Stochastic search selection for heterogeneous panel data models

Andreas Pick*
Erasmus University Rotterdam

January 27, 2026

Abstract

This paper presents a method for selecting variables and determining parameter heterogeneity in Bayesian hierarchical panel data models. Mixture distributions are used as priors for the mean and the variance of the individuals' parameters. Selection indicators determine the best fitting component of each mixture distribution and indicate whether the mean parameter is non-zero and whether the parameters are heterogeneous. The method is applied to two panel data sets. The first is on inflation of US CPI sub-indices and the results suggest that a heterogeneous panel AR model with lagged, first principal component is the preferred model. A second application to house price inflation across US metropolitan statistical areas shows that the model includes either the autoregressive component or the lagged spatial components, but not both at the same time.

JEL classification C11, C33, C52
Keywords Panel data, parameter heterogeneity, variable selection.

*Erasmus School of Economics, Erasmus University Rotterdam, andreas.pick@cantab.net. I would like to thank the late Herman van Dijk, Aryan Gowda, Teresa Leitão, Agnieszka Markiewicz, Hashem Pesaran, and Aisha Schmidt for insightful comments and discussions, and the associate editor, John C. Chao, and two anonymous referees for helpful comments and suggestions.

1 Introduction

This paper proposes a Bayesian method for selecting variables and determining parameter heterogeneity in panel data models. The selection method uses mixture distributions for the priors of two hyperparameters: the mean and the variance of the individuals' parameters. Each mixture distribution contains two components: a narrow distribution that implies a hyperparameter that is effectively zero and a relatively flat distribution that allows for a large hyperparameter. The component of the respective mixture is selected using a binary selection indicator. The posterior probability of the selection indicator reflects the relative fit of each mixture component to the distribution of the hyperparameter, indicating either a negligible hyperparameter size or a relatively large hyperparameter.

The first mixture is for the variance of the individual specific parameters, which is typically modeled using an inverse Gamma or, in a multivariate context, an inverse Wishart distribution. A complication with these distributions is that they do not have parameterizations that imply uninformative priors (Gelman, 2006). Since one of the two components of the prior mixture needs to allow for comparatively large mean parameters or variances, it needs to be a relatively vague prior. I, therefore, model parameter variances using the scale mixture formulation of Huang and Wand (2013). Using the scale mixture, the prior distribution of each parameter's standard deviation is the half-t distribution. The scale parameter of the half-t distribution can be set arbitrarily large, to yield a vague prior.

The second mixture is for the mean of the individuals' parameters. Here, I use the idea of stochastic search variable selection (SSVS) of George and McCulloch (1993), which uses a mixture of two normal distributions: a normal distribution that has substantial support only for a narrow region around zero and another that has support for a large region of the real line. Together with the mixture for the parameter variance, this is a panel SSVS method: variables with zero mean and parameter homogeneity can then be excluded from the model.

The proposed method is applied to a panel data set of monthly inflation rates of US CPI sub-indices and a panel data set of quarterly house price inflation of US metropolitan statistical areas (MSAs). In the first application, panel SSVS suggests a heterogeneous panel AR model with the lagged common factor. In the second application on house price inflation, the autoregressive component again appears to have a non-zero mean and heterogeneous parameters. Furthermore, spatial lags and lags of regional and country-wide averages only appear important when the autoregressive coefficient is not in the model.

This paper addresses a long standing issue in the panel data literature: parameter heterogeneity affects inference and forecasting when using panel data. While in the static linear model the mean parameter can be estimated

without bias (Zellner, 1969), Pesaran and Smith (1995) show that this is not true in heterogeneous dynamic models where neglected parameter heterogeneity will lead to biased estimates. Pesaran et al. (2000) demonstrate this in the context of modeling international private savings. Pesaran et al. (2026) analyze the forecast performance of pooled and heterogeneous panel data models. They show that parameter heterogeneity, if not accounted for, can severely reduce forecast accuracy; see Pick and Timmermann (2024) for a survey of the literature on forecasting with panel data.

In the Bayesian context, the hierarchical model of Lindley and Smith (1972) has been widely used to model parameter heterogeneity, in particular since it lends itself to estimation via the Gibbs sampler (Gelfand et al., 1990). Hsiao et al. (1999) demonstrate the usefulness of the hierarchical model for dynamic heterogeneous panel data models. Albert and Chib (1993) apply the idea of the hierarchical model to generalized linear models, such as the probit model, and the method proposed here can equally be employed in such models.

The heterogeneity model, discussed by Frühwirth-Schnatter et al. (2004), extends the hierarchical model to allow for clustering of distinct groups of cross-section units. In each group, the parameter vectors have hyperdistributions as in the hierarchical model of Lindley and Smith (1972) with group specific mean vectors and covariance matrices. See Frühwirth-Schnatter (2011) and Grün (2018) for reviews of the heterogeneity model. The variable and heterogeneity selection of this paper can therefore either be applied to each group separately or the selection can be pooled across groups.

The approach developed here is inspired by SSVS of George and McCulloch (1993) for variable selection in cross-sectional models. SSVS has previously been used for panel data models, where SSVS is used on each of the parameters separately, which contrasts with the modeling of parameter heterogeneity via the covariance matrix in this paper. For instance, Gilbride et al. (2006) apply SSVS to the parameters of a panel regression for each cross-section unit separately. Cuaresma et al. (2016) and Dovern et al. (2016) use the SSVS in Bayesian global VAR models, where the SSVS prior on each coefficient is used in an approach akin to shrinkage.

Similarly, Koop and Korobilis (2016) address model uncertainty in panel VAR models by using SSVS on the different parameters of the panel VAR. They distinguish parameter restrictions on three distinct groups of parameters: first, restrictions that set the parameters of lags of variable of one unit in the equation for the variables of another unit to zero; second, restrictions that equate the covariance matrix of errors that determine contemporaneous correlations between variables of different units to zero; and finally, restrictions that, across units, equate the parameters of lags of the dependent variables for a given unit. These restrictions can be modeled with the panel SSVS suggested here. For example, the last restriction of parameter homogeneity is set up via numerous pairwise comparisons. The mixture of inverse

Wishart distributions suggested in this paper targets the same restriction in a more parsimonious manner.

SSVS priors have also been used as shrinkage priors for forecasting. Cross et al. (2020) use the SSVS priors on VAR models, as introduced by George et al. (2008), and find that they are competitive with other shrinkage priors. Extensions of the approach here to forecasting are equally possible.

A nonparametric version of SSVS for the panel data model has been proposed by Kim et al. (2009). They use spiked Dirichlet process priors for the individual parameters to determine inclusion of the respective regressor for each cross-section unit separately. The same idea can be extended for the mean parameter and parameter variance in the approach in this paper. This is, however, beyond the scope of the current paper.

