote individual countries. Values below one indicate better performance; the dashed line marks parity with the benchmark.

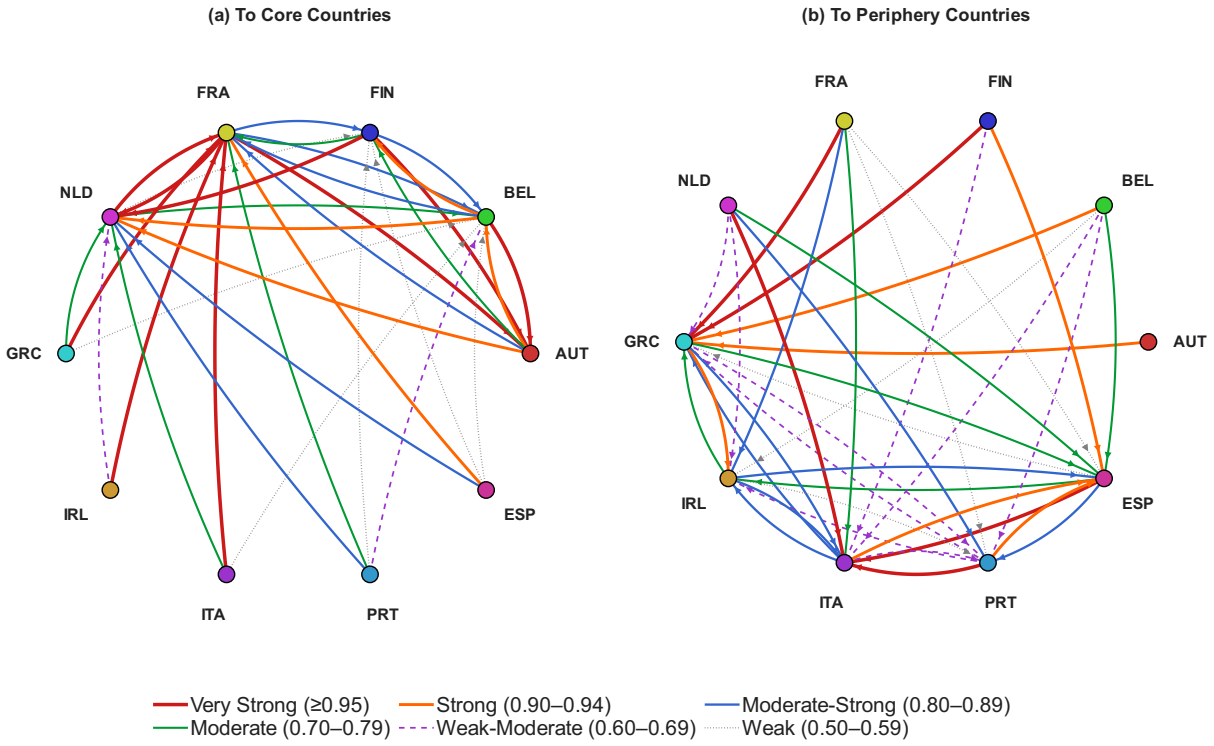


Figure 5: Posterior Dynamic Interdependencies Across the Euro Area

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.

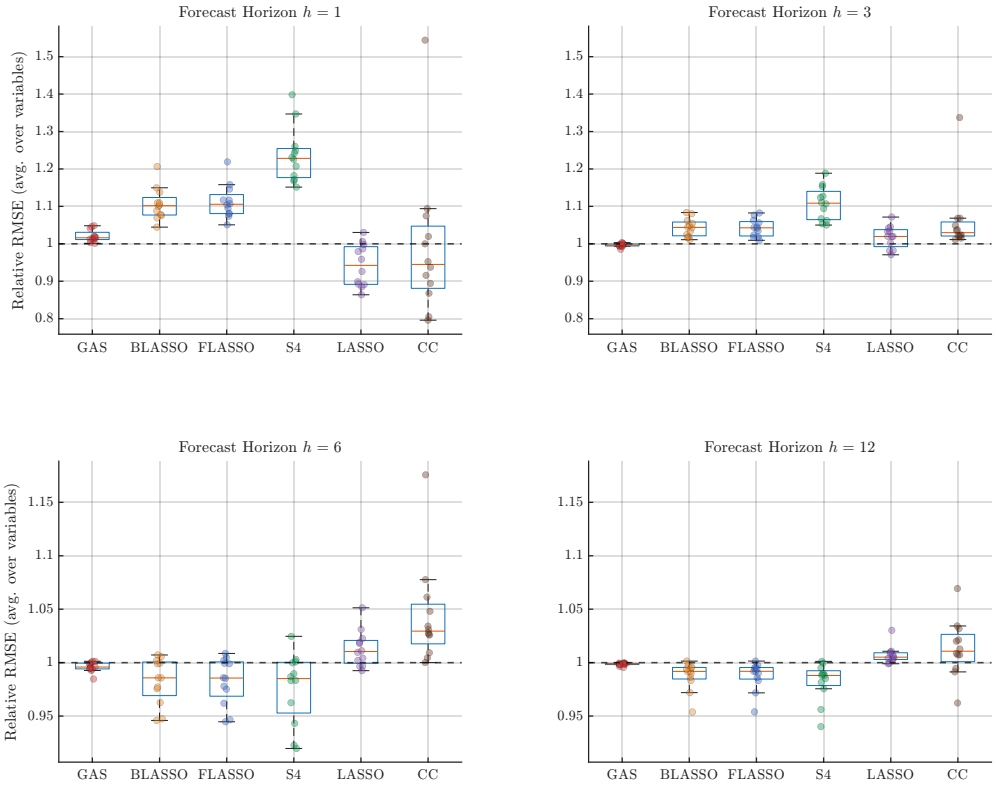


Figure 6: Country-Level Distribution of Relative MSFE for Forecast Horizons $h = 1, 3, 6, 12$.

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 MSFE / Benchmark (GVS) values across all countries. Each dot represents a single country's relative MSFE 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.

