

A Tutorial on Forecasting with Bayesian Panel Vector Autoregressions Using the R Package bpvars

Miguel Sanchez-Martinez 
International Labour Organization

Tomasz Woźniak 
University of Melbourne

Abstract

This document is a tutorial on a family of Bayesian Panel Vector Autoregressions with multi-level hierarchical prior distribution to forecast dynamic panel data of gross domestic product, employment, unemployment, and labour market participation rates for 189 countries. It is accompanied by an R package **bpvars** by [Woźniak \(2025\)](#). It provides the basics on Bayesian inference and essential techniques employed in the models. The detailed specification for each of the models with its variations is followed by the presentation of workflows for the estimation and forecasting for R. Importantly, this document constitutes supplementary material for the manuscript by [Sanchez-Martinez and Woźniak \(2025\)](#) that provides a detailed description of the package and statistical methods implemented in it.

Keywords: Dynamic Panel Data, Unemployment Rate, Employment Rate, Labour Market Participation Rate.

Contents

Non-Technical Summary	2
Key Features of the Proposed Approach	3
A Family of Hierarchical Panel Vector Autoregression	8
Panel Vector Autoregressions with Global Prior	8
Panel Vector Autoregressions with Country Grouping	15
Panel Vector Autoregressions with Global Prior Grouping	18
Vector Autoregressions for Dynamic Panel Data	20
Forecast Performance Evaluation in R	22
Computations for Forecast Performance Evaluation	22
Density Forecast Performance Comparison	26
Point Forecast Performance Comparison	27
Forecast Reconciliation	29

1. Non-Technical Summary

This paper proposes a new forecasting model for labour market outcomes of 189 countries. The variables to be forecasted include the unemployment, employment, and labour force participation rates. The model also includes other variables that are useful for forecasting the labour market block, such as the real gross domestic product, all of which are collected over time. Predictive analyses of such constructed dynamic panel data face multiple challenges arising from complicated dynamic properties of the series. The proposed model addresses these challenges by implementing known and new modelling strategies.

The main features of the developed framework include:

- system modelling – considering the target and ancillary variables as jointly determined within the dynamic system,
- dynamic approach – capable of capturing the temporal memory in variables and their interdependencies essential for prediction,
- global-to-local formulation – improving the forecasting precision of country's future outcomes by feeding in concisely summarised global contributions,
- reduced complexity – efficiently extracting the information from the rich data set,
- embedded flexibility – making the predictions less dependent on arbitrary choices,
- risk accountability – controlling for model specification and estimation uncertainty in forecasting.

This note introduces the new model and explains its main features as well as potential extensions in relation to forecast interpretations and performance. It also provides technical details of the Bayesian estimation algorithm and its implementation in the R package **bpvars**.

This note proceeds as follows. Section 2 presents the core Bayesian techniques motivating the new model in an approachable manner. Section 3 introduces the original model and its detailed formulation. Section ?? presents Bayesian forecasting for dynamic panel data. The **bpvars** package workflow detailing each of the stages of the forecasting exercise are highlighted in Section ?. Finally, Appendix ?? presents the detail of the estimation algorithm for the sake of transparency and reproducibility.

2. Key Features of the Proposed Approach

This section provides a brief introduction to the elements of Bayesian inference that are essential for the proposed modelling and forecasting framework using the Hierarchical Panel Vector Autoregressions. These elements include prior and posterior distributions, Bayes' rule, prior shrinkage interpretation, hierarchical modelling for panel data models, estimation and forecasting. Their simple explanation is intended to make the following sections more accessible.

Prior Distributions

Prior distributions are the distinguishing feature of Bayesian inference. These distributions represent the knowledge or belief of the investigator about the parameters prior to using data. Nowadays, the properties of the inference in relationship to the prior specification are well known and used purposefully depending on the objective of the investigation. We follow [Woźniak \(2016\)](#) to look into those uses for the estimation and forecasting.

Bayesian Inference

In Bayesian inference, the central role is played by the posterior distribution, that is, the conditional distribution of the parameters of the model given the data. It is the outcome of the estimation procedure and the basis for statistical inference and forecasting. It is defined through Bayes' rule and most conveniently presented as being proportional up to a normalising constant to the product of the likelihood function and data

$$p(\theta | \mathbf{y}) \propto L(\theta; \mathbf{y})p(\theta) \quad (1)$$

where θ collects the parameters of a model and \mathbf{y} denotes sample data. The proportionality means that the RHS of (1) contains all the information about the parameters given data but it does not integrate to value one, and therefore is not a probability density function.

The Bayes' rule represents a learning mechanism in which the investigator's beliefs, represented by the prior, are updated by the information from the data, the likelihood, to form the updated beliefs, the posterior distribution:

$$\text{prior beliefs} \quad \times \quad \text{information from data} \quad \rightarrow \quad \text{updated beliefs} \quad (2)$$

The update through the data contained in the likelihood function is facilitated by the *likelihood principle* that states that the whole informational content of data with respect to the parameters of the model is expressed by the likelihood function.

The **role of the prior distribution** in forming the posterior distribution is illustrated using a normal prior distribution with the mean $\underline{\theta}$ and variance \underline{v}^2

$$\theta_i \sim \mathcal{N}(\underline{\theta}, \underline{v}^2) \quad (3)$$

for a linear Gaussian model for unit or country i , such as the following linear regression

$$y_{it} = \theta_i x_{it} + \epsilon_{it} \quad \text{where} \quad \epsilon_{it} \sim \mathcal{N}(0, 1) \quad (4)$$

with y_{it} and x_{it} denoting dependent and explanatory variables respectively, and ϵ_{it} being the standard normal error term with the variance fixed to one. In such a simple setup, the posterior distribution is normal with the mean $\bar{\theta}$ and variance \bar{v}^2 . One aspect of the prior distribution is explained in a closer analysis of the posterior mean that can be presented as a weighted mean of the prior mean $\underline{\theta}$ and the maximum likelihood estimator $\hat{\theta}_i$:

$$\bar{\theta}_i = \omega \underline{\theta} + (1 - \omega) \hat{\theta}_i \quad (5)$$

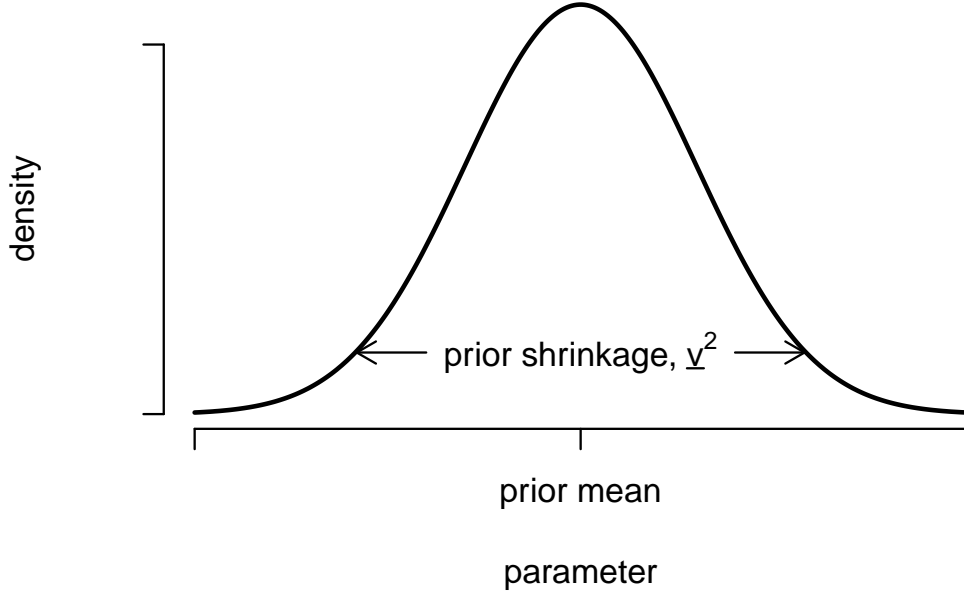


Figure 1: Illustration of prior shrinkage. Source: Woźniak (2016).

where $\omega \in [0, 1]$ and

$$\omega = \frac{\frac{v^{-2}}{T}}{\frac{v^{-2}}{T} + \frac{\sum_{t=1}^T x_{it}^2}{T}} \quad \text{and} \quad (1 - \omega) = \frac{\frac{\sum_{t=1}^T x_{it}^2}{T}}{\frac{v^{-2}}{T} + \frac{\sum_{t=1}^T x_{it}^2}{T}} \quad (6)$$

Therefore the weight on the prior mean is proportional to the prior precision divided by the sample size $\frac{v^{-2}}{T}$ and that on the maximum likelihood is proportional to the data informational content as measured by $\frac{\sum_{t=1}^T x_{it}^2}{T}$.

Two points arise from this analysis. With increasing sample size T the role of prior diminishes and in the limit, when the sample size goes to infinity, the posterior estimate is equal to the maximum likelihood estimator $\bar{\theta}_i = \hat{\theta}_i$. Moreover, in finite samples, with fixed T , stronger shrinkage (smaller prior variance \underline{v}^2) will bias the posterior estimate towards the prior mean, whereas with weaker shrinkage (relatively larger \underline{v}^2) the posterior estimate gets closer to $\hat{\theta}_i$. In the limit, when the prior precision \underline{v}^{-2} goes to zero, the posterior estimate is equal to the maximum likelihood estimator $\bar{\theta}_i = \hat{\theta}_i$.