The paper proceeds as follows. The next section discusses the Bayesian hierarchical panel data model. Section 3 introduces the stochastic search selection for heterogeneity and variable selection. Section 4 applies the method to two panel data sets, inflation rates of sub-indices of US CPI and house price inflation rates for US MSAs. Finally, Section 5 concludes. The Gibbs sampler is described in Appendix A.

2 The Bayesian panel data model

Consider the panel data model

$$y_{it} = \boldsymbol{\beta}_i' \mathbf{x}_{it} + \varepsilon_{it}, \quad \varepsilon_{it} \sim \mathrm{N}\left(0, \eta_i^{-1}\sigma_\varepsilon^2\right) \tag{1}$$

where $\boldsymbol{\beta}_i$ is a $K \times 1$ vector of parameters, $\mathbf{x}_{it}$ is a $K \times 1$ vector of regressors, including the intercept, $i = 1, 2, \ldots, N$ denotes cross section units, and $t = 1, 2, \ldots, T$ denotes time periods. Furthermore, we stack the observations to obtain $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iT})'$ and $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{iT})'$.

The coefficient vector, $\boldsymbol{\beta}_i$, is modeled via the hierarchical model of Lindley and Smith (1972). Hyperpriors connect the individuals' coefficient vectors, such that

$$\boldsymbol{\beta}_i \sim \mathrm{N}\left(\bar{\boldsymbol{\beta}}, \boldsymbol{\Sigma}\right) \tag{2}$$

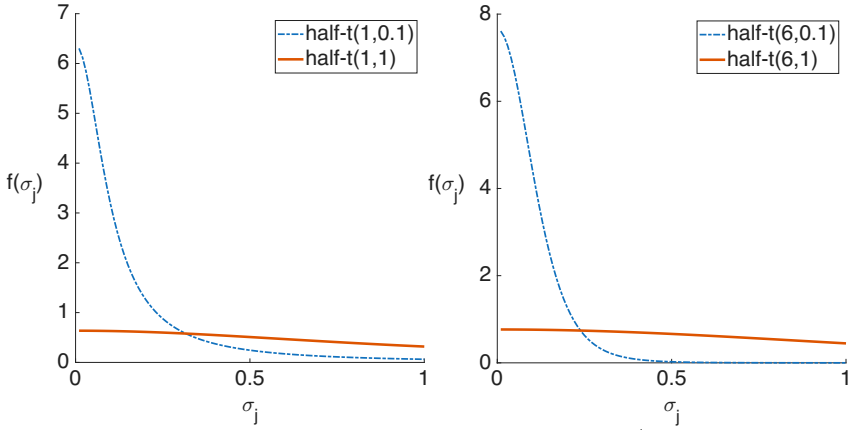
where $\bar{\boldsymbol{\beta}}$ is a $K \times 1$ vector of mean parameters and $\boldsymbol{\Sigma}$ a $K \times K$ covariance matrix. When using conjugate prior distributions, the parameters in the hierarchical model are easily estimated using the Gibbs sampler (Gelfand et al., 1990).

The prior for the mean parameter vector, $\bar{\boldsymbol{\beta}}$, is the normal with prior parameters $\boldsymbol{b}_0$ and $\boldsymbol{S}_0$,

$$\bar{\boldsymbol{\beta}} \sim \mathrm{N}\left(\boldsymbol{b}_0, \boldsymbol{S}_0\right) \tag{3}$$

where $\boldsymbol{b}_0$ is a $K \times 1$ vector of mean parameters and $\boldsymbol{S}_0$ a $K \times K$ covariance matrix.

Figure 1: Prior mixtures of half-t distributions



Note: The left plot displays two half-Cauchy distributions (half-t distributions with $\nu_\sigma = 1$) with $a_j = 0.1$ and $a_j = 1$ and the right plot displays two half-t distributions with $\nu_\sigma = 6$, $a_j = 0.1$ and $a_j = 1$.

The prior for the parameter covariance matrix, $\mathbf{\Sigma}$, is typically the inverse Wishart distribution. It is, however, difficult to formulate uninformative priors for the inverse Wishart prior (Gelman, 2006). For this reason, I use the prior scale mixture of Huang and Wand (2013), where

$$\mathbf{\Sigma}|\boldsymbol{A} \sim \text{invWishart}\left(\nu_\sigma + K - 1, 2\nu_\sigma \boldsymbol{A}^{-1}\right) \tag{4}$$

where ν_σ is the degrees of freedom, $\boldsymbol{A}$ is a diagonal scale matrix. The jth element on the diagonal of $\boldsymbol{A}$, α_j, is independently distributed as

$$\alpha_j \sim \text{invGamma}\left(1/2, 1/a_j^2\right) \tag{5}$$

where the scale parameters, a_j, $j = 1, 2, \ldots, K$ are the prior parameters that determine the shape of the scale mixture distribution. Huang and Wand show that the mixture implies a half-$t(\nu_\sigma, a_j)$ marginal distribution for each standard deviation in $\mathbf{\Sigma}$. This, in turn, suggests that arbitrarily large values of a_j lead to relatively uninformative priors for the standard deviation.

Figure 1 plots two half-t distributions: in the left plot, ν_σ is set to 1, which implies a half-Cauchy distribution, and $a_j = 0.1$ or 1. In the right plot $\nu_\sigma = 6$ and, again, $a_j = 0.1$ or 1. In both plots, the larger a_j leads to a comparatively flat distribution, which support substantially larger values of σ_j compared to the distributions with $a_j = 0.1$.

Finally, I model the distribution of the error term, ε_{it}, via a mixture distribution with inverse Gamma-2 priors for the variance, σ_ε^2, and Gamma-2 priors for the mixture term, $\eta_{\varepsilon i}$,

$$\sigma_\varepsilon^2 \sim \text{invGamma}_2\left(\nu_\varepsilon, \nu_\varepsilon s_\varepsilon^2\right) \quad \text{and} \quad \eta_{\varepsilon i} \sim \text{Gamma}_2\left(\nu_\eta, \nu_\eta\right)$$

with prior parameters ν_ε, s_ε^2 and ν_η. Geweke (1993) shows that this implies a t-distribution for the error term.

3 Stochastic search selection

3.1 Parameter heterogeneity

A question is whether the parameters in the panel model (2) are, in fact, heterogeneous, that is, whether the variance parameters are non-zero. To answer this question, I use a two component mixture distribution as the prior of each diagonal element in $\mathbf{\Sigma}$, σ_j^2, which is defined using the scale coefficient a_j in the prior distribution in (4) and (5).

Since priors for the covariance matrix in (4) and (5) imply half-$t(\nu_\sigma, a_j)$ distributions for the standard deviations, σ_j, we can define the first mixture component with a small value of a_j, say a_{0j}, to have most of its mass close to zero. The second mixture component with a large value of a_j, say a_{1j} yields a distribution that offers substantial support for larger σ_j. The first component is then a prior that implies a negligible parameter variance, such that homogeneity of the parameter is a reasonable conclusion, whereas the second component implies parameter heterogeneity.

