Computational Economics
https://doi.org/10.1007/s10614-024-10718-x

High-Dimensional Dynamic Panel with Correlated Random Effects: A Semiparametric Hierarchical Empirical Bayes Approach

Antonio Pacifico[1]

Accepted: 10 September 2024

Abstract
A novel multivariate dynamic panel data analysis with correlated random effects is proposed for estimating high-dimensional parameter spaces. A semiparametric hierarchical Bayesian strategy is used to jointly address incidental parameters, endogeneity issues, and model mis-specification problems. The underlying methodology involves an `ad-hoc` model selection based on conjugate informative proper mixture priors to select promising subsets of predictors affecting outcomes. Monte Carlo algorithms are then conducted on the resulting submodels to construct empirical Bayes estimators and investigate ratio-optimality and posterior consistency for forecasting purposes and policy issues. An empirical approach is applied to a large panel of economies, describing the functioning of the model. Simulations based on Monte Carlo designs are also performed to account for relative regrets dealing with cross-sectional heterogeneity.

Keywords Multidimensional data · Bayesian inference · Conditional forecasting · Incidental parameters · Tweedie correction · Multicountry analysis

1 Introduction

Dynamic panel data (DPD) models are widely used in empirical economics for forecasting individuals' future outcomes (see, e.g., Hirano, 2002; Gu & Koenker, 2017b; Liu, 2018; Liu et al., 2020) and for controlling unobserved time-invariant individual heterogeneity (see, e.g., Chamberlain, 1984; Arellano & Bond, 1991 (linear case); Chamberlain, 2010; Arellano & Bonhomme, 2011 (non-linear case)). Such heterogeneity is an important issue, and failure to control for it results in misleading

✉ Antonio Pacifico
antonio.pacifico@unimc.it; antonio.pacifico86@gmail.com

[1] Department of Economics and Law, University of Macerata, Piazza Strambi 1, Macerata, Italy

Published online: 01 October 2024

inferences. This problem is even more severe when the unobserved heterogeneity may be correlated with covariates.

Consider a simple DPD model:

$$y_{it} = w_{i,t-1}\mu_i + \beta y_{i,t-1} + u_{it} \tag{1}$$

where $i = 1, \ldots, N$, $t = 1, \ldots, T$, y_{it} and $y_{i,t-1}$ are $NT \cdot 1$ and $N(T-1) \cdot 1$ vectors denoting the outcomes and their first lags, μ_i is a $N \cdot 1$ vector referring to individual–specific intercept with $w_{i,t-1} = 1$, and $u_{it} \sim N(0, \sigma^2)$ is a $NT \cdot 1$ vector of independent and identically distributed (*i.i.d.*) shocks.

Despite its widespread use, the model (1) faces some main limitations. This paper focuses on three related open research questions: the incidental parameters problem; high-dimensional settings; and nonlinear relationships. (i) Concerning the former, this arises from the presence of individual–specific effects (μ_i) in panel data models. These effects need to be estimated along with the other parameters of the model, and the problem becomes particularly pronounced when the time dimension (T) is small relative to the cross-sectional dimension (N). The challenges due to incidental parameters are highly severe in dynamic panels where behavioural effects over time are jointly measured with individual–specific effects. While the incidental parameters to be estimated are consistent in least squares methods, maximum likelihood estimation leads to inconsistent estimates affecting the dynamics of data (see, for instance, Nickell, 1981). Solutions such as difference generalized method of moments (GMM), system GMM, and bias-corrected estimators involve analytical or bootstrap corrections to reduce the incidental parameters bias (Arellano & Bond, 1991; Arellano & Bover, 1995; Blundell & Bond, 1998; Alvarez & Arellano, 2003; Arellano & Hahn, 2016, 2007; Bester & Hansen, 2009; Fernandez-Val, 2009; Hahn & Kuersteiner, 2011). (ii) A DPD model is not well-suited for high-dimensional settings due to issues related to overfitting, leading to poor generalization to new data and hence resulting in unreliable and unstable estimates, multicollinearity, leading to imprecise estimates and difficulties in interpreting the effects of individual predictors, and computational inefficiency, slowing down the analysis and making it difficult to apply standard techniques to very large datasets. Moreover, standard DPD models rely on parametric assumptions and linear relationships between variables that, in high-dimensional settings, may be too restrictive, leading to model mis-specification and biased estimates. Finally, as the dimensionality increases, the volume of the space increases exponentially, making the data sparser and more challenging to analyze. (iii) DPD models also fail to capture the complex and nonlinear relationships present in real-world data due to linearity assumption, overlooking interaction effects, limited flexibility, incurring model mis-specification problems, endogeneity issues, without dealing with unobserved heterogeneity and omitted variable biases, and static structure, ignoring the possibility of time-varying relationships or state-dependent dynamics. Methods like hierarchical Bayesian approaches, dynamic factor models, and semiparametric/nonparametric models generally provide powerful tools to address the challenges of high-dimensionality and complex nonlinear relationships (Hirano, 2002; Gelman & Hill, 2012; Forni et al., 2000; Norets & Pelenis, 2012; Liu et al., 2020).