Lessons for Bayesian Forecasting

The practice of Bayesian forecasting with Bayesian Vector Autoregressions makes an informed use of these phenomena. Following the formulation of the Minnesota prior by [Doan, Litterman, and Sims \(1984\)](#) the prior mean is set to reflect the parameters of a random walk model for a multivariate unit-root non-stationary system

$$\mathbf{y}_t = \mathbf{y}_{t-1} + \boldsymbol{\epsilon}_t. \quad (7)$$

This choice of the prior mean for the vector autoregressive parameters provides a hedging mechanism. As long as signal from data is strong, the future values are forecasted using the estimates $\hat{\theta}_i$. However, in the opposite case, to some extent, the forecasting is performed using the random walk model from the prior mean, which is a reasonable practice by all means. In this context, the choice of the level of shrinkage driven by \underline{v}^2 provides an additional possibility of improving the forecasting performance for particular data set.

Hierarchical Modelling

Hierarchical modelling expands the possibility of making the model better fitted to data through the estimation of the prior distribution hyper-parameters such as the prior mean $\underline{\theta}_i$ or its level of shrinkage \underline{v}^2 . Such estimation is possible by establishing the prior hierarchy using the conditional and marginal distributions. Consider the following extension of the prior distribution in (3)

$$\theta_i | \underline{\theta}, \underline{v}^2 \sim \mathcal{N}(\underline{\theta}, \underline{v}^2), \quad (8)$$

$$\underline{\theta} \sim \mathcal{N}(\underline{m}_\theta, \underline{v}_\theta^2), \quad (9)$$

$$\underline{v}^2 \sim \mathcal{IG2}(\underline{s}_\theta, \underline{v}_\theta). \quad (10)$$

In this specification, the dependence on the hyper-parameters is explicitly marked in (8) by the introduction of statistical conditioning. Then, prior distributions are assumed for these hyper-parameters in (9) and (10). Using this formulation results in all the three parameters, θ_i , $\underline{\theta}$, and \underline{v}^2 , being estimated.

The major benefit is the reduced reliance on arbitrary choices to be made in (3) that requires fixing hyper-parameters $\underline{\theta}$ and \underline{v}^2 , as opposed to the specification (8)–(10) where they are estimated. For instance, estimating the level of shrinkage \underline{v}^2 makes the Bayesian estimate, $\bar{\theta}_i$, more aligned with data as the weights in (6) are determined more by the information in the data. Therefore, the hierarchical modelling allows the data to decide on the level of shrinkage of the parameters towards the prior mean. This leads to improved fit of the model to the data and better forecasting performance.

Bayesian estimation of the hyper-parameters $\underline{\theta}$ and \underline{v}^2 is not directly informed by the data, but rather by the values of parameter θ_i sampled from its posterior distribution. The updated values of $\underline{\theta}$ and \underline{v}^2 sampled from their respective posterior distribution feed in the new values of parameter θ_i in a joint estimation proceeding in multiple iterations.

Bayesian Panel Data Modelling

Bayesian panel data modelling is also facilitated through hierarchical modelling by estimating the prior mean of the parameters as a consequence of setting the country-invariant prior mean $\underline{\theta}$ (with no subscript i), assuming priors for all parameters, θ_i and $\underline{\theta}$, and estimating them. This modelling strategy of using hierarchical prior distributions constitutes the backbone of Bayesian fixed and random effects models (see a recent review by [Rendon 2013](#)) and has been adapted to Bayesian Panel Vector Autoregressions by [Jarociński \(2010\)](#). It has also been used by [Raftery, Zimmer, Frierson, Startz, and Liu \(2017\)](#) and [Gerland, Raftery, Ševčíková, Li, Gu, Spoorenberg, Alkema, Fosdick, Chunn, Lalic et al. \(2014\)](#) in projects on the first probabilistic forecasting of global CO₂ emissions and the world's population for United Nations' Intergovernmental Panel on Climate Change.

Bayesian Estimation

Bayesian estimation for the Hierarchical Panel Vector Autoregression is performed using the Gibbs sampler. The Gibbs sampler is an algorithm to obtain random draws from the posterior distribution of the parameters of the model given the data. The algorithm is explained on an example of the simple model with parameters θ_1 and θ_2 and hyper-parameters $\underline{\theta}$ and \underline{v}^2 considered in the current section. In order to sample from the joint posterior distribution $p(\theta_1, \theta_2, \underline{\theta}, \underline{v}^2 | \mathbf{Y})$ the Gibbs sampler proceeds by sampling from full-conditional posterior distributions of each parameter given data and all the other parameters, denoted by $p(\theta_1 | \theta_2, \underline{\theta}, \underline{v}^2, \mathbf{Y})$, $p(\theta_2 | \theta_1, \underline{\theta}, \underline{v}^2, \mathbf{Y})$, ... These distributions must have an analytical expression,

and should belong to a family of distributions for which sampling random numbers is relatively simple.

To obtain S draws from the posterior distribution, the Gibbs sampler proceeds as follows:

1. Set the initial values of the parameters $\theta_2^{(0)}$, $\underline{\theta}^{(0)}$, and $\underline{v}^{2(0)}$
2. At each iteration s :
 - (a) Sample $\theta_1^{(s)}$ from $p(\theta_1 | \theta_2^{(s-1)}, \underline{\theta}^{(s-1)}, \underline{v}^{2(s-1)}, \mathbf{Y})$
 - (b) Sample $\theta_2^{(s)}$ from $p(\theta_2 | \theta_1^{(s)}, \underline{\theta}^{(s-1)}, \underline{v}^{2(s-1)}, \mathbf{Y})$
 - (c) Sample $\underline{\theta}^{(s)}$ from $p(\underline{\theta} | \theta_1^{(s)}, \theta_2^{(s)}, \underline{v}^{2(s-1)}, \mathbf{Y})$
 - (d) Sample $\underline{v}^{2(s)}$ from $p(\underline{v}^2 | \theta_1^{(s)}, \theta_2^{(s)}, \underline{\theta}^{(s)}, \mathbf{Y})$
3. Repeat step 2. S times.
4. Return draws $\{\theta_1^{(s)}, \theta_2^{(s)}, \underline{\theta}^{(s)}, \underline{v}^{2(s)}\}_{s=1}^S$ as a sample drawn from the posterior distribution $p(\theta_1, \theta_2, \underline{\theta}, \underline{v}^2 | \mathbf{Y})$.

Note that at each of the iterations in point 2. the particular parameter is sampled from its full conditional posterior distribution where the other parameters are replaced by their most up-to-date draw. These values are used to compute the parameters determining the full conditional posterior distribution such as the mean from equations (5) and (6). This numerical integration technique facilitates the formulation of the joint posterior distribution as in point 4.

Gibbs sampler, being an iterative procedure, has been proven to converge to the stationary posterior density with certainty irrespective of the starting values (see [Casella and George 1992](#)). To obtain a sample of draws from the target posterior distribution, first, a *burn-in* run of the Gibbs sampler is performed to obtain the convergence. It is followed by the final run of S iterations started from the last draw of the burn-in run.

The sample of draws from the posterior distribution, $\{\theta_1^{(s)}, \theta_2^{(s)}, \underline{\theta}^{(s)}, \underline{v}^{2(s)}\}_{s=1}^S$, is used to report the posterior estimates, such as the mean, standard deviation, highest posterior density intervals, as well as to compute any interpretable quantities of interest, or to forecast the dependent variable.

Bayesian Missing Observations Treatment

Bayesian Forecasting

Bayesian forecasting is based on the conditional predictive density of the dependent variable y_{it} given the explanatory variable x_{it} and the parameter θ_i which is determined by the model as that in equation (4):

$$p(y_{it} | x_{it}, \theta_i) = \mathcal{N}(\theta_i x_{it}, 1). \quad (11)$$

This density is not different from that used in frequentist forecasting. In Bayesian forecasting one can construct the predictive density of the dependent variable y_{it} given the explanatory variable x_{it} as:

$$p(y_{it} | x_{it}) = \int p(y_{it}, \theta_i | x_{it}) d\theta_i = \int p(y_{it} | x_{it}, \theta_i) p(\theta_i | y_{it}, x_{it}) d\theta_i. \quad (12)$$

In the first equality in (12), the predictive density, $p(y_{it} | x_{it})$, is obtained by integrating out the parameter θ_i from the joint density of the parameter and dependent variable, $p(y_{it}, \theta_i | x_{it})$.

The second equality in (12) decomposes the joint density in the conditional predictive density, $p(y_{it} | x_{it}, \theta_i)$, and the posterior density of the parameter, $p(\theta_i | y_{it}, x_{it})$. Therefore, the predictive density is obtained by integrating out the parameters using their posterior density and, thus, Bayesian forecasting incorporates the uncertainty with respect to the parameter estimation in the forecasting process.

In practice, the predictive density is usually not known in a closed-form solution and appropriate numerical integration techniques are employed. Their objective is to provide a sample drawn from the predictive density in equation (12). These draws are then used to report the point, interval, and density forecasts.

Forecast Performance Evaluation

3. A Family of Hierarchical Panel Vector Autoregression

In what follows, we present a tutorial on each of the models included in the R package **bpvars** with their variations and implementation scripts for R. The first model is fully explained in Section 3.1 and the remaining three models include extensive references to some of its parts. The exposition here should allow the user to gain full knowledge about the models and their use in R.