Formally, the mixture is

$$\sigma_j|\kappa_j \sim \kappa_j \text{half-}t(\nu_\sigma, a_{1j}) + (1-\kappa_j)\text{half-}t(\nu_\sigma, a_{0j}) \tag{6}$$

where κ_j is a latent, binary selection indicator with $\mathrm{P}(\kappa_j = 1) = 1 - \mathrm{P}(\kappa_j = 0) = p_j$. Thus, $\kappa_j = 0$ implies half-$t(\nu, a_{0j})$ and, when $\kappa_j = 1$, the prior distribution for parameter j is half-$t(\nu, a_{1j})$. The posterior probability for $\kappa_j = 1$ is then the probability that parameter j is heterogeneous in the model.

Two crucial parameters are a_{0j} and a_{1j}, since they determine the magnitude of parameter heterogeneity that is allowed under the half-t prior distributions. While the half-t distribution with $a_j = 1$, plotted in Figure 1, is comparatively flat, the distribution using $a_j = 0.1$ only has meaningful mass for small values of σ_j. Choices for these two hyperparameters are discussed in Section 3.3.

A second choice concerns the value of the degrees of freedom, ν_σ. The half-Cauchy distribution on the left of Figure 1 is less informative when a_j is large compared to a half-t distribution with $\nu_\sigma = 6$. The disadvantage of the half-Cauchy distribution is, however, that the distinction between the distributions with different a_j is less pronounced at the crossing point, which is given by

$$\phi_j = \left(\nu_\sigma \frac{a_{0j}^{2/(\nu_\sigma+1)} - a_{1j}^{2/(\nu_\sigma+1)}}{a_{1j}^{-2\nu_\sigma/(\nu_\sigma+1)} - a_{0j}^{-2\nu_\sigma/(\nu_\sigma+1)}} \right)^{1/2}$$

This may lead to a less sharp distinction between homogeneity and heterogeneity of a parameter. Since uninformativeness of the prior with large a_j is not a requirement for the selection, larger ν_σ could be used in order to maintain the relatively sharper distinction between the two distributions around the crossing point.

Furthermore, Huang and Wand (2013) show for the example of a 2×2 covariance matrix, that the implied marginal prior for the correlation coefficient is

$$p(\rho) \propto (1-\rho^2)^{\nu_\sigma/2-1}$$

For $\nu_\sigma = 1$, this is a U-shaped distribution with most of the support for values close to -1 and 1, for $\nu_\sigma = 2$ the distribution is uniform, and for $\nu_\sigma > 2$ the distribution has an inverted U-shape where the support is increasingly concentrated around 0 as ν_σ increases. This suggests that values of ν_σ greater than unity may be preferable.

The model is completed with priors for the latent variable, κ_j, for which I assume independent Bernoulli priors

$$p(\boldsymbol{\kappa}) = \prod_{j=1}^{K} p_j^{\kappa_j}(1-p_j)^{(1-\kappa_j)}$$

where $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \ldots, \kappa_K)'$.

Since the distributions are conjugate, estimation via the Gibbs sampler is straightforward, and estimates of the posterior probability is easy to obtain. The conditional distributions required for the Gibbs sampler are detailed in Appendix A.

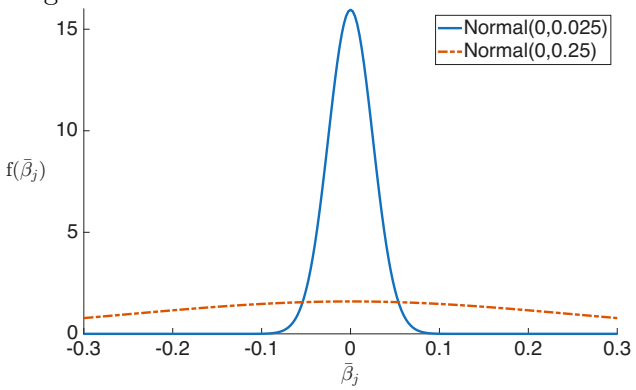
3.2 Variable selection

The mixture for the parameter variances can be combined with mixture distributions for the mean parameter vector, $\bar{\boldsymbol{\beta}}$, in order to analyze variable inclusion. The analysis for the mean parameter vector uses the idea of SSVS of George and McCulloch (1993). Since the exclusion of variables from the model requires parameter homogeneity and a zero mean parameter, SSVS is applied to each mean parameter, $\bar{\beta}_j,\ j = 1, 2, \ldots, K$. The prior for the mean parameter in (3) is replaced by a mixture of normals

$$\bar{\beta}_j|\gamma_j \sim \gamma_j \mathrm{N}\left(0, h_{1j}^2\right) + (1-\gamma_j)\mathrm{N}\left(0, h_{0j}^2\right) \tag{7}$$

where γ_j is a binary selection indicator, similar to κ_j in the mixture (6). If $\gamma_j = 1$, the first mixture with variance h_{1j}^2 is selected and the second component with variance h_{0j}^2 is selected when $\gamma_j = 0$. We again set h_{0j} to a small value, which leads to a narrow distribution around 0. If h_{0j} yields a sufficiently narrow distribution, it can be assumed that $\bar{\beta}_j$ is zero. The first

Figure 2: Prior mixture of normal distributions



Note: The plot displays two normal distributions that enter the mixture for $\bar{\beta}_j$ with standard deviation 0.025 for the first distribution in the mixture and 0.25 for the second distribution in the mixture.

component with large h_{1j} implies a flatter tailed distribution around zero, such that non-zero values for $\bar{\beta}_j$ are supported.

Note that the two distributions intersect at

$$|\xi_j| = h_{1j} h_{0j} \sqrt{2 \frac{\log(h_{1j}/h_{0j})}{h_{1j}^2 - h_{0j}^2}} \tag{8}$$

The choice of the intersection point, which is determined by h_{0j} and h_{1j}, inform the prior magnitude of the coefficient that justifies inclusion or exclusion of the variable. We will turn to the choice of hyperparameters in Section 3.3.

Similar to the case of half-t distributions above, one could use a prior distribution that is uninformative, such as extremely large h_{1j}^2 or a Cauchy distribution. Also here, however, an uninformative prior will not necessarily improve selection compared to a prior that allows for larger variances and has a sharper distinction to the narrower distribution around the crossing point. For this reason, I will use normal distributions in the mixture for $\bar{\boldsymbol{\beta}}_j$, which is in line with the literature on SSVS cited above.

The hierarchical model can be extended by putting priors on h_{0j}, h_{1j}, and $q_j = \mathrm{P}(\gamma_j)$ (Cross et al., 2020) and on a_{0j}, a_{1j}, and p_j. For simplicity, however, I will keep these parameters fixed in the Gibbs sampler.