This paper aims to develop a computational approach to jointly deal with high-dimensionality, incidental parameters problem, and nonlinearities. It focuses on a structural semiparametric hierarchical Bayesian approach to conduct inference in high-dimensional dynamic panel data with cross-sectional heterogeneity. The model is called the hierarchical dynamic panel Bayesian model with correlated random effects (HDPB-CRE) and is achieved by combining an implemented version of Pacifico (2020)'s analysis, which develops a robust open Bayesian (ROB) procedure for improving Bayesian model averaging (BMA) in multiple high-dimensional linear regression models, and the Liu et al. (2020)'s framework, which constructs point predictors using Tweedie's formula for the posterior mean of heterogeneous coefficients under a correlated random effects distribution.

The ROB procedure is used to select the `best` number of predictors that affects outcomes when dealing with large paramater and model space and overfitting problems. It uses conjugate informative proper mixture (CIPM) priors to obtain posterior estimates. The term 'proper' refers to assigning different prior moments based on the observed data, which is particularly useful for accounting for dynamics when studying time-varying data. Multicollinearity and complex relationships are also avoided because the shrinkage is performed on a set of informative priors (`robust`) and the posterior distributions are computed via Markov chains based on the Posterior model probabilities (PMPs) across every model solution (or combinations of predictors). Given the observed data and prior beliefs, the PMPs refer to the probability that a particular model is the `best` one, where `best` stands for the model solution better fitting the data and hence predicting the outcomes. A PMP close to 1 implies strong evidence that a model solution is correct. In Pacifico (2020)'s analysis, the ROB procedure has been applied in a multivariate context to evaluate a structural panel VAR model. In a hierarchical framework, the procedure can be implemented within a multidimensional panel setting.

Empirical Bayes (EB) estimators combine the strengths of Bayesian methods with computational efficiency and simplicity. Indeed, empirical Bayes methods involve a two-step process to estimate hyperparameters by using maximum likelihood or similar techniques, and then estimating the remaining parameters conditionally based on these hyperparameters. It is more efficient compared to the iterative sampling required—for example—in full Bayesian methods. EB approaches also simplify the prior specification by estimating hyperparameters directly from the data, making the procedure more straightforward to implement. For instance, in a high-dimensional dynamic panel data model, estimating the variances and covariances of random effects directly from the data simplifies the modeling process compared to specifying complex hierarchical priors. CIPM priors allow for defining a simpler hierarchical structure to model time-varying coefficients, the covariance matrix, and the volatility process. The latter is modeled by a stochastic process to reduce dimensionality and replace volatility changes with coefficient changes. The lagged distributions of the parameters are estimated using AutoRegressive coefficients in the procedure. Finally, correlated random effects are used to compute posterior means, addressing cross-sectional heterogeneity that affects outcomes.