Supplementary Material

for

Adaptive Bayesian Shrinkage of High-Dimensional Panel VARs

Zhiruo Zhang Firmin Doko Tchatoka Qazi Haque

School of Economics, College of Business and Law, Adelaide University, Australia

A.1 Model Setup

This appendix provides further details on the model specification used in the paper. 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

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

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. $A^{p}_{ij,t}$ 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

$$y_{it} = A_{i,1}Y_{t-1} + \cdots + A_{i,p}Y_{t-p} + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathcal{N}(0, \Sigma_{ii}), \tag{A.1.1}$$

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

A.1.1 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 $\Gamma_{i,j}$:

$$\Gamma = \prod_{i=1}^{N-1} \prod_{j=i+1}^{N} \Gamma_{i,j}. \tag{A.1.2}$$

$\Gamma_{i,j}$ 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 $[N^2 \times N^2]$. 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:

$$A = \begin{bmatrix} A^1 \\ A^2 \\ A^3 \end{bmatrix} = \begin{bmatrix} A^{11} & A^{12} & A^{13} \\ A^{21} & A^{22} & A^{23} \\ A^{31} & A^{32} & A^{33} \end{bmatrix}. \tag{A.1.3}$$

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:

$$\bar{\mathbf{A}} = \begin{bmatrix} \bar{\mathbf{A}} & 0 & 0 \\ 0 & \bar{\mathbf{A}} & 0 \\ 0 & 0 & \bar{\mathbf{A}} \end{bmatrix}. \tag{A.1.4}$$

In this example, the dimensions of the restriction matrix of Γ_{ij} are shown in equation (A.1.5) are [9×9]. 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.1.5), 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.1.5), 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.

$$\Gamma_{1,3}A = \begin{bmatrix} \gamma_{13}^{CSH} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-\gamma_{13}^{CSH} \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} A^{11} - \bar{\mathbf{A}} \\ A^{12} \\ A^{13} \\ A^{21} \\ A^{22} - \bar{\mathbf{A}} \\ A^{23} \\ A^{31} \\ A^{32} \\ A^{33} - \bar{\mathbf{A}} \end{bmatrix} = \begin{bmatrix} A^{33} - \bar{\mathbf{A}} \\ A^{12} \\ A^{13} \\ A^{21} \\ A^{22} - \bar{\mathbf{A}} \\ A^{23} \\ A^{31} \\ A^{32} \\ A^{33} - \bar{\mathbf{A}} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{A}} \\ A^{12} \\ A^{13} \\ A^{21} \\ A^{22} \\ A^{23} \\ A^{31} \\ A^{32} \\ \bar{\mathbf{A}}. \end{bmatrix} \tag{A.1.5}$$

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.2 Proofs for the DI and CSH Penalties

Notation conventions used throughout this appendix. We write $\theta^{DI} = \{\theta_{ij}^{DI}\}_{i \neq j}$ and $\theta^{CSH} = \{\theta_{ii}^{CSH}\}_{i=j}$ for the collections of block-level inclusion parameters, with each component lying in $(0, 1)$ by Definitions 1 and 2. The symbols Φ_0, Φ_1 denote the spike and slab densities of whichever penalty is under discussion: in DI proofs $\Phi_k(a) = (\lambda_k/2)\exp(-\lambda_k\|a\|_1)$, in CSH proofs $\Phi_k(v) = (\eta_k/2)\exp(-\eta_k\|v\|_1)$. Throughout these proofs, Φ_0 and Φ_1 are interpreted as the block-level scalar-norm Laplace kernels defined in Definitions 1 and 2; hence the ratios Φ_0/Φ_1 contain the factors λ_0/λ_1 and η_0/η_1, respectively. This local re-use mirrors the parallel role of the spike–slab construction in the two penalties and never causes ambiguity within a given proof. All derivatives are taken with respect to the scalar norm $s := \|\mathbf{A}_p^{ij}\|_1$ (or $\|\Delta\mathbf{A}_p^{ii}\|_1$): we view pen_{DI} and pen_{CSH} as functions of this scalar through the dependence of Φ_0, Φ_1 on it.

A.2.1 Proof of Proposition 1

Proof of Proposition 1. To simplify notation, write $a := \mathbf{A}_p^{ij}$, $\theta := \theta_{ij}^{DI} \in (0,1)$, and $p_\gamma(\cdot) := \mathbb{P}_{\gamma_{p,ij}^{DI}=1}(\cdot)$. By Definition 1,

$$\pi(a \mid \theta) = \theta\, \Phi_1(a) + (1-\theta)\, \Phi_0(a), \qquad \Phi_k(a) = \tfrac{\lambda_k}{2} \exp(-\lambda_k \|a\|_1), \quad k \in \{0,1\}.$$

First, factoring $\Phi_1(a)$ from the numerator and $\Phi_1(\mathbf{0})$ from the denominator,

$$\log \frac{\pi(a \mid \theta)}{\pi(\mathbf{0} \mid \theta)} = \log \frac{\Phi_1(a)}{\Phi_1(\mathbf{0})} + \log \frac{\pi(a \mid \theta)/\Phi_1(a)}{\pi(\mathbf{0} \mid \theta)/\Phi_1(\mathbf{0})}. \tag{A.2.1}$$

Second, since $\Phi_1(a)/\Phi_1(\mathbf{0}) = \exp(-\lambda_1 \|a\|_1)$,

$$\log \frac{\Phi_1(a)}{\Phi_1(\mathbf{0})} = -\lambda_1 \|a\|_1. \tag{A.2.2}$$

Third, the definition of the posterior inclusion probability (2.20) gives $\pi(a \mid \theta)/\Phi_1(a) = \theta / p_\gamma(a)$, which is valid since $\theta \in (0,1)$ and $\Phi_1 > 0$. Applying this identity at both a and $\mathbf{0}$, the common factor θ cancels and

$$\log \frac{\pi(a \mid \theta)/\Phi_1(a)}{\pi(\mathbf{0} \mid \theta)/\Phi_1(\mathbf{0})} = \log \frac{p_\gamma(\mathbf{0})}{p_\gamma(a)}. \tag{A.2.3}$$

Substituting (A.2.2) and (A.2.3) into (A.2.1) and restoring the original notation,

$$\mathrm{pen}_{DI}(\mathbf{A}_p^{ij} \mid \theta_{ij}^{DI}) = -\lambda_1 \, \|\mathbf{A}_p^{ij}\|_1 + \log \frac{\mathbb{P}_{\gamma_{p,ij}^{DI}=1}(\mathbf{0})}{\mathbb{P}_{\gamma_{p,ij}^{DI}=1}(\mathbf{A}_p^{ij})}. \tag{A.2.4}$$

which is (2.23).