The K univariate mixtures can be combined into a multivariate mixture distribution,

$$\bar{\boldsymbol{\beta}} | \boldsymbol{\gamma} \sim \mathrm{N}\left[\mathbf{0}, \boldsymbol{H}(\boldsymbol{\gamma}) \boldsymbol{R} \boldsymbol{H}(\boldsymbol{\gamma})\right] \tag{9}$$

where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_K)'$ is a $K \times 1$ vector of binary indicators, $\boldsymbol{R}$ is a prior correlation matrix, and $\boldsymbol{H}(\boldsymbol{\gamma})$ is a $K \times K$ diagonal matrix with j-th diagonal element h_{1j} if $\gamma_j = 1$ and h_{0j} if $\gamma_j = 0$. In the applications below, I will assume that the mean parameters are a priori uncorrelated and set $\boldsymbol{R} = \boldsymbol{I}_K$, where $\boldsymbol{I}_K$ is the K-dimensional identity matrix.

Finally, the model requires a prior distribution for $\boldsymbol{\gamma}$ and I assume prior independence of γ_j from γ_l and κ_l, $\forall j, l$

$$\mathrm{p}(\boldsymbol{\gamma}) = \prod_{j=1}^{K} q_j^{\gamma_j}(1-q_j)^{1-\gamma_j}$$

If $\gamma_j = \kappa_j = 0$, the variable can be removed from the model since all parameters, $\bar{\beta}_j$, are homogenous and zero. If $\gamma_j = 1$ and $\kappa_j = 0$, variable j should be included but the parameter can be modeled as homogeneous. If $\kappa = 1$, the variable has heterogeneous coefficients and is part of the model irrespective of the value of γ_j.

3.3 Choosing the hyperparameters

The hyperparameters h_{0j} and h_{1j} for the mean parameter vector and the hyperparameters a_{0j}, a_{1j} and ν_σ for the parameter covariance matrix need to be chosen by the user.

Mean parameters The crossing points of the two hyperprior distributions, $|\xi_j|$, is given in (8). For $\bar{\boldsymbol{\beta}}$, we can lean on the insights of George and McCulloch (1993; 1995; 1997) and Chipman et al. (2001) for SSVS in the linear model. George and McCulloch (1993) consider intersections of the marginal distributions of β_j: $\mathrm{N}(0, \sigma^2_{\beta,j} + h^2_{0j})$ and $\mathrm{N}(0, \sigma^2_{\beta,j} + h^2_{1j})$, using least squares estimate of the parameter variances, $\sigma^2_{\beta,j}$ together with the tuning parameters h_{0j} and h_{1j}.

In the case of the hierarchical model, a similar approach would use the least squares estimate of the parameter covariance matrix

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{LS}} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\beta}}_i - \bar{\hat{\boldsymbol{\beta}}}\right)\left(\hat{\boldsymbol{\beta}}_i - \bar{\hat{\boldsymbol{\beta}}}\right)' - \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1} \tag{10}$$

where $\hat{\boldsymbol{\beta}}_i$ is the least squares estimate of $\boldsymbol{\beta}_i$ and $\bar{\hat{\boldsymbol{\beta}}} = \frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\beta}}_i$ (Maddala et al., 1997). The intersection of normal distributions with variances $[\hat{\boldsymbol{\Sigma}}_{\mathrm{LS}}]_{jj} + h^2_{0j}$ and $[\hat{\boldsymbol{\Sigma}}_{\mathrm{LS}}]_{jj} + h^2_{1j}$ could then be chosen such that parameters smaller than the intersection point are economically unimportant, where $[\cdot]_{jj}$ denotes the jth element on the diagonal of a matrix.

A downside is that $\hat{\boldsymbol{\Sigma}}_{\mathrm{LS}}$ need not be positive definite in finite sample. Using

$$\tilde{\boldsymbol{\Sigma}}_{\mathrm{LS}} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\beta}}_i - \bar{\hat{\boldsymbol{\beta}}}\right)\left(\hat{\boldsymbol{\beta}}_i - \bar{\hat{\boldsymbol{\beta}}}\right)' \tag{11}$$

guarantees positive definiteness but may overestimate the parameter variances in finite sample.

An alternative is to estimate the intersection of the posterior distributions of the mean parameter using the Gibbs sampler, where $\bar{\beta}_j$ is sampled for fixed h_{0j} and h_{1j}. The intersection of the posterior distributions is then calculated for each draw in the Gibbs sampler. Since the posterior distributions need not be centred around the same posterior mean, only one intersection point may be meaningful whereas the other can be at a point that is in the tails of both distributions. I determine the relevant intersection point by plotting the posterior densities of $\bar{\beta}_j$. In the applications below, I report the intersection that is at a high probability region of the posterior together with the intersection of the prior distributions. Choices of h_{0j} can be proportional to $[\hat{\boldsymbol{\Sigma}}_{\text{LS}}]_{jj}^{1/2}$ and $h_{1j} = c \cdot h_{0j}$, for $c = 10$ or 100.

Parameter variances As discussed above, the choice of ν_σ determines the informativeness of the prior distribution, the sharpness of the distinction at the crossing point of the two prior distributions in the mixture, and the implied correlation of the Wishart distribution. In order to balance these three considerations, I will use $\nu_\sigma = 6$ in both applications.

One can then choose a_{0j} and a_{1j} to obtain a desired intersection of the prior distributions in the mixture in (3.1). Additionally, as in the case of the mean parameter vector, it is informative to consider the implied intersection of posterior distributions. Conditional on a_{0j} and on a_{1j}, we can draw two parameter variances and determine the intersection of their distribution in the Gibbs sampler. Since the conditional posterior of σ_j is an inverse Gamma distribution, the intersection is $(d_0 - d_1)/[\nu \log(d_0/d_1)]$, where $d_0 = Ns^2_{\beta,j} + 2\nu_\sigma [\boldsymbol{A}_0^{-1}]_{jj}$, $d_1 = Ns^2_{\beta,j} + 2\nu_\sigma [\boldsymbol{A}_1^{-1}]_{jj}$, $s^2_{\beta,j}$ is the j-th diagonal element of $\frac{1}{N}\sum_{i=1}^{N} \left(\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}}\right)\left(\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}}\right)'$, and $\nu = N + \nu_\sigma$.

Choices for a_{0j} prior to the evaluation using the Gibbs sampler can be based on the observation that, in the absence of parameter heterogeneity and for large N, an estimate of the parameter covariance matrix is $\boldsymbol{S} = \frac{1}{N}\sum_{i=1}^{N} \sigma_i^2 (\boldsymbol{X}_i' \boldsymbol{X}_i)^{-1}$. Therefore, a_{0j}^2 can be set proportional to the j-th diagonal element of $\boldsymbol{S}$, and a_{1j}^2 to $c \cdot a_{0j}^2$, where $c = 100$ or 1000.