3.1. Panel Vector Autoregressions with Global Prior

Consider a benchmark model that consists of a set of country-specific Vector Autoregressions (VARs) with the global prior specification. In this model, the labour market outcomes are forecasted with the country-specific VAR parameters. These parameters, however, entertain a panel model feature by sharing the same prior distribution, which we call a global prior distribution. Such formulation of the model can be understood as a country-specific VAR whose parameters follow a global VAR under the prior mean. The main features of using this model include:

- country-specific VAR modelling,
- global prior distribution where the country-specific autoregressive and error term covariance parameter matrices have the prior mean given by these parameters' global counterparts,
- this exchangeable prior supplies the model-specific equations with information from the whole panel of countries through the global parameters as the prior mean,
- a multi-level hierarchical prior structure that grants flexibility, make the model adaptable to various types of data, and reduces the reliance on arbitrary choices,
- Bayesian estimation through an efficient and fast Gibbs sampler.

Below, the country specific model is presented and the global prior is explained.

Country-Specific Vector Autoregressions

Let an N -vector $\mathbf{y}_{c,t} = [gdp_{c,t} \quad ur_{c,t} \quad er_{c,t} \quad pr_{c,t}]'$ collect the dependent variables for country c at time t , where the country indicator takes values $c \in \{1, \dots, C\}$ with a total number of C countries in the sample data, and the time indicator $t \in \{1, \dots, T_c\}$, with the country-specific sample size T_c . These variables follow a multivariate dynamic specification, namely, the Gaussian VAR model (see Sims 1980) given by

$$\mathbf{y}_{c,t} = \mathbf{A}_{c,1}\mathbf{y}_{c,t-1} + \dots + \mathbf{A}_{c,p}\mathbf{y}_{c,t-p} + \mathbf{A}_{c,d}\mathbf{d}_{c,t} + \boldsymbol{\epsilon}_{c,t}, \quad (13)$$

$$\boldsymbol{\epsilon}_{c,t} | \mathbf{y}_{c,t-1}, \dots, \mathbf{y}_{c,t-p} \sim iid \mathcal{N}_N(\mathbf{0}_N, \boldsymbol{\Sigma}_c), \quad (14)$$

where $\mathbf{A}_{c,l}$ are $N \times N$ autoregressive matrices at lag l , $\mathbf{d}_{c,t}$ is the D -vector of deterministic terms, and $\mathbf{A}_{c,d}$ is the $N \times D$ matrix of the corresponding coefficients, $\boldsymbol{\epsilon}_{c,t}$ is the N -vector of error terms that is normally distributed with covariance matrix $\boldsymbol{\Sigma}_c$. Additionally, this dynamic model relies on initial conditions $\mathbf{y}_{c,0}, \dots, \mathbf{y}_{c,-(p-1)}$ that are estimated.

In the model from expressions (13) and (14), all the variables are treated as endogenous through the specification of the joint conditional normal distribution with covariance $\boldsymbol{\Sigma}_c$ and their dynamics and temporal inter-dependencies are captured by the autoregressive parameters $\mathbf{A}_{c,l}$. These features decide on the improved forecasting performance of VAR models compared to their univariate or static alternatives.

Rewrite this model in a matrix notation. Define a $T_c \times N$ matrix $\mathbf{Y}_c = [\mathbf{y}_{c,1} \ \dots \ \mathbf{y}_{c,T_c}]'$, a $T_c \times K$ matrix $\mathbf{X}_c = [\mathbf{x}_{c,1} \ \dots \ \mathbf{x}_{c,T_c}]'$, where $\mathbf{x}_{c,t} = [\mathbf{y}'_{c,t-1} \ \dots \ \mathbf{y}'_{c,t-p} \ \mathbf{d}'_{c,t}]'$, and $K = Np + D$, a $T_c \times N$ matrix $\mathbf{E}_c = [\mathbf{e}_{c,1} \ \dots \ \mathbf{e}_{c,T_c}]'$, and a $K \times N$ matrix $\mathbf{A}_c = [\mathbf{A}_{c,1} \ \dots \ \mathbf{A}_{c,p} \ \mathbf{A}_{c,d}]'$. Then, the model from (13)–(14) can be written in an equivalent form as

$$\mathbf{Y}_c = \mathbf{X}_c \mathbf{A}_c + \mathbf{E}_c, \quad (15)$$

$$\mathbf{E}_c | \mathbf{X}_c \sim \mathcal{MN}_{T_c \times N}(\mathbf{0}_{T_c \times N}, \mathbf{I}_{T_c}, \boldsymbol{\Sigma}_c), \quad (16)$$

where $\mathcal{MN}_{T_c \times N}()$ denotes a matrix-variate normal distribution for a $T_c \times N$ matrix (see Bauwens, Lubrano, and Richard 1999; Woźniak 2016).¹ Note that the assumptions in matrix specification (16) correspond to those from assumption (14). It is further used to introduce the prior structure for the model.

Global Hierarchical Prior Distribution

Bayesian inference and estimation requires the specification of prior distributions of the model's parameters. These distributions are specified based on the feasibility of estimation, interpretability, and to optimise forecasting performance.

The main objectives motivating the prior distributions selection are:

- to ensure flexible prior specification for the country-specific parameters allowing them to vary substantially for different countries,
- to facilitate the estimation of the global parameters, thereby giving the model a panel data model interpretation,
- to grant the prior distribution the interpretability of the Minnesota prior or pooled estimator, which has been proven to improve forecasting performance for macroeconomic aggregates and panel models,
- to allow the data to flexibly determine the level of prior shrinkage,
- to result in efficient Bayesian estimation through Gibbs sampler.

The prior distribution for the country-specific parameters \mathbf{A}_c and $\boldsymbol{\Sigma}_c$ is specified in a hierarchical manner, which allows for the estimation of the prior hyper-parameters addressing the objectives stated above. Its main characteristic is the interpretation of the prior means for country-specific autoregressive parameters $\mathbb{E}[\mathbf{A}_c] = \mathbf{A}$ and error term covariance matrix $\mathbb{E}[\boldsymbol{\Sigma}_c] = \boldsymbol{\Sigma}$ as global parameters, that is, invariant over countries c . These prior expected values imply a VAR model with global parameters is given by:

$$\mathbf{Y}_c = \mathbf{X}_c \mathbf{A} + \mathbf{E}_c, \quad (17)$$

$$\mathbf{E}_c | \mathbf{X}_c \sim \mathcal{MN}_{T_c \times N}(\mathbf{0}_{T_c \times N}, \mathbf{I}_{T_c}, \boldsymbol{\Sigma}). \quad (18)$$

This prior specification is implemented by assuming a convenient matrix-variate normal inverse Wishart distribution (see Karlsson 2013; Woźniak 2016) given by

$$\mathbf{A}_c, \boldsymbol{\Sigma}_c | \mathbf{A}, \mathbf{V}, \boldsymbol{\Sigma}, \nu \sim \mathcal{MNIW}_{K \times N}(\mathbf{A}, \mathbf{V}, (N - \nu - 1)\boldsymbol{\Sigma}, \nu) \quad (19)$$

¹Let operator $\text{vec}()$ stack the columns of \mathbf{E}_c one under another in a $T_c N \times 1$ vector $\text{vec}(\mathbf{E}_c)$. Then the distribution specification in (16) with the mean matrix $\mathbf{0}_{T_c \times N}$, the row-specific covariance parameter $\boldsymbol{\Sigma}_c$, and the column-specific covariance parameter \mathbf{I}_{T_c} , is equivalent to the multivariate normal distribution $\text{vec}(\mathbf{E}_c) \sim \mathcal{N}_{T_c N}(\mathbf{0}_{T_c N \times 1}, \boldsymbol{\Sigma}_c \otimes \mathbf{I}_{T_c})$, where \otimes denotes the Kronecker product of two matrices.

that additionally includes hyper-parameters determining the scale and shape of the prior distribution for \mathbf{A}_c and Σ_c , namely, a $K \times K$ column-specific covariance matrix \mathbf{V} and the shape parameter ν . Additional advantage of this specification is that it leads to a convenient Gibbs sampler for the estimation of the model.

The global parameters are estimated, which is facilitated by Bayesian hierarchical modelling. Therefore, the global autoregressive matrix, \mathbf{A} , follows a matrix-variate normal distribution with the mean $K \times N$ matrix $m\mathbf{M}$, $K \times K$ column-specific covariance \mathbf{V} , and $N \times N$ row-specific covariance $s\mathbf{S}$ denoted by

$$\mathbf{A} \mid \mathbf{V}, m, s \sim \mathcal{MN}_{K \times N}(m\mathbf{M}, \mathbf{V}, s\mathbf{S}), \quad (20)$$

where \mathbf{M} and \mathbf{S} are fixed matrices of appropriate sizes and types while scalar hyper-parameters m and s are further estimated.

The global error term covariance matrix, Σ , follows a Wishart distribution with $N \times N$ scale matrix $s\mathbf{S}_\Sigma$ and shape parameter μ_Σ

$$\Sigma \mid s, \nu \sim \mathcal{W}_N(s\mathbf{S}_\Sigma, \mu_\Sigma), \quad (21)$$

where the matrix \mathbf{S}_Σ and shape parameter μ_Σ are fixed, while the positive scalar s is estimated.

Hierarchical Prior Distribution

In the hierarchical Panel VAR model proposed here, all of the hyper-parameters of the prior in (19) are estimated. This gives the model the advantage of fitting the data closely, while avoiding the necessity of making arbitrary choices regarding the values of these hyper-parameters.