Finally, by the conditional independence of the DI prior across blocks, the joint prior factorizes as $\pi(\mathbf{A} \mid \theta^{DI}) = \prod_{i \neq j} \prod_p \pi(\mathbf{A}_p^{ij} \mid \theta_{ij}^{DI})$, so the log-ratio in (2.21) separates additively over blocks. Summing (A.2.4) over all (i,j,p) with $i \neq j$ yields the global form

$$\mathrm{pen}_{DI}(\mathbf{A} \mid \theta^{DI}) = -\lambda_1 \, \|\mathbf{A}\|_1 + \sum_{i \neq j}^{R} \sum_{p=1}^{P} \log \frac{\mathbb{P}_{\gamma_{p,ij}^{DI}=1}(\mathbf{0})}{\mathbb{P}_{\gamma_{p,ij}^{DI}=1}(\mathbf{A}_p^{ij})}, \tag{A.2.5}$$

where $\|\mathbf{A}\|_1 = \sum_{i \neq j}^{R} \sum_{p=1}^{P} \|\mathbf{A}_p^{ij}\|_1$.

This completes the proof.

A.2.2 Proof of Corollary 1

Proof of Corollary 1. To simplify notation, write $a := \mathbf{A}_p^{ij}$ and $\theta := \theta_{ij}^{DI}$. We view pen_{DI} as a function of the scalar $s := \|a\|_1 \geq 0$ through the dependence of $\Phi_0(a), \Phi_1(a)$ on s, and differentiate with respect to s for $s > 0$. The constant term $\log \pi(\mathbf{0} \mid \theta^{DI})$ does not depend on s, and by the block-additive decomposition established in (A.2.5) only the (i, j, p)-block term in $\log \pi(\mathbf{A} \mid \theta^{DI})$ depends on s; hence the derivative reduces to that of $\log \pi(a \mid \theta)$.

First, since $\partial \Phi_k(a)/\partial s = -\lambda_k \Phi_k(a)$ for $k \in \{0, 1\}$,

$$\begin{aligned} \frac{\partial \,\mathrm{pen}_{DI}(\mathbf{A} \mid \theta^{DI})}{\partial s} &= \frac{1}{\pi(a \mid \theta)} \frac{\partial \,\pi(a \mid \theta)}{\partial s} \\ &= \frac{-\theta\, \lambda_1\, \Phi_1(a) - (1-\theta)\, \lambda_0\, \Phi_0(a)}{\pi(a \mid \theta)}. \end{aligned} \tag{A.2.6}$$

Second, by the definition of the posterior inclusion probability in (2.20), the weight $\theta\, \Phi_1(a)/\pi(a \mid \theta)$ equals $\mathbb{P}_{\gamma_{p,ij}^{DI}=1}(a) =: w(a)$, and the complementary weight satisfies $(1-\theta)\, \Phi_0(a)/\pi(a \mid \theta) = 1 - w(a)$.

Finally, splitting the numerator in (A.2.6) and substituting the weights identified in the previous step,

$$\frac{\partial \,\mathrm{pen}_{DI}(\mathbf{A} \mid \theta^{DI})}{\partial \,\|\mathbf{A}_p^{ij}\|_1} = -\Big[\lambda_1\, w(\mathbf{A}_p^{ij}) + \lambda_0 \left(1 - w(\mathbf{A}_p^{ij})\right)\Big] = -\,\lambda^*_{\theta_{ij}^{DI}}(\mathbf{A}_p^{ij}),$$

which is (2.24)–(2.25).

A.2.3 Proof of Proposition 2

Proof of Proposition 2. To simplify notation, write $v := \Delta\mathbf{A}_p^{ii}$, $\theta := \theta_{ii}^{CSH} \in (0, 1)$, and $p_\gamma(\cdot) := \mathbb{P}_{\gamma_{p,ii}^{CSH}=1}(\cdot)$. By Definition 2,

$$\pi(v \mid \theta) = \theta\, \Phi_1(v) + (1-\theta)\, \Phi_0(v), \qquad \Phi_k(v) = \tfrac{\eta_k}{2} \exp(-\eta_k \|v\|_1), \quad k \in \{0, 1\},$$

where, throughout this proof, Φ_0, Φ_1 denote the CSH spike and slab densities (with parameters η_k), not their DI counterparts.

First, factoring $\Phi_1(v)$ from the numerator and $\Phi_1(\mathbf{0})$ from the denominator,

$$\log\frac{\pi(v \mid \theta)}{\pi(\mathbf{0} \mid \theta)} = \log\frac{\Phi_1(v)}{\Phi_1(\mathbf{0})} + \log\frac{\pi(v \mid \theta)/\Phi_1(v)}{\pi(\mathbf{0} \mid \theta)/\Phi_1(\mathbf{0})}. \tag{A.2.7}$$

Second, since $\Phi_1(v)/\Phi_1(\mathbf{0}) = \exp(-\eta_1\|v\|_1)$,

$$\log\frac{\Phi_1(v)}{\Phi_1(\mathbf{0})} = -\eta_1\|v\|_1. \tag{A.2.8}$$

Third, the definition of the posterior inclusion probability (2.28) gives $\pi(v \mid \theta)/\Phi_1(v) = \theta/p_\gamma(v)$, which is valid since $\theta \in (0,1)$ and $\Phi_1 > 0$. Applying this identity at both v and $\mathbf{0}$, the common factor θ cancels and

$$\log\frac{\pi(v \mid \theta)/\Phi_1(v)}{\pi(\mathbf{0} \mid \theta)/\Phi_1(\mathbf{0})} = \log\frac{p_\gamma(\mathbf{0})}{p_\gamma(v)}. \tag{A.2.9}$$

Substituting (A.2.8) and (A.2.9) into (A.2.7) and restoring the original notation,

$$\mathrm{pen}_{CSH}(\Delta\mathbf{A}_p^{ii} \mid \theta_{ii}^{CSH}) = -\eta_1\,\|\Delta\mathbf{A}_p^{ii}\|_1 + \log\frac{\mathbb{P}_{\gamma_{p,ii}^{CSH}=1}(\mathbf{0})}{\mathbb{P}_{\gamma_{p,ii}^{CSH}=1}(\Delta\mathbf{A}_p^{ii})},$$

which is (2.30).