4 Application to CPI sub-indices and house prices

This section applies the panel SSVS method outlined above to a panel of inflation rates of sub-indices of US CPI and to a panel of house price inflation in metropolitan statistical areas (MSAs) in the USA. The data sets have previously been employed by Pesaran et al. (2026). The results for each application are based on 20,000 iterations of the Gibbs sampler, detailed in Appendix A, with the first 5000 discarded as burn-in.

4.1 Application to inflation of US CPI sub-indices

The first application analyses the inflation rate of sub-indices of US CPI obtained from the FRED data base. I restrict attention to the period January 2006 to December 2022, since the data for all sub-indices is complete with this exception of two sub-indices, which I exclude. This results in a panel data set of $N = 184$ sub-indices for $T = 202$ observations after accounting for the pre-sample.

Pesaran et al. (2026) use an autoregressive panel model with the addition of the lagged first common factor of the data set obtained via PCA for the entire data set, the lagged default yield, and the lagged term spread:

$$y_{it} = \beta_{0i} + \beta_{1i} y_{i,t-1} + \beta_{2i} \hat{f}_{t-1} + \beta_{3i} dy_{t-1} + \beta_{4i} ts_{t-1} + \varepsilon_{it} \tag{12}$$

where y_{it} denotes the monthly inflation rate of sub-index i in period t, $\hat{f}_{t-1}$ is the lagged common factor extracted from the CPI sub-indices using the information up to period $t-1$, dy_{t-1} is the lagged default yield and ts_{t-1} the lagged term spread. For further details see Pesaran et al. (2026).

Since the inclusion of a heterogeneous intercept is not in question, I demean the data for each individual separately in order to focus on the remaining parameters. The inclusion priors are set to $p(\gamma_j) = p(\kappa_j) = 1/K = 0.25$ for each mean parameter and parameter variance. The additional priors are $\nu_\varepsilon = 10$, $s_\varepsilon^2 = 1$, and $\nu_\eta = 2$.

I set a_{0j}^2 to $c_{a0}[\boldsymbol{S}]_{jj}$ with c_{a0} either 0.001 or 0.1 and $a_{1j}^2 = c_{a1} a_{0j}^2$ with $c_{a1} = 1000$, except the second variable where I set c_{a0} either 0.00001 or 0.001. Furthermore, $h_{0j} = c_{h0}[\hat{\boldsymbol{\Sigma}}_{\text{LS}}]_{jj}^{1/2}$ with c_{h0} either 0.02 or 0.2, and $h_{1j} = c_{h1} h_{0j}$ with $c_{h1} = 100$. These choices are made to obtain mixtures that lead to inclusion and exclusion outcomes for each variable, assuming this can be achieved for reasonably sized parameter values. The contrast will allow me to judge the size of intersection of the posteriors that leads to one outcome over another.

Table 1 reports the results, where the upper panel shows the results for the mean parameters and the lower panel shows the results for the parameter variances. For each parameter, the results consist of three lines. The first line reports the (positive) intersection of the prior distributions. In the second line is the (relevant) intersection of the posterior distributions. Finally, the third line gives the posterior probabilities of $\gamma_j = 1$ for the mean parameters and $\kappa_j = 1$ for the parameter variances. For each variable, the table has four columns that correspond to the different settings of the tuning parameters.

A first observation is that for the mean parameter the specification of the prior for the variance has relatively little influence on the inclusion probabilities. Equally, for the inclusion probabilities of the variance component in the lower panel, the priors for the mean parameter play only a minor role.

Table 1: Inflation of CPI sub-indices: inclusion probabilities

Mean parameters, $\tilde{\beta}_j$					
		$c_{h0} = 0.02$		$c_{h0} = 0.2$	
		$c_{a0} = 0.001$	$c_{a0} = 0.1$	$c_{a0} = 0.001$	$c_{a0} = 0.1$
AR	prior inter.	0.014	0.014	0.143	0.143
	posterior inter.	0.121	0.124	0.270	0.273
	$p(\gamma_j = 1)$	1.000	1.000	0.255	0.257
PCA	prior inter.	0.021	0.021	0.214	0.214
	posterior inter.	0.013	0.015	0.075	0.084
	$p(\gamma_j = 1)$	0.478	0.599	0.245	0.246
Default yield	prior inter.	0.003	0.003	0.030	0.030
	posterior inter.	−0.003	−0.004	−0.019	−0.017
	$p(\gamma_j = 1)$	0.555	0.661	0.251	0.249
Term spread	prior inter.	0.013	0.013	0.132	0.132
	posterior inter.	0.033	0.034	0.158	0.135
	$p(\gamma_j = 1)$	1.000	1.000	0.253	0.252
Parameter variance, σ_j^2					
		$c_{a0} = 0.001$		$c_{a0} = 0.1$	
		$c_{h0} = 0.02$	$c_{h0} = 0.2$	$c_{h0} = 0.02$	$c_{h0} = 0.2$
AR	prior inter.×100	0.005	0.005	0.451	0.451
	posterior inter.	0.021	0.009	0.021	0.010
	$p(\kappa_j = 1)$	0.864	1.000	0.253	0.247
PCA	prior inter.×100	0.000	0.000	0.029	0.029
	posterior inter.	0.112	0.104	0.113	0.106
	$p(\kappa_j = 1)$	1.000	1.000	0.550	0.487
Default yield	prior inter.×100	0.003	0.003	0.341	0.341
	posterior inter.	0.002	0.002	0.002	0.002
	$p(\kappa_j = 1)$	0.250	0.248	0.252	0.253
Term spread	prior inter.×100	0.014	0.014	1.410	1.410
	posterior inter.	0.041	0.036	0.040	0.036
	$p(\kappa_j = 1)$	0.285	0.285	0.254	0.254

Note: The top panel of the table reports the results for the mean parameters. The first line for each variable gives the (positive) intersection of the prior distributions, the second line the intersection of the posterior distributions, and the third line the posterior probability of $\gamma_j = 1$. The bottom panel reports the results for the parameter variances. The first line for each variable, again, reports the intersection of the prior distributions, the second line the intersection of the posterior distributions, and the third line the posterior probabilities for $\kappa_j = 1$. The results are for four combinations of tuning parameters, c_{h0} and c_{a0}. 'AR' denotes the autoregressive coefficients, 'PCA' the coefficients of the first principal component, 'Default yield' the coefficients of the default yield, and 'Term spread' the coefficients of the term spread.