Earlier works regarding empirical Bayes methods with parametric priors on heterogeneous parameters refer to Robbins (1964), Robert (1994), Brown and

Greenshtein (2009), and Jiang and Zhang (2009). More recently, nonparametric approaches have been developed by Liu et al. (2020) and Liu et al. (2019) (hereafter LMS) and Gu and Koenker (2017a, b) (hereafter GK). LMS aim to forecast a collection of short time-series using cross-sectional information. Then, they construct point forecasts predictors using Tweedie's formula[1] for the posterior mean of heterogeneous individual–specific factors under a correlated random effects distribution. They show that the ratio optimality of point forecasts asymptotically converge to the one based on a nonparametric kernel estimate of the Tweedie correction. However, they replace the μ_i's distribution with a kernel density estimator that would perform less accurate forecasts than alternative estimates of the Tweedie correction such as nonparametric maximum likelihood estimation and finite mixture of normal distributions. They also estimate relative regrets for these two alternative approaches via Markov chains simulations, but specifying bounds for the domain of the μ_i's and partitioning it into default setting bins. This compromises the estimates because of weak empirical forecast optimality limited to restrictive and constrained classes of models. GK use Tweedie's formula to construct an approximation to the posterior mean of the heterogeneous parameters. They build on Kiefer and Wolfowitz (1956) and implement the empirical Bayes predictor based on a nonparametric maximum likelihood estimator of the cross-sectional distribution of the sufficient statistics. However, no theoretical optimality results are provided. In addition, neither LMS nor GK face variable selection problems and causality relationships in the shrinkage of large panel parameter spaces.

In this study, the multivariate panel data model is unbalanced and includes large cross-sectional dimension N and sufficiently large time-series T. Methodologically, Markov Chain Monte Carlo (MCMC) algorithms and implementations are used to construct posterior distributions and then perform cross-country conditional forecasts and policy issues. Theoretically, ratio-optimality and posterior consistency are also investigated to account for relative regrets when modelling individual–specific heterogeneity.

The main contributions of this paper can be summarized as follows. First, in a hierarchical framework, multivariate conjugate informative proper mixture (mvCIPM) priors are used to select the `best` promising subset of covariates according to their PMPs, which denote the probability to better explain and thus fit the data in high-dimensional model classes. The mvCIPM priors are an implementation of the conjugate informative priors in Pacifico (2020) by adapting the prior specification strategy to large multidimensional (panel) setups with incidental parameters. The main thrust is to jointly deal with variable selection problems and causal relationships. The former stand for endogeneity issues (because of omitted factors and unobserved heterogeneity), structural model uncertainty (because of some functional forms of mis-specification), and overfitting (when complex[2] models always provide a somewhat better fit to the data than simpler models). Causality in dynamic panel data is assessed according to the Granger (non-)causality test (see, for instance, Dumitrescu & Hurlin 2012).

[1] The formula is attributed to the astronomer Arthur Eddington and the statistician Maurice Tweedie.

[2] The 'complexity' stands, for example, for the number of unknown parameters.

Second, to accomodate the correlated random coefficients model, I involve in the previous shrinking process an EB procedure, where the posterior mean of the μ_i's is expressed in terms of the marginal distribution of a sufficient statistic ($\hat{\mu}_i(\beta)$) estimated from the cross-sectional whole information (Tweedie's formula). Dealing with variable selection problems and causal inference, I implement the Liu et al. (2020)'s framework by constructing nonparametric Bayesian statistics through finite mixture approximation of multivariate (FMM) distributions. The latter are evaluated via MCMC integrations to (i) maximize the log likelihood function of the estimation procedure (expectation-maximization (EM)), (ii) use EB estimators to draw posteriors for $\hat{\mu}_i(\beta)$ from the joint distribution between some sufficient statistics designed for the μ_i's and indivudal outcomes (Metropolis–Hastings algorithm), and (iii) analytically compute posterior distributions for the time-varying estimates (Kalman-Filter algorithm). In this context, lagged covariates and outcomes from AutoRegressive (AR) processes are introduced on the right-hand side of the estimation model as `external` instruments to account for (potential) correlation between predictors and residual errors (see, e.g., Arellano & Bond, 1991).

Third, better conditional forecasts are involved in HDPB-CRE because of three main features: (i) the use of a semiparametric Bayesian approach modeling either time-varying and fixed effects; (ii) the use of a hierarchical framework to construct proper informative priors disentangling heterogeneous and common parameters; and (iii) the observation of incidental parameters treated as random variables possibly correlated with some of the predictors within the system.