A.2.4 Proof of Corollary 2

Proof of Corollary 2. To simplify notation, write $v := \Delta\mathbf{A}_p^{ii}$ and $\theta := \theta_{ii}^{CSH}$; throughout, Φ_0, Φ_1 denote the CSH spike and slab densities. We view pen_{CSH} as a function of the scalar $s := \|v\|_1 \geq 0$ through the dependence of $\Phi_0(v), \Phi_1(v)$ on s, and differentiate with respect to s for $s > 0$. The constant term $\log\pi(\mathbf{0} \mid \theta^{CSH})$ does not depend on s, and by the block-additive decomposition (2.31) only the (i,p)-block term in $\log\pi(\Delta\mathbf{A} \mid \theta^{CSH})$ depends on s; hence the derivative reduces to that of $\log\pi(v \mid \theta)$.
First, since $\partial\Phi_k(v)/\partial s = -\eta_k\Phi_k(v)$ for $k \in \{0,1\}$,

$$\begin{aligned}\frac{\partial\,\mathrm{pen}_{CSH}(\Delta\mathbf{A} \mid \theta^{CSH})}{\partial s} &= \frac{1}{\pi(v \mid \theta)}\,\frac{\partial\,\pi(v \mid \theta)}{\partial s} \\ &= \frac{-\theta\,\eta_1\,\Phi_1(v) - (1-\theta)\,\eta_0\,\Phi_0(v)}{\pi(v \mid \theta)}.\end{aligned} \tag{A.2.10}$$

Second, by the definition of the posterior inclusion probability in (2.28), the weight

$\theta\,\Phi_1(v)/\pi(v \mid \theta)$ equals $\mathbb{P}_{\gamma_{p,ii}^{CSH}=1}(v) =: w(v)$, and the complementary weight satisfies $(1-\theta)\,\Phi_0(v)/\pi(v \mid \theta) = 1 - w(v)$.

Finally, splitting the numerator in (A.2.10) and substituting the weights identified in the previous step,

$$\frac{\partial\, \mathrm{pen}_{CSH}(\Delta\mathbf{A} \mid \theta^{CSH})}{\partial\, \|\Delta\mathbf{A}_p^{ii}\|_1} = -\Big[\eta_1\, w(\Delta\mathbf{A}_p^{ii}) + \eta_0\,\big(1 - w(\Delta\mathbf{A}_p^{ii})\big)\Big] = -\,\eta^*_{\theta_{ii}^{CSH}}(\Delta\mathbf{A}_p^{ii}),$$

which is (2.33)–(2.34).

Remark The derivatives in Corollaries 1 and 2 are taken with respect to the scalar norm $s = \|\mathbf{A}_p^{ij}\|_1$ or $s = \|\Delta\mathbf{A}_p^{ii}\|_1$. For $s > 0$, the displayed derivatives are ordinary derivatives. At $s = 0$, they are interpreted as right derivatives with respect to the scalar norm. If one instead differentiates with respect to individual coordinates, the usual ℓ_1 subgradient applies: $\partial|a_\ell| = \{\mathrm{sign}(a_\ell)\}$ for $a_\ell \neq 0$ and $\partial|a_\ell| = [-1, 1]$ for $a_\ell = 0$.

A.3 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

$$\begin{aligned}\pi\left(\mathbf{A}, \Sigma \mid \mathbf{Y}_t, \mathbf{Z}_t\right) &\propto \pi(\mathbf{Y}_t \mid \Sigma)\pi(\Sigma) \prod_{i \neq j}^{\mathcal{R}} \prod_{p=1}^{P} \pi\left(\mathbf{A}_p^{ij} \mid \gamma_{p,ij}^{DI}\right) \prod_{\mathbf{i}=\mathbf{1}}^{\mathbf{N}} \prod_{p=1}^{P} \pi\left(\mathbf{A}_p^{\mathbf{ii}} \mid \gamma_{p,\mathbf{ii}}^{CSH}\right) \\ &= \pi(\mathbf{Y}_t \mid \Sigma)\pi(\Sigma) \Bigg(\prod_{i \neq j}^{\mathcal{R}} \prod_{p=1}^{P} \left[\theta_{ij}^{DI} \Phi_1\left(\mathbf{A}_p^{ij} \mid \lambda_1\right) + (1-\theta_{ij}^{DI}) \Phi_0\left(\mathbf{A}_p^{ij} \mid \lambda_0\right)\right] \\ &\quad \times \prod_{\mathbf{i}=\mathbf{1}}^{\mathbf{N}} \prod_{p=1}^{P} \left[\theta_{\mathbf{ii}}^{CSH} \Phi_1\left(\mathbf{A}_p^{\mathbf{ii}} \bar{\mathbf{A}}_p \mid \eta_1\right) + (1-\theta_{\mathbf{ii}}^{CSH}) \Phi_0\left(\mathbf{A}_p^{\mathbf{ii}} - \bar{\mathbf{A}}_p \mid \eta_0\right)\right] \Bigg).\end{aligned} \tag{A.3.1}$$

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

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

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

$$V_A = \left((D'D)^{-1} + \Sigma \otimes Z_t' Z_t \right)^{-1}, \tag{A.3.2}$$

and μ_A as

$$\mu_A = V_A \left(\left(\Sigma^{-1} \otimes Z_t' Z_t \right) \mathbf{A}_{OLS} \right). \tag{A.3.3}$$