Consequently, a prior distribution is assumed for the column-specific covariance \mathbf{V} that controls the level of shrinkage of the autoregressive parameters \mathbf{A}_c and \mathbf{A} around their respective prior means. It is set to the inverse-Wishart distribution with scale $w\mathbf{W}$ and shape η

$$\mathbf{V} \mid w \sim \mathcal{IW}_N(w\mathbf{W}, \eta), \quad (22)$$

with the $K \times K$ scale matrix \mathbf{W} and the shape parameter η being fixed and the positive scalar w estimated. The shape parameter ν follows an exponential distribution with mean $\underline{\lambda}$ denoted by

$$\nu \sim \exp(\underline{\lambda}). \quad (23)$$

The prior specification is complemented by the average global persistence hyper-parameters m , and scaling factors w and s following the normal, gamma, and inverted gamma 2 prior distributions respectively

$$m \sim \mathcal{N}(\underline{\mu}_m, \sigma_m^2), \quad (24)$$

$$w \sim \mathcal{G}(\underline{s}_w, a_w), \quad (25)$$

$$s \sim \mathcal{IG2}(\underline{s}_s, v_s). \quad (26)$$

To summarise, the joint prior distribution for the parameters of the model is given by

$$p(\mathbf{A}_c, \Sigma_c, \mathbf{A}, \mathbf{V}, \Sigma, \nu, m, w, s) = p(\mathbf{A}_c, \Sigma_c \mid \mathbf{A}, \mathbf{V}, \Sigma, \nu) p(\mathbf{A}, \mathbf{V} \mid m, w, s) p(\Sigma \mid s) \\ \times p(\nu)p(m)p(w)p(s), \quad (27)$$

where the particular distributions are as follows:

$$\mathbf{A}_c, \mathbf{\Sigma}_c | \mathbf{A}, \mathbf{V}, \mathbf{\Sigma}, \nu \sim \mathcal{MNITW}_{K \times N}(\mathbf{A}, \mathbf{V}, (N - \nu - 1)\mathbf{\Sigma}, \nu) \quad (28)$$

$$\mathbf{A}', \mathbf{V} | m, w, s \sim \mathcal{MNITW}_{N \times K}(m\mathbf{M}', s\mathbf{S}, w\mathbf{W}, \eta) \quad (29)$$

$$\mathbf{\Sigma} | s \sim \mathcal{W}_N(s\mathbf{S}, \underline{\mu}_{\mathbf{\Sigma}}) \quad (30)$$

$$\nu \sim \exp(\underline{\lambda}) \quad (31)$$

$$m \sim \mathcal{N}(\underline{\mu}_m, \underline{\sigma}_m^2) \quad (32)$$

$$w \sim \mathcal{G}(\underline{s}_w, \underline{a}_w) \quad (33)$$

$$s \sim \mathcal{IG2}(\underline{s}_s, \underline{\nu}_s). \quad (34)$$

Fixed Prior Hyper-Parameters

These prior distributions depend on fixed hyper-parameters which in our notation are underscored. In what follows, we provide a justification for their default values, which are then utilized in the **bpvars** package.

The package offers two alternative values of the matrix contributing to the global autoregressive parameters prior mean, \mathbf{M} . The first choice is motivated by the interpretability of the Minnesota Prior proposed by Doan *et al.* (1984), in which case this matrix is set to $\begin{bmatrix} \mathbf{I}_N & \mathbf{0}_{N \times K-N} \end{bmatrix}'$. It implies that the prior mean of the own lag for each of the variables is estimated by the value of another hyper-parameter m pre-multiplying this matrix in (29). The alternative choice draws on the idea by Zellner and Hong (1989) where this matrix is set to the pooled estimator equal to $(\sum_{c=1}^C \mathbf{X}_c' \mathbf{X}_c)^{-1} (\sum_{c=1}^C \mathbf{X}_c' \mathbf{Y}_c)$. The prior mean specification is complemented by a normal prior distribution for the hyper-parameter m with mean $\underline{\mu}_m = 1$ and the variance $\underline{\sigma}_m^2 = 1$. This distribution centres the prior around the matrix \mathbf{M} , that is around random walk process in the Minnesota prior, and the pooled estimator otherwise.

The \mathbf{V} parameter prior scale matrix is set to $\mathbf{W} = \text{diag}(\mathbf{I}_N \otimes \mathbf{p}^{-2} \quad 100)$, where \mathbf{p} is a p -vector of values from 1 to p , implements the feature of the Minnesota prior for global parameters where the shrinkage towards the prior mean becomes exponentially stronger for autoregressive matrices \mathbf{A}_i with increasing lag order $i = 1, \dots, p$. The estimated hyper-parameter pre-multiplying this matrix, namely w , features a gamma prior with the scale $\underline{s}_w = 1$ and shape $\underline{a}_w = 1$. These values make the prior distribution little informative and lets the data decide on the underlying estimate. The row-specific covariance matrix of \mathbf{A} is a product of the identity matrix $\mathbf{S} = \mathbf{I}_N$ and the estimated hyper-parameter s , featuring the inverted gamma 2 prior with the scale and shape set to $\underline{s}_s = 1$ and $\underline{\nu}_s = 3$ respectively.

Similar choices are made for the prior scale of the global covariance matrix $\mathbf{\Sigma}$ being the identity matrix $\mathbf{S}_{\mathbf{\Sigma}} = \mathbf{I}_N$ pre-multiplied by the estimated s . The shape parameter of this Wishart distribution is set to $\underline{\mu}_{\mathbf{\Sigma}} = N + 1$, which ensures finite prior variance of $\mathbf{\Sigma}$. Similarly, the value of the shape parameter for the prior distribution in (29) is set to $\underline{\eta} = N + 1$.

Finally, the exponential prior for the degrees of freedom parameter ν is set to $\underline{\lambda} = 72$, which assigns 50% of the prior probability to the degrees of freedom parameter being less than 50. This choice makes the prior span the part of the parameter space implying Student-t like distribution for low values of ν , as well as close approximations of the normal distribution for values of $\nu > 30$.

Estimation and Forecasting in R

The code below introduces the basic steps to specify and estimate the model. Load the package using `library(bpvars)`. Specify the model using the function `specify_bvarPANEL$new()` and provide the necessary arguments, including data set to the list of matrices provided in the package in object `ilo_dynamic_panel`, specification of deterministic terms `exogenous` set to the list of matrices provided in the package `ilo_exogenous_variables`, and the qualifier `stationary` where this type set to `FALSE` means the corresponding variable is unit-root non-stationary, which determines the Minnesota prior mean in the matrix \mathbf{M} . Estimate the model using method `estimate()` in two rounds: the first to achieve convergence with the specification of the model in object `spec` running 5000 iterations of the Gibbs sampler, and the second for the final estimation accepting estimation output from the first round in object `burn` that extracts its last draw and continues the MCMC for 10000 iterations. The estimation output is stored in object `post` that is used to compute and display estimation summary using method `summary()`.

```
R> library(bpvars)                                # load the package
R>
R> spec      = specify_bvarPANEL$new(              # specify the model
+   data      = ilo_dynamic_panel,                 # include data
+   stationary = c(FALSE, FALSE, FALSE, FALSE),    # declare variables as non-stationary
+   exogenous  = ilo_exogenous_variables           # include deterministic terms
+ )
R> burn      = estimate(spec, 5000)                 # initial estimation for convergence
R> post      = estimate(burn, 10000)                # run the final estimation
R>
R> post_summ = summary(post)                       # compute the estimation summary
R> post_summ$POL                                  # print estimation summary for Poland
```

To forecast the variables, provide the final estimation output object `post` to method `forecast()` along with the forecast horizon set to 5 and the future values of the deterministic terms in object `ilo_exogenous_forecasts` that is provided in the package. Then plot the forecasts for individual countries using method `plot()` or analyse the point forecasts using method `summary()`.

```
R> fore      = forecast(                           # run the forecast
+   post,                                          # provide estimation output
+   horizon = 5,                                  # specify forecast horizon
+   exogenous = ilo_exogenous_forecasts          # include deterministic terms
+ )
R> plot(fore, "POL")                              # plot the forecast for Poland
R> fore_summ = summary(fore, "POL")               # compute forecast summary for Poland
R> fore_summ                                       # print the point forecast
```

Finally, save the estimation and forecast outputs in a file for future use.

```
R> save(                                           # save the outputs
+   post, fore,                                   # estimation and forecast outputs
+   file = "bvarPANEL.rda"                       # file name
+ )
```

Variation: Prior Mean of the Global Prior

As mentioned above, the package offers two alternative settings for the prior mean of the global autoregressive parameters, that is, matrix \mathbf{M} . The first choice is motivated by the interpretability

of the Minnesota Prior proposed by [Doan et al. \(1984\)](#), in which case this matrix is set to

$$\underline{\mathbf{M}} = \begin{bmatrix} \mathbf{I}_N \\ \mathbf{0}_{K-N \times N} \end{bmatrix}. \quad (35)$$

It implies that the prior mean of the own lag for each of the variables is estimated by the value of another hyper-parameter m pre-multiplying this matrix in (29). The alternative choice draws on the idea by [Zellner and Hong \(1989\)](#) where this matrix is set to the pooled estimator equal to

$$\underline{\mathbf{M}} = \left(\sum_{c=1}^C \mathbf{x}'_c \mathbf{x}_c \right)^{-1} \left(\sum_{c=1}^C \mathbf{x}'_c \mathbf{y}_c \right). \quad (36)$$

Prior Mean of the Global Prior in R

The Minnesota prior is the default setup in the package. To check the prior mean specification, use:

```
R> spec$prior$M # check the hyperparameter
```

To set this parameter to the pooled estimator, specify the model and modify it:

```
R> spec = specify_bvarPANEL$new( # specify the model
+ data = ilo_dynamic_panel, # include data
+ stationary = c(FALSE, FALSE, FALSE, FALSE), # declare variables as non-stationary
+ exogenous = ilo_exogenous_variables # include deterministic terms
+ )
R> spec$set_global2pooled() # change the prior mean
```

Extension: A Model by Jarociński

The benchmark specification presented in Section 3 has an alternative prior specification proposed by [Jarociński \(2010\)](#). This model facilitates the estimation of the country-specific parameters and the global autoregressive matrix assuming a minimal prior structure. It features equations (15) and (16) with the prior distribution for the country-specific parameters given by

$$\mathbf{A}_c | \Sigma_c, \mathbf{A}, s \sim \mathcal{MN}_{K \times N}(\mathbf{A}, s \underline{\mathbf{W}}, \Sigma_c) \quad (37)$$

$$\Sigma_c \propto \det(\Sigma_c)^{-\frac{N+1}{2}} \quad (38)$$

In this model, the global autoregressive parameters follow an improper prior distribution:

$$p(\mathbf{A}) \propto 1, \quad (39)$$

and s follows the inverted gamma 2 prior distribution as in (26) with $\underline{s}_s = \underline{v}_s = 0.001$.