Table 2: Inflation of CPI sub-indices: selected models

$c_{h0} = 0.02$, $c_{a0} = 0.001$			$c_{h0} = 0.02$, $c_{a0} = 0.1$		
p	$\gamma_j = 1$	$\kappa_j = 1$	p	$\gamma_j = 1$	$\kappa_j = 1$
16.57	1,4	1,2	10.07	1,2,3,4	2
13.65	1,2,3,4	1,2	8.41	1,2,3,4	–
9.98	1,3,4	1,2	5.31	1,3,4	2
6.43	1,2,4	1,2	4.17	1,4	–
$c_{h0} = 0.2$, $c_{a0} = 0.001$			$c_{h0} = 0.2$, $c_{a0} = 0.1$		
16.99	–	1,2	6.93	–	–
6.54	–	1,2,4	6.43	–	2
5.81	1	1,2	2.44	1	–
5.75	3	1,2	2.36	–	2,3

Note: The table reports the four models that are chosen most often in the Gibbs sampler for different combinations of tuning parameters. The column p reports the percentage of draws for which the model is selected, where a model is defined as the j for which either γ_j or κ_j equals one. The numbers in the second, third, fifth and sixth columns denote the variables in the order they are given in Table 1.

The results for the autoregressive parameter show that the mean parameter has a posterior inclusion probability, $p(\gamma_j = 1|\boldsymbol{y})$, of 1 for the smaller h_{0j} where the intersection of the posteriors is 0.12. For the larger h_{0j}, where the intersection of the posteriors is 0.27, the inclusion probability is around the level of the prior. The results for the parameter variance of the autoregressive parameter in the lower panel show that the heterogeneity is selected even if the intersection of the posterior distribution implies modest heterogeneity.

The mean coefficient for the PCA variable has an inclusion probability around 0.5 for the smaller h_{0j} where the intersection of the posteriors is 0.13 but an inclusion probability very similar to the prior for the larger h_{0j}. Parameter heterogeneity is selected for posterior intersections of about 0.1. The default yield is included only for very small values and the inclusion probability of parameter heterogeneity does not exceed the value of the prior. This suggests that the default yield does not play an important role in the model. The term spread has an inclusion probability of 1 for the small h_{0j} with an intersection of the posterior of about 0.03 and parameter heterogeneity inclusion probabilities does not exceed the level of the prior, which suggests at most a minor importance in the model.

Table 2 considers the inclusion probabilities for the variables jointly. For each combination of the tuning parameters, it gives the probabilities that

a model is selected in an iteration of the Gibbs sampler, where a model is defined by the γ_j and κ_j that equals one. It can be seen that for the smaller c_{h0}, many mean parameters are included but for the larger c_{h0} the most often chosen model sets all mean parameters to zero. Furthermore, for the smaller value of c_{a0} the autoregressive and the PCA coefficients are selected as heterogeneous. For the larger c_{a0}, the PCA coefficient is still selected as heterogeneous relatively frequently.

4.2 Application to US house price inflation

The second application analyses variable inclusion and parameter heterogeneity in a spatial autoregressive model for a panel of quarterly house price inflation in US MSAs from 1975Q1 to 2023Q1. I consider the model of Pesaran et al. (2026)

$$y_{it} = \alpha_i + \beta_i y_{i,t-1} + \gamma_i y_{i,t-1}(\boldsymbol{w}) + \delta_{Ri}\bar{y}_{i,t-1}^{(R)} + \delta_{Ci}\bar{y}_{t-1}^{(C)} + \varepsilon_{it} \tag{13}$$

where y_{it} is the house price inflation for MSA i in quarter t, $y_{it}(\boldsymbol{w}) = \sum_{k=1,k\neq i}^{N} w_{ik}y_{kt}$ is the spatially weighted average of house price inflation, where $w_{ik} = v_{ik}/\sum_{l=1}^{N} v_{il}$ and $v_{ik} = 1$ if MSA k is within 100 miles of MSA i and zero otherwise. Next, $\bar{y}_{it}^{(R)}$ is the average house price inflation in the Bureau of Economic Analysis region of MSA i, and $\bar{y}_{t}^{(C)}$ are the country-wide average house price inflation. The spatial weights are those of Yang (2021) and I exclude MSAs that have no neighboring MSA within 100 miles, which leaves $N = 362$ MSAs for $T = 188$ periods after accounting for the pre-sample.

As in the previous application, I assume the intercept should always be included and heterogeneous and therefore demean the data for each individual separately. For the remaining variables, I set the priors to $p(\gamma_j) = p(\kappa_j) = 1/K = 0.25$. The additional priors are $\nu_\varepsilon = 10$, $s_\varepsilon^2 = 1$, and $\nu_\eta = 2$. Additionally, I set a_{0j}^2 to $c_{a0}[\boldsymbol{S}]_{jj}$ with $c_{a0} \cdot 100$ either 0.025 or 0.5 and $a_{1j}^2 = c_{a1}a_{0j}^2$ with $c_{a1} = 1000$. Furthermore, $h_{0j} = c_{h0}[\hat{\boldsymbol{\Sigma}}_{\text{LS}}]_{jj}^{1/2}$ with c_{h0} either 0.05 or 0.15, except the first variable where c_{h0} is multiplied by 1.25, and $h_{1j} = c_{h1}h_{0j}$ with $c_{h1} = 100$. Again these choices are made to obtain a clear separation of inclusion and exclusion outcomes.

The results are in Table 3, which is structured in the same way as Table 1. The results in the upper panel suggest that the mean parameter of the autoregressive component has zero posterior inclusion probability when c_{h0} is small but a posterior inclusion probability of 1 when c_{h0} is large and the intersection of the posterior distribution is 0.867. This result is explained when considering the inclusion probability of the spatial autoregressive component, which is 1 for small c_{h0}, when the autoregressive coefficient was excluded, and is 0 for the large c_{h0}, when the autoregressive coefficient was included. This suggests that the two autoregressive coefficients serve a very

Table 3: House price inflation: inclusion probabilities

Mean parameters, β_j					
		$c_{h0} = 0.02$		$c_{h0} = 0.2$	
		$c_{a0} = 0.001$	$c_{a0} = 0.1$	$c_{a0} = 0.001$	$c_{a0} = 0.1$
AR	prior inter.	0.024	0.024	0.073	0.073
	posterior inter.	0.539	0.539	0.867	0.867
	$p(\gamma_j = 1)$	0.000	0.000	1.000	1.000
SAR	prior inter.	0.034	0.034	0.102	0.102
	posterior inter.	0.016	0.016	0.017	0.018
	$p(\gamma_j = 1)$	1.000	1.000	0.246	0.246
Reg.ave.	prior inter.	0.044	0.044	0.131	0.131
	posterior inter.	0.058	0.058	0.061	0.061
	$p(\gamma_j = 1)$	0.783	0.762	0.252	0.253
Country ave.	prior inter.	0.023	0.023	0.068	0.068
	posterior inter.	0.041	0.038	0.007	0.010
	$p(\gamma_j = 1)$	0.337	0.338	0.242	0.243
Parameter variance, σ_j^2					
		$c_{a0} = 0.001$		$c_{a0} = 0.1$	
		$c_{h0} = 0.02$	$c_{h0} = 0.2$	$c_{h0} = 0.02$	$c_{h0} = 0.2$
AR	prior inter.$\times 100$	0.001	0.001	0.014	0.014
	posterior inter.	0.027	0.013	0.027	0.013
	$p(\kappa_j = 1)$	0.361	0.527	0.254	0.256
SAR	prior inter.$\times 100$	0.002	0.002	0.031	0.031
	posterior inter.	0.039	0.028	0.039	0.029
	$p(\kappa_j = 1)$	0.719	0.610	0.246	0.244
Reg.ave.	prior inter.$\times 100$	0.002	0.002	0.048	0.048
	posterior inter.	0.047	0.042	0.047	0.043
	$p(\kappa_j = 1)$	0.499	0.554	0.252	0.255
Country ave.	prior inter.$\times 100$	0.002	0.002	0.031	0.031
	posterior inter.	0.011	0.011	0.011	0.011
	$p(\kappa_j = 1)$	0.508	0.469	0.250	0.254