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

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

and

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

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**

$$\lambda_0^2 \mid \text{rest} \sim \text{Gamma}\left(n + r, \frac{1}{2} \sum_{i \neq j}^{R} \tau_{0,ij}^2 + \delta \right),$$
$$\eta_0^2 \mid \text{rest} \sim \text{Gamma}\left(C + r, \frac{1}{2} \sum_{i \neq j}^{C} \kappa_{0,ij}^2 + \delta \right).$$

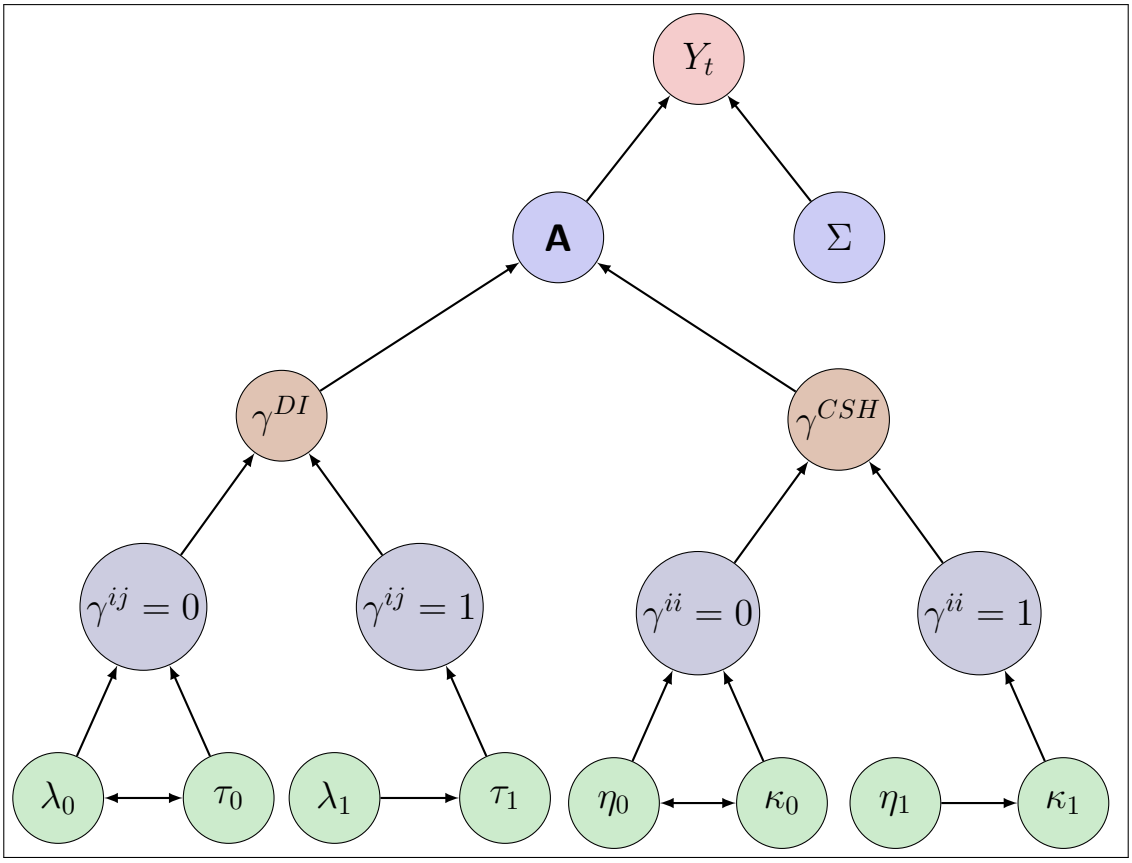


Figure A.4.1: Panel Spike-and-Slab Lasso model for inference on PVARs

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

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

A.4 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.

Depending on the data structure, we consider two alternative approaches for initialization. A simple choice sets $\lambda_0, \eta_0 \in [1, 10, 50]$, and fix $\lambda_1^2 = \eta_1^2 = 0.1$. Alternatively, in a more data-driven way, so as to better reflect the scale and sparsity of the data. The initial values can follow the formulas suggested by Park and Casella (2008):

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

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

Here, $\hat{A}^{\mathrm{OLS}}$ and $\hat{\Sigma}^{\mathrm{OLS}}$ 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:

$$\lambda_1 = \delta_\lambda \cdot \lambda_0, \quad \eta_1 = \delta_\eta \cdot \eta_0, \quad \text{where } \delta_\lambda, \delta_\eta \in [10^{-1}, 10^{-4}]. \tag{A.4.1}$$

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:

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

A.5 Additional Simulation Results

This appendix provides supplementary evidence on the finite-sample and forecasting performance of the proposed methods. We first examine estimation accuracy and bias to assess finite-sample properties. Finally, we report medium- and long-horizon forecasting results using simulated data.

A.5.1 Estimation Accuracy

Table A.5.1 reports average estimation errors for three DGPs when the group size is $G = 2$. Results are normalized to the proposed estimator (GAS). Values below one indicate improved estimation accuracy relative to GAS.

Table A.5.1: Estimation Accuracy Across Three DGPs at $G = 2$

	Spike & Slab		**Regularization**					**Frequentist**
	GAS	**GVS**	**BL**	**BFL**	**CC**	**GVAR**	**S4**	**Lasso** $(\lambda, c, \alpha, \Sigma)$
DGP 1								
$N = 5$	1.00	1.27	2.74	3.25	14.60	14.26	6.85	1.61
$N = 10$	1.00	1.05	2.41	3.04	9.31	153.00	4.86	0.98
DGP 2								
$N = 5$	1.00	1.16	3.61	4.15	16.26	16.56	7.23	1.99
$N = 10$	1.00	1.19	2.75	3.26	9.01	89.02	4.44	1.15
DGP 3								
$N = 5$	1.00	0.94	1.38	1.41	5.22	5.34	1.60	1.35
$N = 10$	1.00	8.78	2.02	5.13	64.08	85.95	5.80	5.84

Notes: The table reports average estimation errors across three DGPs (G = 2) and two sample sizes (N = 5, 10), normalized to GAS (values < 1 indicate better performance). GAS and GVS are spike-and-slab Bayesian selection methods; BL and BFL denote Bayesian shrinkage (Bayesian Lasso and Fused Lasso); CC is a factor-based model; GVAR is a global VAR with weighted foreign variables; S4 is a stochastic search variable selection method; Lasso is a frequentist penalized regression with hyperparameters $(\lambda, c, \alpha, \Sigma)$, incorporating DI and CSH penalties with shrinkage on Σ.