A Model by Jarociński in R

To set the model by [Jarociński \(2010\)](#) specify the model and modify it:

```
R> spec = specify_bvarPANEL$new( # specify the model
+ data = ilo_dynamic_panel, # include data
+ stationary = c(FALSE, FALSE, FALSE, FALSE), # declare variables as non-stationary
+ exogenous = ilo_exogenous_variables # include deterministic terms
+ )
R> spec$set_to_Jarocinski() # change the specification
```

Estimation and forecasting proceed the same way as explained above.

A Model for Data with Missing Observations in R

Each of the models here can be estimated with data that have missing observations. The package implements a data augmentation algorithm that treats the missing observations as additional parameters to be estimated. This estimation is performed using the model-coherent sampling densities. The package provides data with missing observations in object `ilo_dynamic_panel_missing`. For instance, the series for Afghanistan are characterised by relatively short sample with many missing observations in the sample period, where the missing observations take value NA:

```
R> ilo_dynamic_panel_missing$AFG
```

Time Series:

Start = 2014

End = 2024

Frequency = 1

	gdp	UR	EPR	LFPR
2014	23.66034	7.915	43.3681	47.0957
2015	23.67474	NA	NA	NA
2016	23.69710	NA	NA	NA
2017	23.72322	11.184	42.0144	47.3050
2018	23.73504	NA	NA	NA
2019	23.77341	NA	NA	NA
2020	23.74962	11.710	36.7104	41.5793
2021	23.59248	NA	NA	NA
2022	23.52805	14.100	32.3328	37.6400
2023	23.55046	NA	NA	NA
2024	23.57259	NA	NA	NA

To estimate the model with such data, specify the model normally, using the `specify_bvarPANEL$new()` function, and provide for the data argument the list of matrices with missing observations. The package will detect the missing observations and apply the estimation algorithms accordingly.

```
R> spec      = specify_bvarPANEL$new(           # specify the model
+   data      = ilo_dynamic_panel_missing,      # include data with missing observations
+   stationary = c(FALSE, FALSE, FALSE, FALSE), # declare variables as non-stationary
+   exogenous  = ilo_exogenous_variables        # include deterministic terms
+ )
```

The estimation and forecasting proceed the same way as explained above.

```
R> burn      = estimate(spec, 5000)              # run the burn-in
R> post      = estimate(burn, 10000)            # estimate the model
R> fore      = forecast(                        # run the forecast
+   post,                                       # provide estimation output
+   horizon = 5,                              # specify forecast horizon
+   exogenous = ilo_exogenous_forecasts       # include deterministic terms
+ )
```

3.2. Panel Vector Autoregressions with Country Grouping

As a variation on the country-specific VAR model, we consider a model with country grouping. Consider a set of C countries grouped into G groups, where each group $g \in \{1, \dots, G\}$ contains C_g countries. Each group g has its own parameters, \mathbf{A}_g and Σ_g , defining the VAR model with country grouping for country c from group g given by

$$\mathbf{Y}_c = \mathbf{X}_c \mathbf{A}_g + \mathbf{E}_c, \quad (40)$$

$$\mathbf{E}_c | \mathbf{X}_c \sim \mathcal{MN}_{T_c \times N}(\mathbf{0}_{T_c \times N}, \mathbf{I}_{T_c}, \Sigma_g). \quad (41)$$

In other words, this model is specified by imposing restrictions on the country-specific parameters such that $\mathbf{A}_c = \mathbf{A}_g$ and $\Sigma_c = \Sigma_g$. The group allocations can be fixed and determined, for example, by geographical location or economic development, or they can be estimated from the data given a pre-specified number of groups G .

The group-specific parameters follow the hierarchical prior distribution given by

$$\mathbf{A}_g, \Sigma_g | \mathbf{A}, \mathbf{V}, \Sigma, \nu \sim \mathcal{MN} \mathcal{IW}_{K \times N}(\mathbf{A}, \mathbf{V}, (N - \nu - 1)\Sigma, \nu), \quad (42)$$

that is complemented by the same hierarchical structure as presented in parts *Hierarchical Prior Distribution* and *Fixed Prior Hyper-Parameters* of Section 3.1. The package implements two variations in the prior specification with the matrix \mathbf{M} specified either as in the Minnesota Prior proposed by Doan *et al.* (1984) or as the pooled estimator by Zellner and Hong (1989).

The main objectives motivating such a model are:

- to allow the possibility of country-specific variables being predicted using group-specific parameters that better reflect regional or group trends,
- to facilitate more precise model estimation benefiting from the larger number of observations used to estimate the group-specific parameters,
- to facilitate the estimation of the global parameters, thereby giving the model a panel data model interpretation,
- to grant the prior distribution the interpretability of the Minnesota prior or pooled estimator, which has been proven to improve forecasting performance for macroeconomic aggregates and panel models,
- to allow the data to flexibly determine the level of prior shrinkage,
- to result in efficient Bayesian estimation through Gibbs sampler.

Variation: A Model with Fixed Country Grouping

This model can be estimated using fixed and pre-specified by the user country groupings. Such a grouping determines which countries, indexed by c , belong to which group, indexed by g .

A Model with Fixed Country Grouping in R

In order to specify the model with fixed country grouping use the `specify_bvarGroupPANEL$new()` function providing the `group_allocation` argument setting it to a vector of length C with group indices for each country. The package provides a grouping based on geographical regions in object `country_grouping_region` used below. Other such groupings of countries are included in vectors `country_grouping_subregionbroad`, `country_grouping_subregiondetailed`, and

country_grouping_incomegroup. Their detailed specification can be revised by accessing package documentation ?country_grouping_incomegroup. The estimation and forecasting proceed the same way as explained above.

```
R> spec      = specify_bvarGroupPANEL$new(          # specify the model
+   data      = ilo_dynamic_panel,                  # include data
+   group_allocation = country_grouping_region,      # specify country grouping
+   stationary = c(FALSE, FALSE, FALSE, FALSE)      # declare variables as non-stationary
+ )
```

Variation: A Model with Estimated Country Grouping

An alternative option is to specify the number of groups G and allow the package to estimate the country groupings from the data. At each iteration of the Gibbs sampler and for each of the countries, the group allocation is sampled from the multinomial distribution with probabilities proportional to the posterior kernel ordinate for each of the groups.

A Model with Fixed Estimated Grouping in R

In order to specify the model with fixed country grouping use the `specify_bvarGroupPANEL$new()` function providing a positive integer to the G argument setting it to the desired number of groups. The function requires the specification of either the argument G or `group_allocation`. The estimation and forecasting proceed the same way as explained above.

```
R> spec      = specify_bvarGroupPANEL$new(          # specify the model
+   data      = ilo_dynamic_panel,                  # include data
+   G = 2,                                           # specify number of groups
+   stationary = c(FALSE, FALSE, FALSE, FALSE)      # declare variables as non-stationary
+ )
```

Prior Mean of the Global Prior in R

The Minnesota prior is the default setup in the package. To check the prior mean specification, use:

```
R> spec$prior$M                                     # check the hyperparameter
```

To set this parameter to the pooled estimator, specify the model and modify it:

```
R> spec      = specify_bvarGroupPANEL$new(          # specify the model
+   data      = ilo_dynamic_panel,                  # include data
+   group_allocation = country_grouping_region,      # specify country grouping
+   stationary = c(FALSE, FALSE, FALSE, FALSE)      # declare variables as non-stationary
+ )
R> spec$set_global2pooled()                          # change the prior mean
```

A Model for Data with Missing Observations in R

To estimate the model with missing observations, specify the model normally, using the `specify_bvarGroupPANEL$new()` function, and provide for the data argument the list of matrices with missing observations. The package will detect the missing observations and apply the estimation algorithms accordingly.

[illegible]

3.3. Panel Vector Autoregressions with Global Prior Grouping

Another variation is a model featuring equations (15) and (16) in which the country-specific parameters have group-specific global parameters. This model specification is implemented by setting the prior expectations to $\mathbb{E}[\mathbf{A}_c] = \mathbf{A}_g$ and $\mathbb{E}[\mathbf{\Sigma}_c] = \mathbf{\Sigma}_g$ and to let the country groupings to be fixed and specified by the user or estimated for a fixed group number G . Consequently, the country-specific parameters follow the hierarchical prior distribution given by

$$\mathbf{A}_c, \mathbf{\Sigma}_c | \mathbf{A}_g, \mathbf{V}, \mathbf{\Sigma}_g, \nu \sim \mathcal{MNITW}_{K \times N}(\mathbf{A}_g, \mathbf{V}, (N - \nu - 1)\mathbf{\Sigma}_g, \nu) \quad (43)$$

with the priors for the group-specific global parameters given by:

$$\mathbf{A}_g | \mathbf{V}, m, s \sim \mathcal{MN}_{K \times N}(m\mathbf{\underline{M}}, \mathbf{V}, s\mathbf{\underline{S}}), \quad (44)$$

$$\mathbf{\Sigma}_g | s, \nu \sim \mathcal{W}_N(s\mathbf{\underline{S}}, \mu_{\underline{\Sigma}}). \quad (45)$$

Similarly, to other specifications, the remaining prior hierarchy is as described in parts *Hierarchical Prior Distribution* and *Fixed Prior Hyper-Parameters* of Section 3.1 with the same choices for the values of matrix $\mathbf{\underline{M}}$.