An empirical application is conducted to highlight the functioning and performance of the methodology. It builds on a pool of advanced and emerging economies and evaluates a large set of data including socioeconomic–demographic factors, policy tools, and economic–financial issues during the period 1990–2021. Forecasting analysis is addressed to perform policy-relevant strategies safeguarding against (future) sudden outbreak on the global economy.

A simulated experiment using MCMC-based designs is also addressed to highlight the performance of the estimating procedure with related works.

The remainder of this paper is organized as follows. Section 2 introduces the econometric model and the estimating procedure. Section 3 displays prior specification strategy and posterior distributions accounting for empirical Bayes estimator (Tweedie correction), ratio-optimality, and Markov Chain algorithms. Section 4 describes the data and the empirical analysis. Section 5 presents the simulated experiment dealing with relative regrets for Tweedie correction. The final section contains some concluding remarks.

2 Dynamic Panel Data and Shrinking Process

2.1 Econometric Model

The baseline hierarchical DPD model, stacking for the sub-index parameters, is:

$$y_{it} = \beta_l y_{i,t-l} + \alpha x_{it} + \gamma_l z_{i,t-l} + \mu_i + u_{it} \tag{2}$$

where the subscripts $i = 1, 2, \ldots, N$ are country indices, $t = 1, 2, \ldots, T$ denotes time, y_{it} is a $NT \cdot 1$ vector of outcomes, $y_{i,t-l}$ and $z_{i,t-l}$ are $N(T-\lambda) \cdot M$ and $N(T-\lambda) \cdot P$ matrices of predetermined ($m = 1, 2, \ldots, M$) and directly observed (endogenous) variables ($p = 1, 2, \ldots, P$) for each i, respectively, with $l = 0, 1, 2, \ldots, \lambda$, $\beta_{\tilde{l}}$ and $\gamma_{\tilde{l}}$ are the autoregressive coefficients to be estimated for each i, with $\tilde{l} = 1, \ldots, \lambda$, x_{it} is a $NT \cdot K$ vector of strictly exogenous factors for each i, with $k = 1, 2, \ldots, K$ and α denoting the regression coefficients to be estimated, μ_i is a $N \cdot 1$ heterogeneous intercept containing—for example—time-constant differences (such as territorial competitiveness, infrastructural system, competitiveness developments, macroeconomic imbalances), and $u_{it} \sim i.i.d.N(0, \sigma_u^2)$ is a $NT \cdot 1$ vector of unpredictable shock (or idiosyncratic error term), with $E(u_{it}) = 0$ and $E(u_{it} \cdot u_{js}) = \sigma_u^2$ if $i = j$ and $t = s$, and $E(u_{it} \cdot u_{js}) = 0$ otherwise. In this study, I consider the same lag order (or optimal lag length) for both predetermined ($y_{i,t-l}$) and observed variables ($z_{i,t-l}$).

Here, some considerations are in order: (i) the predetermined variables contain lagged control variables (e.g., economic status) and lagged outcomes (capturing, for example, the persistence); (ii) the μ_i's denote cross-sectional heterogeneity affecting the outcomes; (iii) correlated random effects matter and then μ_i's are possibly correlated with some of the covariates within the system; (iv) the roots of $\tilde{l}(L) = 0$ lie outside the unit circle so that the AR processes involved in the model (2) are stationary, with L denoting the lag operator; (v) the x_{it}'s strictly exogenous factors contain dummy variables to test—for example—the presence of structural breaks or policy shifts; and (vi) the instruments are fitted values from AR parameters based on all the available lags of the time-varying variables. In this study, the order of integration and the optimal lag length have been set using the the Augmented Dickey–Fuller (ADF) test for each i and the Arellano's test (see, for instance, Arellano, 2003; Arellano & Honore, 2001), respectively.