A.5.2 Additional Forecasting Performance

We next report additional forecasting results. Figure **??** present forecasting results for the $G = 2$ setting. In all cases, results are normalized to the proposed estimator (GAS for

Table A.5.2: Average Relative RMSFE Across Forecast Horizons for DGP1–DGP3, $N \in \{5, 10\}$, and $G = 2$

	PSSL		Other Bayesian Methods			Factor	Lasso
Horizon	GVS	GAS	BL	BFL	S4	CC	$(\lambda, c, \alpha, \Sigma)$
			Panel A: $N = 5$				
DGP1							
$h = 1$	1.00	1.05	1.58	1.56	1.08	5.28	1.16
$h = 3$	1.00	1.14	1.44	1.21	1.13	1.85	1.21
$h = 6$	1.00	1.15	1.22	1.25	0.89	1.71	1.02
$h = 12$	1.00	1.39	1.09	1.18	1.03	1.54	1.15
DGP2							
$h = 1$	1.00	1.02	1.55	1.33	0.97	6.81	1.09
$h = 3$	1.00	1.11	1.30	1.27	1.06	2.35	1.22
$h = 6$	1.00	1.33	1.25	1.21	1.19	1.49	1.14
$h = 12$	1.00	0.93	1.20	1.33	0.95	0.90	0.92
DGP3							
$h = 1$	1.00	0.91	0.70	0.81	0.69	7.63	0.68
$h = 3$	1.00	1.17	1.03	0.89	1.07	1.51	0.95
$h = 6$	1.00	1.06	0.92	0.89	0.92	1.06	0.78
$h = 12$	1.00	1.13	0.96	0.99	0.88	1.03	0.66
			Panel B: $N = 10$				
DGP1							
$h = 1$	1.00	2.29	2.42	2.51	1.10	6.80	1.01
$h = 3$	1.00	1.56	2.05	2.15	1.32	2.99	1.21
$h = 6$	1.00	1.18	1.22	1.25	0.90	1.24	0.90
$h = 12$	1.00	1.25	2.89	10.61	1.36	1.47	1.20
DGP2							
$h = 1$	1.00	1.30	1.63	1.69	1.08	4.63	1.01
$h = 3$	1.00	1.08	1.12	1.18	1.00	1.27	1.13
$h = 6$	1.00	0.88	1.01	0.99	0.82	0.99	0.94
$h = 12$	1.00	0.96	3.94	1.20	0.99	1.03	1.05
DGP3							
$h = 1$	1.00	1.00	1.05	1.41	1.63	870.32	1.00
$h = 3$	1.00	0.91	1.03	1.27	1.35	6.83	0.96
$h = 6$	1.00	0.86	0.92	1.20	1.05	3.06	1.15
$h = 12$	1.00	1.06	1.19	1.33	0.88	2.22	1.07

Notes: The table reports average relative root mean squared forecast errors (RMSFE) across countries and variables. Each entry is computed as $\text{RMSFE}_m/\text{RMSFE}_{\text{GVS}}$, so values below one indicate an improvement relative to the GVS benchmark. GAS denotes an alternative PSSL specification; BL, BFL, and S4 denote alternative Bayesian shrinkage methods; CC is a factor-based model; GVAR is a global VAR with weighted foreign variables; and Lasso is a penalized regression model with tuning parameters $(\lambda, c, \alpha, \Sigma)$.

$G = 2$), with values below one indicating superior performance relative to the baseline.

A.5.3 Data-Based Simulation Table

To complement the boxplots presented in the main text, this appendix provides the corresponding numerical results from the data-based simulations. The tables report relative MSFE values across specifications, with entries normalized to the proposed estimator. Values below one indicate better forecasting performance relative to the baseline.

Table A.5.3: Country-Level Relative MSFE at $h = 1$ (Normalized to GVS)

Country	**PSSL**		**Bayesian Methods**					**Lasso**
	GVS	**GAS**	**BL**	**BFL**	**S4**	**CC**	**GVAR**	$\lambda, c, \alpha, \Sigma$
Austria	1.00	1.06	1.26	1.25	1.35	1.08	0.71	1.07
Belgium	1.00	0.93	1.12	1.07	0.99	0.85	1.63	1.03
Finland	1.00	1.01	0.95	0.94	0.91	0.76	0.75	0.68
France	1.00	0.88	1.02	1.01	1.04	1.79	1.57	1.66
Greece	1.00	0.93	1.27	1.26	1.40	1.47	1.44	1.17
Ireland	1.00	1.25	1.42	1.42	1.44	1.75	2.40	1.52
Italy	1.00	0.96	0.90	0.89	0.99	0.97	3.35	1.30
Netherlands	1.00	0.97	1.33	1.33	1.36	1.59	2.06	1.56
Portugal	1.00	1.10	0.85	0.85	0.86	0.91	1.05	1.07
Spain	1.00	1.11	1.84	1.78	1.90	1.60	5.16	2.06
United Kingdom	1.00	1.05	1.05	1.04	1.03	0.88	0.42	0.96
United States	1.00	1.03	1.04	1.03	1.09	0.45	0.59	0.81

Note: The table reports country-level relative Mean Squared Forecast Errors (MSFE) at forecast horizon $h = 1$. Entries are computed as $\text{MSFE}_{\text{model}}/\text{MSFE}_{\text{GVS}}$, where GVS serves as the benchmark model. Methods include PSSL specifications (GVS and GAS), Bayesian shrinkage methods (BL, BFL, and S4), factor-based model (CC), global VAR with weighted foreign variables (GVAR), and a frequentist Lasso specification with shrinkage over hyperparameters λ, c, α, and Σ.