The main objectives motivating such a model are:

- to support the estimation of country-specific parameters using group-specific global prior parameters that better reflect regional or group trends,
- to facilitate the estimation of the group-specific global parameters, thereby giving the model a panel data model interpretation,
- to grant the prior distribution the interpretability of the Minnesota prior or pooled estimator, which has been proven to improve forecasting performance for macroeconomic aggregates and panel models,
- to allow the data to flexibly determine the level of prior shrinkage,
- to result in efficient Bayesian estimation through Gibbs sampler.

Variation: A Model with Fixed Country Grouping

This model can be estimated using fixed and pre-specified by the user country groupings. Such a grouping determines which countries, indexed by c , belong to which group, indexed by g .

A Model with Fixed Country Grouping in R

In order to specify the model with fixed country grouping use the `specify_bvarGroupPriorPANEL$new()` function providing the `group_allocation` argument setting it to a vector of length C with group indices for each country. The package provides a grouping based on geographical regions in object `country_grouping_region` used below. Its detailed specification can be revised by accessing package documentation `?country_grouping_region`. The estimation and forecasting proceed the same way as explained above.

```
R> spec      = specify_bvarGroupPriorPANEL$new(      # specify the model
+   data      = ilo_dynamic_panel,                  # include data
+   group_allocation = country_grouping_region,      # specify country grouping
+   stationary = c(FALSE, FALSE, FALSE, FALSE)      # declare variables as non-stationary
+ )
```

An alternative option is to specify the number of groups G and allow the package to estimate the country groupings from the data. At each iteration of the Gibbs sampler and for each of the countries, the group allocation is sampled from the multinomial distribution with probabilities proportional to the posterior kernel ordinate for each of the groups.

In order to specify the model with fixed country grouping use the `specify_bvarGroupPriorPANEL$new()` function providing a positive integer to the `G` argument setting it to the desired number of groups. The function requires the specification of either the argument `G` or `group_allocation`. The estimation and forecasting proceed the same way as explained above.

The Minnesota prior is the default setup in the package. To check the prior mean specification, use:

To estimate the model with missing observations, specify the model normally, using the `specify_bvarGroupPriorPANEL$new()` function, and provide for the `data` argument the list of matrices with missing observations. The package will detect the missing observations and apply the estimation algorithms accordingly.

[illegible]

3.4. Vector Autoregressions for Dynamic Panel Data

Finally, the package allows the estimation of VAR models for individual countries. This facility is provided for comparisons with the Hierarchical Panel models. In this model, the country specific parameters are estimated independently for each country c by setting the model equations as in (15) and (16), and assuming the following prior distribution:

$$\mathbf{A}_c, \mathbf{\Sigma}_c | m, s, w, v \sim \mathcal{MNITW}_{K \times N}(\underline{m}\mathbf{M}, w\underline{\mathbf{W}}, s\underline{\mathbf{S}}, v), \quad (46)$$

where the hyper-parameters m, s, w , and v may be pre-specified, or estimated. In the latter case, the hyper-parameters m and v follow the distributions specified in (32) and (31), respectively, while those for s and w are:

$$w \sim \mathcal{IG2}(\underline{s}_w, \underline{av}_w) \quad (47)$$

$$s \sim \mathcal{G}(\underline{s}_s, \underline{a}_s). \quad (48)$$

The main objectives motivating such a model are:

- to facilitate easy estimation of a country-specific VAR models fitted to dynamic panel data,
- to allow coherent forecast performance evaluation that can document benefits of using panel data models,
- to grant the prior distribution the interpretability of the Minnesota prior or pooled estimator, which has been proven to improve forecasting performance for macroeconomic aggregates and panel models,
- to allow the data to flexibly determine the level of prior shrinkage,
- to result in efficient Bayesian estimation through Gibbs sampler.

Estimation and Forecasting in R

Specify the model using the function `specify_bvars$new()` and provide the necessary arguments for customisation. The functions `estimate()` and `forecast()` are then used to estimate the model and produce forecasts.

```
R> spec      = specify_bvars$new(                # specify the model
+   data      = ilo_dynamic_panel,              # include data
+   stationary = c(FALSE, FALSE, FALSE, FALSE), # declare variables as non-stationary
+   exogenous  = ilo_exogenous_variables        # include deterministic terms
+ )
R> burn      = estimate(spec, 5000)              # run the burn-in
R> post      = estimate(burn, 10000)            # estimate the model
R> fore      = forecast(                        # run the forecast
+   post,                                       # provide estimation output
+   horizon = 5,                              # specify forecast horizon
+   exogenous = ilo_exogenous_forecasts      # include deterministic terms
+ )
```

Variation: A Model with Difuse Prior

An alternative prior specification for these country-specific models complements the model equations as in (15) and (16) by the diffuse prior set as:

$$p(\mathbf{A}_c, \Sigma_c) \propto \det(\Sigma_c)^{-\frac{N+1}{2}}, \quad (49)$$

ensuring that the posterior mean estimator for the country-specific parameters is equal to the corresponding maximum likelihood estimator (see [Karlsson 2013](#)).

A Model with Difuse Prior in R

In order to specify this model use the function `specify_bvars$new()` and modify it.

```
R> spec      = specify_bvars$new(                # specify the model
+   data      = ilo_dynamic_panel,              # include data
+   stationary = c(FALSE, FALSE, FALSE, FALSE), # declare variables as non-stationary
+   exogenous  = ilo_exogenous_variables        # include deterministic terms
+ )
R> spec$set_prior2objective()                    # change the prior to diffuse
```

The functions `estimate()` and `forecast()` are then used to estimate the model and produce forecasts.

Variation: A Model with the Prior Mean Set the Pooled Estimator

One more modification of the model sets the prior mean of autoregressive parameters to the pooled estimator by [Zellner and Hong \(1989\)](#). This change is implemented by setting the matrix \mathbf{M} to the pooled estimator equal to $(\sum_{c=1}^C \mathbf{X}'_c \mathbf{X}_c)^{-1} (\sum_{c=1}^C \mathbf{X}'_c \mathbf{Y}_c)$.

A Model with the Prior Mean Set the Pooled Estimator in R

In order to specify this model use the function `specify_bvars$new()` and modify it.

```
R> spec      = specify_bvars$new(                # specify the model
+   data      = ilo_dynamic_panel,              # include data
+   stationary = c(FALSE, FALSE, FALSE, FALSE), # declare variables as non-stationary
+   exogenous  = ilo_exogenous_variables        # include deterministic terms
+ )
R> spec$set_global2pooled()                     # change the prior to diffuse
```

A Model for Data with Missing Observations in R

To estimate the model with missing observations, specify the model normally, using the `specify_bvars$new()` function, and provide for the data argument the list of matrices with missing observations. The package will detect the missing observations and apply the estimation algorithms accordingly.

```
R> spec      = specify_bvars$new(                # specify the model
+   data      = ilo_dynamic_panel_missing,      # include data with missing observations
+   stationary = c(FALSE, FALSE, FALSE, FALSE)
+ )
```



```
+ horizons = 1,                                # set forecast horizon
+ training_sample = 25                          # set the initial training sample
+ )
```

Finally, the function `forecast_poos_recursively()` executes the pseudo-out-of-sample forecasting exercise, while the function `compute_forecast_performance()` computes the forecast performance measures including the point forecast performance measures, namely the Root-Squared-Mean Forecast Error (RMSFE) and Mean-Absolute Forecast Error (MAFE), and density forecast performance measure, namely the predictive log-score (PLS). The results are saved in object `fpm1`.

```
R> fore = forecast_poos_recursively(spec, poos)    # execute the forecasting exercise
R> fpm1 = compute_forecast_performance(fore)      # compute performance measures
R> save(                                          # save the results
+   fpm1,
+   file = "poos_bvarPANEL1.rda"
+ )
```

Above the forecasting performance is measured for a Bayesian Panel VAR model with Global Prior featuring the Minnesota prior. The following script implements the same forecasting exercise for the same model but with the prior mean set to the pooled estimator by [Zellner and Hong \(1989\)](#). The difference in coding is only in the model specification step, where the function `set_global2pooled()` is used to modify the prior mean.

```
R> spec = specify_bvarPANEL$new(
+   data = ilo_dynamic_panel,
+   stationary = rep(FALSE, 4)
+ )
R> spec$set_global2pooled()
R> poos = specify_poosf_exercise$new(
+   spec,
+   S = S,
+   S_burn = S_burn,
+   horizons = 1,
+   training_sample = 25
+ )
R> fore = forecast_poos_recursively(spec, poos)
R> fpm2 = compute_forecast_performance(fore)
R> save(
+   fpm2,
+   file = "poos_bvarPANEL2.rda"
+ )
```