Note: 'AR' denotes the autoregressive coefficients, 'SAR' the spatial autoregressive coefficients, 'Reg.ave.' the coefficients of regional averages, and 'Country ave.' the coefficients of country wide averages. For further information see the footnote of Table 1.

Table 4: House price inflation: selected models

$c_{h0} = 0.02$, $c_{a0} = 0.01$			$c_{h0} = 0.02$, $c_{a0} = 0.1$		
p	$\gamma_j = 1$	$\kappa_j = 1$	p	$\gamma_j = 1$	$\kappa_j = 1$
10.12	2,3	2,3,4	15.29	2,4	–
8.37	2,3	2,4	8.79	2,3,4	–
6.23	2,3	2,3	5.57	2	–
5.23	2,3,4	2,3	5.19	2,3	4
$c_{h0} = 0.2$, $c_{a0} = 0.01$			$c_{h0} = 0.2$, $c_{a0} = 0.1$		
6.32	1	2,3,4	13.18	1	–
5.39	1	1,2,3	4.75	1	3
3.86	1	1	4.63	1	1
3.61	1	1,2,3,4	4.53	1,3	–

Note: The numbers in the second, third, fifth and sixth columns denote the variables in the order they are given in Table 3. For further details, see the footnote of Table 2.

similar function in the model and only one is required. The regional averages are included for small c_{h0} with the intersection of the posterior density at 0.058. The country wide averages do not have a high inclusion probability for either value of c_{h0}.

The results for the parameter variances in the lower panel of Table 3 show relatively small inclusion probabilities for parameter heterogeneity of the autoregressive coefficients for both c_{a0}. The inclusion of parameter heterogeneity for the spatial coefficient is larger at higher intersections of the posterior densities. The regional averages also have somewhat higher inclusion probabilities for heterogeneity at posterior intersections around 0.47. For the country wide averages, inclusion probabilities are higher only at very small intersection points.

The results in Table 4, again, summarize the results per model and mirror the results above. If the spatial autocorrelation coefficient is included then the autocorrelation coefficient is excluded and vice versa.

5 Conclusion

This paper proposes a Bayesian method for selecting variables and determining parameter heterogeneity in panel data models. In a hierarchical model, the prior for the mean of the individuals' parameters is modeled using a mixture of normal distributions using the idea of SSVS of George and McCulloch (1993). The prior for the variance of the individuals' parameters is modeled using a mixture of half-t distributions. In each of the

mixtures, a binary selection indicator determines the choice of component of the mixture: one component is a distribution with substantial support only close to zero, the second component has wide support for non-zero values. The posterior probability of the selection indicators suggests whether mean parameters are non-zero and parameters are heterogeneous.

The method is applied to two panel data sets. The first application, a panel data set on monthly inflation rates of US CPI sub-indices, suggests that the data can be modeled as a heterogeneous panel AR model with the lagged principal component added with heterogeneous parameters. The second application on quarterly house price inflation in US MSAs shows that different variables play a similar role in the model. The AR component has a non-zero coefficient when the other variables are not in the model. The spatial AR variables and regional averages, in contrast, are non-zero when the autoregressive variable is not included.

Appendix A The Gibbs sampler

The Gibbs sampler iteratively draws from the following conditional distributions, where $|\cdot$ denotes that the distribution is conditional on the other parameters in the Gibbs sampler.