Table A.5.4: Country-Level Relative MSFE at $h = 3$ (Normalized to GVS)

Country	**PSSL**		**Bayesian Methods**					**Lasso**
	GVS	**GAS**	**BL**	**BFL**	**S4**	**CC**	**GVAR**	$\lambda, c, \alpha, \Sigma$
Austria	1.00	1.02	1.01	1.01	1.08	1.26	0.68	1.17
Belgium	1.00	0.96	1.37	1.35	1.31	1.14	1.08	1.28
Finland	1.00	1.00	0.89	0.89	0.96	1.13	0.44	1.07
France	1.00	0.96	1.23	1.23	1.31	1.52	1.23	1.38
Greece	1.00	1.03	1.45	1.45	1.47	1.30	1.24	1.40
Ireland	1.00	1.03	1.18	1.17	1.16	1.02	0.79	1.19
Italy	1.00	1.02	0.99	0.99	1.02	0.97	3.09	1.15
Netherlands	1.00	1.02	1.26	1.26	1.29	1.27	0.93	1.14
Portugal	1.00	1.07	0.93	0.93	1.00	0.99	1.24	1.21
Spain	1.00	1.13	1.23	1.23	1.18	1.09	1.38	1.14
United Kingdom	1.00	1.07	0.94	0.94	0.95	0.94	0.55	0.89
United States	1.00	0.92	1.02	1.03	1.06	0.97	0.53	1.06

Note: The table reports country-level relative Mean Squared Forecast Errors (MSFE) at $h = 3$. 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), factor-based model (CC), global VAR with weighted foreign variables (GVAR), and a frequentist Lasso with shrinkage over hyperparameters λ, c, α, and Σ.

Table A.5.5: Country-Level Relative MSFE at $h = 6$ (Normalized to GVS)

Country	**PSSL**		**Bayesian Methods**					**Lasso**
	GVS	**GAS**	**BL**	**BFL**	**S4**	**CC**	**GVAR**	$\lambda, c, \alpha, \Sigma$
Austria	1.00	0.92	1.28	1.29	1.32	1.27	1.56	1.17
Belgium	1.00	0.95	1.06	1.06	1.09	1.05	0.84	1.08
Finland	1.00	1.01	0.92	0.93	0.99	1.09	0.53	1.06
France	1.00	1.06	0.72	0.73	0.85	1.06	1.73	1.17
Greece	1.00	1.02	0.93	0.94	0.99	1.06	0.72	1.12
Ireland	1.00	1.01	0.90	0.90	0.98	1.08	1.05	1.08
Italy	1.00	1.00	0.98	0.99	1.01	1.01	2.07	1.05
Netherlands	1.00	0.95	1.19	1.19	1.15	1.14	1.14	1.19
Portugal	1.00	1.03	0.93	0.93	0.98	1.02	0.63	1.05
Spain	1.00	0.99	0.96	0.96	1.01	1.05	0.64	1.05
United Kingdom	1.00	0.95	1.10	1.10	1.12	1.11	0.86	1.11
United States	1.00	0.98	1.13	1.13	1.08	0.95	0.73	1.01

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), factor-based model (CC), global VAR with weighted foreign variables (GVAR), and a frequentist Lasso with shrinkage over hyperparameters λ, c, α, and Σ.

Table A.5.6: Country-Level Relative MSFE at $h = 12$ (Normalized to GVS)

Country	**PSSL**		**Bayesian Methods**					**Lasso**
	GVS	**GAS**	**BL**	**BFL**	**S4**	**CC**	**GVAR**	$\lambda, c, \alpha, \Sigma$
Austria	1.00	1.00	1.02	1.02	1.01	1.01	0.52	1.00
Belgium	1.00	0.98	1.11	1.11	1.07	1.03	1.92	1.03
Finland	1.00	1.00	0.93	0.93	0.98	1.03	0.55	1.02
France	1.00	1.00	0.97	0.98	1.00	1.02	0.69	1.02
Greece	1.00	0.99	1.17	1.16	1.08	1.01	0.75	0.99
Ireland	1.00	1.00	0.95	0.96	1.00	1.03	1.36	1.03
Italy	1.00	0.99	1.10	1.09	1.08	1.04	8.96	1.06
Netherlands	1.00	1.00	1.11	1.11	1.06	1.02	0.53	1.02
Portugal	1.00	1.01	0.94	0.94	0.98	1.01	1.18	1.02
Spain	1.00	1.00	1.02	1.02	1.03	1.02	0.93	1.03
United Kingdom	1.00	1.01	0.87	0.88	0.96	1.03	0.60	1.02
United States	1.00	0.99	1.23	1.22	1.08	0.95	1.24	0.95

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), factor-based model (CC), global VAR with weighted foreign variables (GVAR), and a frequentist Lasso with shrinkage over hyperparameters λ, c, α, and Σ.

A.6 Posterior Probability and Strength of Evidence

Table A.6.1 summarizes the interpretation of posterior probabilities in terms of the qualitative strength of evidence.

Table A.6.1: 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

A.7 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 6.

Table A.7.2 displays the posterior probabilities for the dynamic interdependencies directed toward core countries (Austria, Belgium, Finland, France, and the Netherlands), while Table A.7.1 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 A.7.3 to Table A.7.6 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 A.7.1: 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 A.7.2: 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 A.7.3: Country-Level Relative MSFE at $h = 1$ (Normalized to GVS)