Computations for Panel VARs with Country Grouping

Next, we perform the pseudo-out-of-sample forecasting exercise for the Panel VAR model with Country Grouping. The procedure is the same as above, with the only difference in the model specification step. Below, we provide two such specifications. The first one with fixed country grouping where the group allocations are determined by geographical location as in `country_grouping_region`.

```

R> spec      = specify_bvarGroupPANEL$new(
+   data      = ilo_dynamic_panel,
+   group_allocation = country_grouping_region,
+   stationary = rep(FALSE, 4)
+ )
R> poos = specify_poosf_exercise$new(
+   spec,
+   S = S,
+   S_burn = S_burn,
+   horizons = 1,
+   training_sample = 25
+ )
R> fore = forecast_poos_recursively(spec, poos)
R> fpm3 = compute_forecast_performance(fore)
R> save(
+   fpm3,
+   file = "poos_bvarGroupPANEL1.rda"
+ )

```

And the second model specification where the model is assumed to have two groups, $G=2$, and the group allocations are estimated from the data. The results of estimation and recursive forecasting and forecasting performance evaluation are consistently saved on a hard drive.

```

R> spec      = specify_bvarGroupPANEL$new(
+   data      = ilo_dynamic_panel,
+   G         = 2,
+   stationary = rep(FALSE, 4)
+ )
R> poos = specify_poosf_exercise$new(
+   spec,
+   S = S,
+   S_burn = S_burn,
+   horizons = 1,
+   training_sample = 25
+ )
R> fore = forecast_poos_recursively(spec, poos)
R> fpm4 = compute_forecast_performance(fore)
R> save(
+   fpm4,
+   file = "poos_bvarGroupPANEL2.rda"
+ )

```

Computations for Panel VARs with Global Prior Grouping

Another group of models is the Panel VAR model with Global Prior Grouping. The procedure is the same as above, with the only difference in the model specification step. Below, we provide two such specifications. The first one with fixed country grouping where the group allocations are determined by geographical location as in `country_grouping_region`.

```

R> spec      = specify_bvarGroupPriorPANEL$new(
+   data      = ilo_dynamic_panel,

```

```

+   group_allocation = country_grouping_region,
+   stationary = rep(FALSE, 4)
+ )
R> poos = specify_poosf_exercise$new(
+   spec,
+   S = S,
+   S_burn = S_burn,
+   horizons = 1,
+   training_sample = 25
+ )
R> fore = forecast_poos_recursively(spec, poos)
R> fpm5 = compute_forecast_performance(fore)
R> save(
+   fpm5,
+   file = "poos_bvarGroupPriorPANEL1.rda"
+ )

```

And the second model specification where the model is assumed to have two groups, $G=2$, and the group allocations are estimated from the data.

```

R> spec = specify_bvarGroupPriorPANEL$new(
+   data = ilo_dynamic_panel,
+   G = 2,
+   stationary = rep(FALSE, 4)
+ )
R> poos = specify_poosf_exercise$new(
+   spec,
+   S = S,
+   S_burn = S_burn,
+   horizons = 1,
+   training_sample = 25
+ )
R> fore = forecast_poos_recursively(spec, poos)
R> fpm6 = compute_forecast_performance(fore)
R> save(
+   fpm6,
+   file = "poos_bvarGroupPriorPANEL2.rda"
+ )

```

Computations for VAR models for Dynamic Panel Data

Finally, we execute the recursive pseudo-out-of-sample forecasting exercise for the VAR model for dynamic panel data. Here, individual VAR models are estimated for all countries. Of course, we keep the setup of the pseudo-out-of-sample forecasting exercise coherent and only adjust the model specification step.

```

R> spec = specify_bvars$new(
+   data = ilo_dynamic_panel,
+   stationary = rep(FALSE, 4)
+ )
R> poos = specify_poosf_exercise$new(

```

```

+ spec,
+ S = S,
+ S_burn = S_burn,
+ horizons = 1,
+ training_sample = 25
+ )
R> fore = forecast_poos_recursively(spec, poos)
R> fpm7 = compute_forecast_performance(fore)
R> save(
+ fpm7,
+ file = "poos_bvars1.rda"
+ )

```

4.2. Density Forecast Performance Comparison

We collect all of the results and compare the density forecast performance of the models. This is done according to the standards in the macroeconomic forecasting literature. Therefore, we report the differences in the overall predictive log-score computed jointly for all countries of each of the model relative to the benchmark model. We set the benchmark model to be the Panel VAR with Global Prior featuring the Minnesota prior. Therefore, in the computations below the predictive log-scores of the Panel VAR with Global Prior, saved in `fpm1` object, are subtracted from the predictive log-scores of all other models. The higher predictive log-scores the better forecasting performance of a particular model. Hence, positive values of the differences below indicate better density forecast performance relative to the benchmark model, while negative values indicate worse density forecast performance.

```

R> pls_global = cbind(
+ fpm2$PLS$Global - fpm1$PLS$Global,
+ fpm3$PLS$Global - fpm1$PLS$Global,
+ fpm4$PLS$Global - fpm1$PLS$Global,
+ fpm5$PLS$Global - fpm1$PLS$Global,
+ fpm6$PLS$Global - fpm1$PLS$Global,
+ fpm7$PLS$Global - fpm1$PLS$Global
+ )
R> column_names = c(
+ "GP (pool)",
+ "CG (fix)",
+ "CG (est)",
+ "GPG (fix)",
+ "GPG (est)",
+ "VAR"
+ )
R> colnames(pls_global) = column_names
R> round(pls_global, 3)

```

	GP (pool)	CG (fix)	CG (est)	GPG (fix)	GPG (est)	VAR
gdp	0.000	-0.538	-0.945	-0.017	-0.003	-3.332
UR	0.000	-1.132	-1.719	-0.015	-0.002	-1.459
EPR	0.000	-1.752	-1.719	-0.029	0.002	-0.997
LFPR	0.001	-2.134	-2.136	0.002	-0.003	-1.313
joint	0.011	-6.283	-7.420	0.512	-0.039	-11.923

Focusing the analysis on the overall comparisons of the performance measured for all countries and all variables, we look at the last row of the table reported above. Note that this joint measure is not a simple sum of the quantities for individual variables. The reason for that is that the joint predictive log-score is computed on the basis of the joint density forecast, while the individual variable predictive log-scores are computed on the basis of the marginal density forecasts. Such aggregation was possible if the model assumed conditional independence of the variables in the model, which is not a recommended practice in macroeconomic modelling. The last row of the table above includes only two positive numbers for the Panel VAR model with Global Prior and pooled estimator prior mean and the model with the Global Prior Grouping with fixed allocations. The other values are negative. Please, note that density forecasts are compared here on the log scale. Therefore, it is difficult to interpret whether the differences are substantial. Some help is offered here by [Kass and Raftery \(1995\)](#) and based on a plausible interpretation of the reported quantities as the Bayes Factor. As a rule of thumb, differences greater than 3 in absolute terms are considered strong evidence in favour of the model with the higher predictive log-score. Therefore, the models with Country Grouping and especially the VAR models can be conclusively assessed as having substantially worse density forecast performance relative to the benchmark model.

We move on to the analysis of the forecasting performance for a particular country, namely Poland, as an example of the package capacity. The results can be easily computed from the already saved outputs.

```
R> pls_pol = cbind(
+   fpm2$PLS$POL - fpm1$PLS$POL,
+   fpm3$PLS$POL - fpm1$PLS$POL,
+   fpm4$PLS$POL - fpm1$PLS$POL,
+   fpm5$PLS$POL - fpm1$PLS$POL,
+   fpm6$PLS$POL - fpm1$PLS$POL,
+   fpm7$PLS$POL - fpm1$PLS$POL
+ )
R> colnames(pls_pol) = column_names
R> round(pls_pol, 3)
```

	GP (pool)	CG (fix)	CG (est)	GPG (fix)	GPG (est)	VAR
gdp	-0.001	-0.147	-0.679	-0.011	0.011	-3.106
UR	0.002	-0.001	-0.143	0.009	0.000	-0.205
EPR	0.001	-1.030	-1.120	0.002	0.003	-0.464
LFPR	-0.004	-1.750	-1.948	-0.034	0.001	-1.118
joint	-0.004	-2.546	-3.068	-0.041	0.030	-7.878

Similar conclusions apply here. The model with Global Prior Grouping and fixed group allocations performs best, next to the both models with Global Priors. Also, the models with Country Grouping and VAR models perform substantially worse in terms of density forecast performance relative to the benchmark model.