Let the stationarity hold in (2), the time-series regressions would be valid and the estimates feasible. However, some moment restrictions need to hold to address exact identification in a context of correlated random effects and estimate $\beta_{\tilde{l}}$ and $\gamma_{\tilde{l}}$ for $T \geq 3$ (see, for instance, Anderson & Hsiao, 1981; Arellano & Honore, 2001; Blundell & Bond, 1998). In this study, I assume that μ_i and $u_{i,t}$ are independently distributed across i and have the familiar error components structure:

$$E(\mu_i) = 0, E(u_{it}) = 0, E(u_{it} \cdot \mu_i) = 0 \qquad for \quad i = 1, \ldots, N \quad and \quad t = 2, \ldots T \tag{3}$$

$$E(u_{it} \cdot u_{is}) = 0 \qquad for \quad i = 1, \ldots, N \quad and \quad t \neq s \tag{4}$$

Then, I also assume the standard assumption concerning the initial conditions $y_{i,t=1}$:

$$E(y_{i,t=1} \cdot u_{it}) = 0 \qquad for \quad i = 1, \ldots, N \quad and \quad t = 2, \ldots T \tag{5}$$

2.2 Multivariate ROB Procedure in Longitudinal Data

When cross-sectional dimension (N) and time-series (T) are high-dimensional, the estimates of the common parameters (β_l, α γ_l, σ_u^2) in (2) would result biased and inconsistent. Furthermore, leaving the individual heterogeneity unrestricted, the number of individual–specific effects would grow with the sample size and be highly contaminated by the shock u_{it}, leading to inaccurate forecasts. Last but not least, when dealing with time-varying and high-dimensional data, variable selection problems such as overshrinkage/undershrinkage, model mis-specification problems, endogeneity issues, and model uncertainty[3] also matter in DPD models, resulting in inconsistent estimates. The multivariate ROB procedure involved in this study addresses the above-mentioned issues. It moves forward three steps. (i) MCMC-based PMPs are conducted to obtain a reduced subset of promising model solutions (or combination of predictors) fitting the data when dealing with variable selection problems. (ii) A further shrinkage is conducted to obtain a smaller subset of promising submodels with statistically significant predictive capability (accurate forecast). Here, nonparametric Bayesian statistics are also addressed through MCMC implementations to model and quantify correlated random effects in large longitudinal data. (iii) A final shrinkage is addressed according to the Granger (non-)causality test in multivariate dynamic panel data. The idea is to exclude the predictors when no causal link holds across units within the panel (homogeneity under the null hypothesis); conversely, whether highly strong causal links matter for a subgroup of units (heterogeneity under the alternative), the same parameters should be taken into account to deal with overestimation of effect sizes (or individual contributions). In this study, the optimal lag length testing Granger-causality is set using the Arellano's test. This latter step refers to the main novelty with respect to Pacifico (2020)'s analysis.

In a hierarchical panel structure, the use of a multivariate shrinking procedure helps to manage the complexity, mainly when dealing with high-dimensional data, and then improve the reliability of model parameter estimates. Overall, by stabilizing estimates,[4] borrowing strength across groups,[5] and penalizing complexity,[6] the multivariate shrinkage procedure affects the marginal likelihood of models and hence the Posterior Model Probabilities, improving estimation accuracy and interpretability. This differential impact leads to varying magnitudes of posterior probabilities, favoring models that balance fit and parsimony effectively.

Given the HDPB-CRE in (2), I decompose the vectors of the observed endogenous variables: $y_{i,t-l} = \left[y_{i,t-l}^{o'}, y_{i,t-l}^{c'}\right]'$, with $y_{i,t-l}^{o'}$ denoting lagged outcomes to capture the per-

[3] Model uncertainty matters when a given model is set to be true without estimating the evidence for alternative model solutions.

[4] Shrinkage reduces the variance of parameter estimates, leading to more stable and reliable estimates. This can improve the model fit by avoiding overfitting.

[5] By pooling information across groups and individuals, shrinkage allows for better estimation in each subgroup, improving the overall model fit.

[6] Hierarchical priors implicitly penalize model complexity by shrinking parameters towards common values, which can make simpler models more favorable in terms of posterior probability.

sistence and $y_{i,t-l}^{c'}$ including lagged control variables such as general economic conditions; and $z_{i,t-l} = \left[z_{i,t-l}^{s'}, z_{i,t-l}^{p'}\right]'$, referring to other lagged factors such as socioeconomic conditions ($z_{i,t-l}^{s'}$) and policy implications ($z_{i,t-l}^{p'}$). Then, I combine the (non-)homogeneous parameters into the vector $\theta = \left(\beta_l^{o'}, \beta_l^{c'}, \alpha', \gamma_l^{s'}, \gamma_l^{p'}\right)'$.