Country	**PSSL**		**Bayesian Methods**					**Lasso**
	GVS	**GAS**	**BL**	**BFL**	**S4**	**CC**	**GVAR**	$\lambda, c, \alpha, \Sigma$
Austria	1.00	1.02	1.08	1.08	1.18	0.94	0.62	0.98
Belgium	1.00	1.04	1.10	1.10	1.23	0.95	1.36	0.88
Finland	1.00	1.01	1.14	1.15	1.35	0.89	2.05	0.93
France	1.00	1.02	1.09	1.09	1.23	1.07	0.73	1.01
Greece	1.00	1.01	1.04	1.05	1.17	1.02	3.11	1.03
Ireland	1.00	1.00	1.07	1.07	1.15	0.87	11.51	0.89
Italy	1.00	1.01	1.08	1.08	1.17	1.00	0.79	0.99
Netherlands	1.00	1.01	1.10	1.11	1.21	0.81	1.48	0.86
Portugal	1.00	1.00	1.11	1.12	1.25	1.09	2.16	1.00
Spain	1.00	1.05	1.11	1.12	1.24	1.54	0.81	0.96
United Kingdom	1.00	1.05	1.21	1.22	1.40	0.80	1.13	0.90
United States	1.00	1.02	1.15	1.16	1.26	0.92	0.60	0.89

Note: The table reports country-level relative Mean Squared Forecast Errors (MSFE) at forecast horizon $h = 1$. Entries are computed as $\text{MSFE}_{\text{model}}/\text{MSFE}_{\text{GVS}}$, where GVS serves as the benchmark model. Methods include PSSL specifications (GVS and GAS), Bayesian shrinkage methods (BL, BFL, and S4), factor-based model (CC), global VAR with weighted foreign variables (GVAR), and a frequentist Lasso specification with shrinkage over hyperparameters λ, c, α, and Σ.

Table A.7.4: Country-Level Relative MSFE at $h = 3$ (Normalized to GVS)

Country	**PSSL**		**Bayesian Methods**					**Lasso**
	GVS	**GAS**	**BL**	**BFL**	**S4**	**CC**	**GVAR**	$\lambda, c, \alpha, \Sigma$
Austria	1.00	1.00	1.02	1.02	1.06	1.02	0.63	1.00
Belgium	1.00	1.00	1.05	1.04	1.11	1.05	1.80	1.07
Finland	1.00	1.00	1.08	1.08	1.19	1.02	2.82	1.03
France	1.00	0.99	1.04	1.04	1.09	1.02	0.69	1.02
Greece	1.00	0.99	1.04	1.05	1.11	1.02	2.66	0.97
Ireland	1.00	1.00	1.06	1.06	1.13	1.02	19.01	0.98
Italy	1.00	0.99	1.01	1.01	1.07	1.04	1.04	1.03
Netherlands	1.00	0.99	1.03	1.03	1.12	1.04	1.95	1.05
Portugal	1.00	1.00	1.02	1.02	1.05	1.07	2.55	1.02
Spain	1.00	1.00	1.02	1.02	1.05	1.34	0.94	1.04
United Kingdom	1.00	1.00	1.05	1.06	1.15	1.07	0.94	1.02
United States	1.00	1.00	1.08	1.08	1.16	1.01	0.79	0.98

Note: The table reports country-level relative Mean Squared Forecast Errors (MSFE) at $h = 3$. 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), factor-based model (CC), global VAR with weighted foreign variables (GVAR), and a frequentist Lasso with shrinkage over hyperparameters λ, c, α, and Σ.

Table A.7.5: Country-Level Relative MSFE at $h = 6$(Normalized to GVS)

Country	**PSSL**		**Bayesian Methods**					**Lasso**
	GVS	**GAS**	**BL**	**BFL**	**S4**	**CC**	**GVAR**	$\lambda, c, \alpha, \Sigma$
Austria	1.00	1.00	1.01	1.00	1.00	1.01	0.62	1.00
Belgium	1.00	1.00	0.99	0.99	0.99	1.03	1.77	1.02
Finland	1.00	1.00	1.00	1.00	0.99	1.00	2.51	1.00
France	1.00	1.00	1.00	1.00	1.00	1.00	0.58	1.01
Greece	1.00	0.99	0.96	0.96	0.94	1.05	3.06	1.00
Ireland	1.00	0.99	0.99	0.99	0.98	1.06	12.99	0.99
Italy	1.00	1.00	0.98	0.98	0.96	1.03	1.02	1.01
Netherlands	1.00	1.00	0.98	0.98	0.98	1.03	1.99	1.03
Portugal	1.00	0.98	0.95	0.95	0.92	1.08	2.51	1.02
Spain	1.00	0.99	0.95	0.94	0.92	1.18	0.99	1.05
United Kingdom	1.00	1.00	1.01	1.01	1.02	1.03	1.06	1.00
United States	1.00	0.99	1.00	1.00	1.00	1.03	0.73	1.02

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), factor-based model (CC), global VAR with weighted foreign variables (GVAR), and a frequentist Lasso with shrinkage over hyperparameters λ, c, α, and Σ.

Table A.7.6: Country-Level Relative MSFE at $h = 12$ (Normalized to GVS)

Country	**PSSL**		**Bayesian Methods**					**Lasso**
	GVS	**GAS**	**BL**	**BFL**	**S4**	**CC**	**GVAR**	$\lambda, c, \alpha, \Sigma$
Austria	1.00	1.00	1.00	1.00	1.00	0.99	0.62	1.00
Belgium	1.00	1.00	0.99	0.99	0.99	1.02	1.85	1.00
Finland	1.00	1.00	1.00	1.00	0.99	1.01	2.72	1.00
France	1.00	1.00	0.99	0.99	0.98	1.01	0.68	1.01
Greece	1.00	1.00	0.99	0.99	0.99	1.01	2.78	1.00
Ireland	1.00	1.00	0.97	0.97	0.96	1.03	12.96	1.01
Italy	1.00	1.00	1.00	1.00	0.99	1.01	0.97	1.00
Netherlands	1.00	1.00	0.99	0.99	0.99	0.99	1.88	1.01
Portugal	1.00	1.00	0.98	0.98	0.98	1.03	2.61	1.01
Spain	1.00	1.00	0.95	0.95	0.94	1.07	0.95	1.03
United Kingdom	1.00	1.00	1.00	1.00	1.00	0.96	1.23	1.00
United States	1.00	1.00	0.99	0.99	1.00	1.02	0.70	1.01

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), factor-based model (CC), global VAR with weighted foreign variables (GVAR), and a frequentist Lasso with shrinkage over hyperparameters λ, c, α, and Σ.