- $\boldsymbol{\beta}_i|\cdot \sim \mathrm{N}(\boldsymbol{b}_\mathrm{i}, \boldsymbol{S}_\mathrm{i})$, where $\boldsymbol{b}_i = \boldsymbol{S}_i \left(\eta_{\varepsilon i}\sigma_\varepsilon^{-2}\mathbf{X}_i'\mathbf{y}_i + \boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{\beta}}\right)$,
  $\boldsymbol{S}_i = \left(\eta_{\varepsilon i}\sigma_\varepsilon^{-2}\mathbf{X}_i'\mathbf{X}_i + \boldsymbol{\Sigma}^{-1}\right)^{-1}$.
- $\sigma_\varepsilon^2|\cdot \sim \text{invGamma}\left([NT+\nu_\varepsilon]/2, \frac{1}{2}\left[\sum_{i=1}^N \eta_{\varepsilon i}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_i)'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}_i) + \nu_\varepsilon\sigma_\varepsilon^2\right]\right)$.
- $\eta_{\varepsilon i}|\cdot \sim \text{Gamma}\left([\nu_\varepsilon + T]/2, \frac{1}{2}\left[\nu_\varepsilon + \sigma_\varepsilon^{-2}(\mathbf{y}_i - \mathbf{x}_i'\boldsymbol{\beta}_i)'(\mathbf{y}_i - \mathbf{x}_i'\boldsymbol{\beta}_i)'\right]\right)$.
- $\bar{\boldsymbol{\beta}}|\cdot \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{S}_{\bar{\beta}})$, where $\boldsymbol{\mu} = \boldsymbol{S}_{\bar{\beta}}N\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}_i$ and $\boldsymbol{S}_{\bar{\beta}} = \left(\boldsymbol{\Sigma}^{-1} + (\boldsymbol{HRH})^{-1}\right)^{-1}$ where $\boldsymbol{H}$ is a $K \times K$ diagonal matrix with typical element $\gamma_j h_{1j} + (1-\gamma_j)h_{0j}$ and $\boldsymbol{R}$ is a prior correlation matrix. In the applications, I use $\boldsymbol{R} = \boldsymbol{I}_K$, where $\boldsymbol{I}_K$ is the K-dimensional identity matrix.
- $\boldsymbol{\Sigma}|\cdot \sim \text{invWishart}\left(N + K + \nu_\sigma - 1, \sum_{i=1}^N \left(\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}}\right)\left(\boldsymbol{\beta}_i - \bar{\boldsymbol{\beta}}\right)' + 2\nu_\sigma \boldsymbol{A}^{-1}\right)$ where $\boldsymbol{A} = \text{diag}(\alpha_1, \alpha_2, \ldots, \alpha_K)$
- $\alpha_j|\cdot \sim \text{invGamma}\left((\nu_\sigma + K)/2, \nu_\sigma\left(\boldsymbol{\Sigma}^{-1}\right)_{jj} + a_j^{-2}\right)$ where $a_j = \kappa_j a_{1j} + (1-\kappa_j)a_{0j}$.
- $\gamma_j|\cdot \sim \text{Bernoulli}(p_\gamma)$, where $p_\gamma = p_1/(p_1 + p_0)$,
  $p_1 = f_N\left(\bar{\boldsymbol{\beta}}|\boldsymbol{\mu}, [\boldsymbol{H}(\gamma_j = 1)\boldsymbol{RH}(\gamma_j = 1)]^{-1} + N\boldsymbol{\Sigma}^{-1}\right)p(\gamma_j)$,
  $p_0 = f_N\left(\bar{\boldsymbol{\beta}}|\boldsymbol{\mu}, [\boldsymbol{H}(\gamma_j = 0)\boldsymbol{RH}(\gamma_j = 0)]^{-1} + N\boldsymbol{\Sigma}^{-1}\right)[1 - p(\gamma_j)]$,
  $f_N(\cdot)$ is the multivariate normal pdf, $\boldsymbol{H}(\gamma_j = 1)$ is $\boldsymbol{H}$ with the jth element replaced by h_{1j} and $\boldsymbol{H}(\gamma_j = 0)$ is $\boldsymbol{H}$ with the jth element replaced by h_{0j}, $\boldsymbol{R} = \boldsymbol{I}_K$ is the prior correlation matrix.
- $\kappa_j|\cdot \sim \text{Bernoulli}(p_\kappa)$, where $p_\kappa = q_1/(q_1 + q_0)$,
  $q_1 = f_{ig}\left(\alpha_j; (\nu_\sigma + K)/2, \nu_\sigma(\boldsymbol{\Sigma}^{-1})_{jj} + a_{1j}^{-2}\right)p(\kappa_j)$,
  $q_0 = f_{ig}\left(\alpha_j; (\nu_\sigma + K)/2, \nu_\sigma(\boldsymbol{\Sigma}^{-1})_{jj} + a_{0j}^{-2}\right)[1 - p(\gamma_j)]$,
  $f_{ig}(\cdot)$ is the inverse Gamma pdf.

References

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679.

Chipman, H., George, E. I., and McCulloch, R. E. (2001). The practical implementation of Bayesian model selection. In Lahiri, P., editor, *Model Selection*, pages 67–116. Institute for Mathematical Statistics.

Cross, J. L., Hou, C., and Poon, A. (2020). Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity. *International Journal of Forecasting*, 36(3):899–915.

Cuaresma, J., Feldkircher, M., and Huber, F. (2016). Forecasting with global vector autoregressive models: A Bayesian approach. *Journal of Applied Econometrics*, 31(7):1371–1391.

Dovern, J., Feldkircher, M., and Huber, F. (2016). Does joint modelling of the world economy pay off? Evaluating global forecasts from a Bayesian GVAR. *Journal of Economic Dynamics & Control*, 70:86–100.

Frühwirth-Schnatter, S. (2011). Panel data analysis: a survey on model-based clustering of time series. *Advances in Data Analysis and Classification*, 5(4):251–280.

Frühwirth-Schnatter, S., Tüchler, R., and Otter, T. (2004). Bayesian analysis of the heterogeneity model. *Journal of Business and Economic Statistics*, 26(1):2–15.

Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, 85(412):972–985.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–533.

George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.

George, E. I. and McCulloch, R. E. (1995). Stochastic search variable selection. In Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors, *Markov Chain Monte Carlo in Practice*, pages 203–214. Chapman & Hall, London.

George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, 7(2):339–373.

George, E. I., Sun, D., and Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, 142(1):553–580.

Geweke, J. F. (1993). Bayesian treatment of the independent Student-t linear model. *Journal of Applied Econometrics*, 8(S1):S19–S40.

Gilbride, T. J., Allenby, G. M., and Brazell, J. D. (2006). Models for heterogeneous variable selection. *Journal of Marketing Research*, 43(3):420–430.

Grün, B. (2018). Model-based clustering. In Frühwirth-Schnatter, S., Celeux, G., and Robert, C. P., editors, *Handbook of Mixture Analysis*, chapter 8, pages 163–198. CRC Press.

Hsiao, C., Pesaran, M. H., and Tahmiscioglu, A. K. (1999). Bayes estimation of short-run coefficients in dynamic panel data models. In Hsiao, C., Lahiri, K., Lee, L.-F., and Pesaran, M. H., editors, *Analysis of Panel and Limited Dependent Variable Models*, pages 268–296. Cambridge University Press, Cambridge.

Huang, A. and Wand, M. P. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, 8(2):439–452.

Kim, S., Dahl, D. B., and Vannucci, M. (2009). Spiked Dirichlet process prior for Bayesian multiple hypothesis testing in random effects models. *Bayesian Analysis*, 4(4):707–732.

Koop, G. and Korobilis, D. (2016). Model uncertainty in panel vector autoregressive models. *European Economic Review*, 81(1):115–131.

Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B*, 34(1):1–41.

Maddala, G. S., Trost, R. P., Li, H., and Joutz, F. (1997). Estimation of short-run and long-run elasticities of energy demand from panel data using shrinkage priors. *Journal of Business & Economic Statistics*, 15(1):90–100.

Pesaran, M. H., Haque, N. U., and Sharma, S. (2000). Neglected heterogeneity and dynamics in cross-country savings regressions. In Krishnakumar, J. and Ronchetti, E., editors, *Panel Data Econometrics—Future Directions: Papers in the Honour of Professor Balestra*, pages 53–82. Elsevier.

Pesaran, M. H., Pick, A., and Timmermann, A. (2026). Forecasting with panel data: estimation uncertainty versus parameter heterogeneity. *Quantitative Economics*, forthcoming.

Pesaran, M. H. and Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68(1):79–113.

Pick, A. and Timmermann, A. (2024). Panel data forecasting. In Clements, M. and Galvao, A., editors, *Handbook of Research Methods and Applications in Macroeconomic Forecasting*, page 68–89. Edward Elgar.

Yang, C. F. (2021). Common factors and spatial dependence: An application to US house prices. *Econometric Reviews*, 40(1):14–50.

Zellner, A. (1969). On the aggregation problem: A new approach to a troublesome problem. In Fox, K. A., Sengupta, J. K., and Narasimham, G. V. L., editors, *Economic Models, Estimation and Risk Programming*, pages 365–374. Springer Verlag, Berlin.