In order to model the key latent heterogeneities (μ_i) and observed determinants ($y_{i,t-l}$, x_{it}, $z_{i,t-l}$) when dealing with high-dimensional analysis, I define the conditioning set at period t (c_{it}) and the structural density ($D(y_{it}|\cdot)$) as:

$$c_{it} = \left(y_{i,0:t-l}^{o}, y_{i,0:t-l}^{c}, z_{i,0:t-l}^{s}, z_{i,0:t-l}^{p}, x_{i,0:t}\right) \tag{6}$$

and

$$D\left(y_{it}|y_{i,t-l}, x_{it}, z_{i,t-l}, \mu_i\right) = D\left(y_{it}|y_{i,t-l}, x_{it}, z_{i,t-l}, y_{i0}, \mu_i\right) \tag{7}$$

The error terms (u_{it}) are individual-time-specific shocks characterized by zero mean and homoskedastic Gaussian innovations. In a unified and hierarchical framework, I combine the individual heterogeneity into the vector $\phi_i = \left(\mu_i, \sigma_u^2\right)$ under cross-sectional homoskedasticity. Assuming correlated random coefficients model, ϕ_i and c_{i0} could be correlated with each other, with:

$$c_{i0} = \left(y_{i,0}^{o}, y_{i,0}^{c}, z_{i,0}^{s}, z_{i,0}^{p}, x_{i,0:T}\right) \tag{8}$$

Given these primary specifications, the HDPB-CRE model in (2) would be less parsimonious and harder to implement due to high-dimensional parameter spaces.

Let $\mathcal{F}$ be the full panel set containing all (potential) model solutions, the `first step` of the multivariate ROB procedure is addressed by imposing an auxiliary indicator variable χ_h, with $h = 1, 2, \ldots, m$, containing every possible 2^m subset choices, where $\chi_h = 0$ if θ_h is small (absence of h-th covariate in the model) and $\chi_h = 1$ if θ_h is sufficiently large (presence of h-th covariate in the model). According to the Pacifico (2020)'s framework, I match all potential candidate models to shrink both the model space and the parameter space. The shrinking jointly deals with overestimation of effect sizes (or individual contributions) and model uncertainty (implicit in the procedure) by using Posterior Model Probabilities for every candidate model. They can be defined as:

$$\pi(M_h|y) = \frac{\pi(M_h) \cdot \pi(y|M_h)}{\sum_{M_h \in \mathcal{M}} \pi(M_h) \cdot \pi(y|M_h)} \qquad \text{with} \qquad M_h \in \mathcal{M} \tag{9}$$

where $pi(M_h)$ is the marginal prior distribution of M_h, $\pi(y|M_h) = \int_{\mathcal{B}} \pi(y|M_h, \theta_h) \cdot \pi(\theta_h|M_h, y) d\theta_h$ is the marginal likelihood, with $\pi(\theta_h|M_h, y)$ referring to the conditional prior distribution of θ_h. With N high-dimensional and T sufficiently large, the calculation of the integral $\pi(y|M_h)$ is unfeasible and then Markov Chain Monte Carlo algorithms need to be conducted.

The subset containing the `best` model solutions will correspond to:

$$\mathcal{S} = \left\{ M_j : M_j \subset \mathcal{S}, \mathcal{S} \in \mathcal{F}, \Theta_j \subset \Theta_h, \sum_{j=1}^{\varpi} \pi\left(M_j | y_i = y_i, \chi\right) \geq \tau \right\} \tag{10}$$

where M_j denotes the submodel solutions of the HDPB-CRE in (2), with $M_j < M_h$, $j \ll h$, $\{1 \leq j < h\}$, and τ is a threshold chosen arbitrarily for sufficient posterior consistency. With high-dimensional N (> 1000), a threshold sufficiently small ($< 1\%$) would ensure that the PMPs concentrate on the true model. In this study, the convergence is found with $\tau = 0.5\%$.