4.3. Point Forecast Performance Comparison

To assess the point forecast performance of the models we report the overall root-squared-mean forecast errors computed for all countries. According to the standards in the economic forecasting literature we report the ratio of the root-squared-mean forecast errors of each of the models relative to the benchmark model. The benchmark model is again the Panel VAR with Global Prior featuring the Minnesota prior. Therefore, values below 1

indicate better point forecast performance relative to the benchmark model, while values above 1 indicate worse point forecast performance.

```
R> rmsfe_global = cbind(
+   fpm2$RMSFE$Global / fpm1$RMSFE$Global,
+   fpm3$RMSFE$Global / fpm1$RMSFE$Global,
+   fpm4$RMSFE$Global / fpm1$RMSFE$Global,
+   fpm5$RMSFE$Global / fpm1$RMSFE$Global,
+   fpm6$RMSFE$Global / fpm1$RMSFE$Global,
+   fpm7$RMSFE$Global / fpm1$RMSFE$Global
+ )
R> colnames(rmsfe_global) = column_names
R> round(rmsfe_global, 3)
```

	GP (pool)	CG (fix)	CG (est)	GPG (fix)	GPG (est)	VAR
gdp	1.002	1.011	1.655	1.021	1.002	3.123
UR	1.004	0.788	1.982	0.998	1.006	0.880
EPR	1.004	3.795	2.492	0.991	1.009	0.796
LFPR	1.003	4.711	3.221	0.992	1.009	0.791
joint	1.003	3.793	2.691	0.993	1.009	0.821

Focus on the overall results aggregated over all variables and all countries, reported in the last row of the table above, denoted by joint. The model with Global Prior Grouping with the fixed group allocations has 0.7% more precise point forecast than the benchmark model. Two other models have the forecast performance similar to the benchmark model, namely the Global Prior model with pooled estimator for the prior mean and the model with Global Prior Grouping and estimated group allocations. Both of the models with Country Grouping have much worse point forecast performance relative to the benchmark model. However, the VAR models perform surprisingly well according to this exercise and the precision of the point forecast is improved here by nearly 18%.

```
R> rmsfe_pol = cbind(
+   fpm2$RMSFE$POL / fpm1$RMSFE$POL,
+   fpm3$RMSFE$POL / fpm1$RMSFE$POL,
+   fpm4$RMSFE$POL / fpm1$RMSFE$POL,
+   fpm5$RMSFE$POL / fpm1$RMSFE$POL,
+   fpm6$RMSFE$POL / fpm1$RMSFE$POL,
+   fpm7$RMSFE$POL / fpm1$RMSFE$POL
+ )
R> colnames(rmsfe_pol) = column_names
R> round(rmsfe_pol, 3)
```

	GP (pool)	CG (fix)	CG (est)	GPG (fix)	GPG (est)	VAR
gdp	0.997	0.462	0.672	1.009	0.995	1.590
UR	0.961	0.649	1.453	0.971	0.964	0.264
EPR	0.976	1.949	1.542	0.997	0.969	0.541
LFPR	0.992	5.010	4.893	1.031	0.979	1.198
joint	0.971	2.148	2.152	0.989	0.968	0.567

A similar characterisation of the relative point forecasting performance applies to results for Poland reported here for the sake presentation of the possibility a research might entertain.

5. Forecast Reconciliation

Consider the problem of reconciling the forecasts for the absolute numbers of unemployment, employment and labour force of a country. The three variables are linked by the accounting identity:

$$\text{Unemployment} + \text{Employment} - \text{Labour Force} = 0 \quad (50)$$

However, the probabilistic forecasts of these variables are bound to not satisfy the accounting identity above. Forecast reconciliation provides optimal solution by projecting the forecast on the space of the accounting identities with a minimum possible error as presented by [Di Fonzo and Girolimetto \(2023\)](#) and [Girolimetto, Athanasopoulos, Di Fonzo, and Hyndman \(2024\)](#). The solution provides forecasts that satisfy the accounting identities and are as close as possible to the original forecasts (see [Athanasopoulos, Hyndman, Kourentzes, and Panagiotelis 2024](#), for a recent review).

We implement the forecast reconciliation procedure in R using the package **FoReco** by [Girolimetto and Di Fonzo \(2024\)](#). Consider the five year ahead forecasts of labour market for some country collected in a matrix object `labour`:

```
R> labour
```

	unemployment	employment	labour force
1	0.9125837	10.09745	11.03233
2	0.7462074	10.19778	11.37523
3	0.9109497	10.30267	11.34960
4	0.8155660	10.39877	12.21829
5	0.6923606	10.49725	11.21166

Let the accounting identity be represented by a vector for a zero restriction

```
R> identity = matrix(c(1, 1, -1), nrow = 1)
```

```
R> identity
```

	[,1]	[,2]	[,3]
[1,]	1	1	-1

It is easy to verify that the forecasts indeed do not satisfy the accounting identity:

```
R> labour %%% t(identity)
```

	[,1]
1	-0.02229895
2	-0.43124387
3	-0.13598098
4	-1.00395718
5	-0.02204470

which does not give the exact 0 at any forecast horizon.

Apply function `csrec()` from the package **FoReco** to reconcile the forecasts. The function requires the matrix of original forecasts, namely `labour`, and the matrix representing the accounting identities, given in `identity`. The function returns the reconciled forecasts in object `reco`.

```
R> library(FoReco)
```

```
Loading required package: Matrix
```

```
R> reco = csrec(base = labour, cons_mat = identity)
```

```
R> reco
```

```
      unemployment employment labour force
h-1      0.9200167    10.10488    11.02490
h-2      0.8899553    10.34153    11.23148
h-3      0.9562767    10.34799    11.30427
h-4      1.1502184    10.73342    11.88364
h-5      0.6997089    10.50460    11.20431
attr(,"FoReco")
<environment: 0x110021200>
```

Indeed, the reconciled forecasts are very close to the original ones with the root-mean-squared error of 0.164. However, the reconciled forecasts now satisfy the accounting identity nearly exactly, which can be verified by running:

```
R> reco %%% t(identity)
```

```
      [,1]
h-1 0.000000e+00
h-2 0.000000e+00
h-3 -1.776357e-15
h-4 1.776357e-15
h-5 1.776357e-15
```

References

- Athanasopoulos G, Hyndman RJ, Kourentzes N, Panagiotelis A (2024). "Forecast reconciliation: A review." *International Journal of Forecasting*, 40(2), 430–456. doi:<https://doi.org/10.1016/j.ijforecast.2023.10.010>.
- Bauwens L, Lubrano M, Richard JF (1999). *Bayesian Inference in Dynamic Econometric Models*. Advanced Texts in Econometrics. Oxford University Press, Oxford. doi:[10.1093/acprof:oso/9780198773122.001.0001](https://doi.org/10.1093/acprof:oso/9780198773122.001.0001).
- Casella G, George EI (1992). "Explaining the Gibbs Sampler." *The American Statistician*, 46(3), 167–174. doi:[10.1080/00031305.1992.10475878](https://doi.org/10.1080/00031305.1992.10475878).
- Di Fonzo T, Girolimetto D (2023). "Cross-temporal forecast reconciliation: Optimal combination method and heuristic alternatives." *International Journal of Forecasting*, 39(1), 134. doi:[10.1016/j.ijforecast.2021.08.004](https://doi.org/10.1016/j.ijforecast.2021.08.004).
- Doan T, Litterman RB, Sims CA (1984). "Forecasting and Conditional Projection Using Realistic Prior Distributions." *Econometric Reviews*, 3(1), 1–100. doi:[10.1080/07474938408800053](https://doi.org/10.1080/07474938408800053).
- Gerland P, Raftery AE, Ševčíková H, Li N, Gu D, Spoorenberg T, Alkema L, Fosdick BK, Chunn J, Lalic N, *et al.* (2014). "World population stabilization unlikely this century." *Science*, 346(6206), 234–237.

- Girolimetto D, Athanasopoulos G, Di Fonzo T, Hyndman RJ (2024). “Cross-temporal probabilistic forecast reconciliation: Methodological and practical issues.” *International Journal of Forecasting*, **40**(3), 2285. doi:10.1016/j.ijforecast.2023.10.003.
- Girolimetto D, Di Fonzo T (2024). *FoReco: Forecast Reconciliation*. R package version 1.0. doi:10.32614/CRAN.package.FoReco. URL <https://cran.r-project.org/package=FoReco>.
- Jarociński M (2010). “Responses to monetary policy shocks in the east and the west of Europe: a comparison.” *Journal of Applied Econometrics*, **25**(5), 833–868. doi:10.1002/jae.1082.
- Karlsson S (2013). “Forecasting with Bayesian Vector Autoregression.” In *Handbook of Economic Forecasting*, volume 2, pp. 791–897. Elsevier. doi:10.1016/B978-0-444-62731-5.00015-4.
- Kass RE, Raftery AE (1995). “Bayes Factors.” *Journal of the American Statistical Association*, **90**(430), 773–795. doi:10.1080/01621459.1995.10476572.
- Raftery AE, Zimmer A, Frierson DM, Startz R, Liu P (2017). “Less than 2 °C Warming by 2100 Unlikely.” *Nature Climate Change*, **7**(9), 637–641. doi:10.1038/nclimate3352.
- Rendon SR (2013). “Fixed and Random Effects in Classical and Bayesian Regression.” *Oxford Bulletin of Economics and Statistics*, **75**(3), 460–476. doi:10.1111/j.1468-0084.2012.00700.x.
- Sanchez-Martinez M, Woźniak T (2025). “Forecasting with Bayesian Panel Vector Autoregressions Using the R Package bpvars.” *Working Paper*, pp. 1–34.
- Sims CA (1980). “Macroeconomics and Reality.” *Econometrica*, **48**(1), 1–48. doi:10.2307/1912017.
- Woźniak T (2016). “Bayesian Vector Autoregressions.” *Australian Economic Review*, **49**(3), 365–380. doi:10.1111/1467-8462.12179.
- Woźniak T (2025). *bpvars: Forecasting with Bayesian Panel Vector Autoregressions*. R package version 1.0, URL <http://bsvars.org/bpvars/>.
- Zellner A, Hong C (1989). “Forecasting international growth rates using Bayesian shrinkage and other procedures.” *Journal of Econometrics*, **40**(1). doi:10.1016/0304-4076(89)90036-5.

Affiliation:

Miguel Sanchez-Martinez
International Labour Organization
Macro-Economic Policies and Jobs Unit
ILO Research Department
E-mail: sanchezmartinez@ilo.org

Tomasz Woźniak
University of Melbourne
Department of Economics
111 Barry Street
3053 Carlton, VIC, Australia
E-mail: tomasz.wozniak@unimelb.edu.au
URL: <https://github.com/donotdespair>