The `second step` consists of reducing the model space $\mathcal{S}$ to obtain a smaller subset of `best` submodel solutions:

$$\mathcal{E} = \left\{ M_\xi : M_\xi \subset \mathcal{E}, \mathcal{E} \in \mathcal{S}, \sum_{j=1}^{\varpi} \pi\left(M_j | y_i = y_i, \dot{\chi}\right) \geq \dot{\tau} \right\} \tag{11}$$

where $M_\xi \ll M_j$, $\pi(M_j | y_i = y_i, \dot{\chi})$ denotes the PMPs, with $\dot{\chi}$ denoting a new auxiliary variable containing the only `best` model solutions in the subset $\mathcal{S}$ and $\dot{\tau}$ referring to a new arbitrary threshold to evaluate the probability of the model solutions in $\mathcal{S}$ performing the data (PMPs). In this study, I still use $\tau = 0.5\%$—independently of N—for sufficient prediction accuracy explaining the data.

Finally, the multivariate ROB procedure comes to a conclusion (`third step`) once a further shrinkage is conducted according to the panel Granger (Non-)Causality test in order to obtain the smallest final subset of `best` promising submodel solutions ($M_{\xi^*} \subset \mathcal{E}$). More precisely, this last step consists of including the only candidate predictors displaying highly strong causal links for at least a subgroup of units (heterogeneity under the alternative) with p-value $\leq \dot{\tau}$. To deal with endogeneity issues and mis-specified dynamics, all available lags of the `best` candidate predictors—obtained in the previous step—are included as instruments.

The final model solution to have to be considered for performing forecasting and policy-making will correspond to one of the submodels M_{ξ^*} with higher log natural Bayes Factor (lBF):

$$lBF_{\xi^*,\xi} = log\left\{ \frac{\pi(M_{\xi^*} | y_i = y_i)}{\pi(M_\xi | y_i = y_i)} \right\} \tag{12}$$

In this analysis, the lBF is interpreted according to the scale evidence in Pacifico (2020), but with more stringent conditions:

$$\begin{cases} 0.00 \leq lB_{\xi^*,\xi} \leq 4.99 & \text{no evidence for submodel } M_{\xi^*} \\ 5.00 \leq lB_{\xi^*,\xi} \leq 9.99 & \text{moderate evidence for submodel } M_{\xi^*} \\ 10.00 \leq lB_{\xi^*,\xi} \leq 14.99 & \text{strong evidence for submodel } M_{\xi^*} \\ lB_{\xi^*,\xi} \geq 15.00 & \text{very strong evidence for submodel } M_{\xi^*} \end{cases} \tag{13}$$

3 Semiparametric Hierarchical Bayesian Approach

3.1 Prior Specification Strategy and Tweedie's Formula

The variable specification strategy entails estimating χ_h and θ_h as posterior means (the probability that a variable is `in` the model). All observal variables in c_{it} and individual heterogeneity in ϕ_i are hierarchically modelled via mvCIPM priors.

In Bayesian hierarchical models, the prior is designed to reflect the different levels of the hierarchy and the relationships between them. More precisely, let a hierarchical model typically involve parameters at multiple levels, the data have been clustered within groups, where each group includes a pool of parameters defined in the conditioning set c_{it}. According to model (2), two different levels of parameters have been defined: a group–level parameters (higher-level), describing the characteristics of the groups/clusters as a whole (θ) and individual–level parameters (lower-level), denoting individual–specific characteristics within each group (μ_i).

In this context, the MROB procedure involved in the shrinking process is able to deal with variations in cluster characteristics possibly affecting the choice of priors. Indeed, in the case of homogeneous clusters, the inclusion of the thresholds τ and $\dot{\tau}$ in the hierarchical panel set allows the use of tighter priors around the group–level parameters, reflecting the belief that there is not much variation between clusters. Conversely, if there is significant variation between clusters (heterogeneous clusters), greater variability among the group–level parameters is allowed through the use of mvCIPM priors. These latter are used to construct posterior distributions to improve their estimates, especially when the data is limited.

The CIPM priors are so defined:

$$\pi(\theta, \phi, \chi) = \pi(\theta|\chi) \cdot \pi(\mu_i|\chi, y_{i0}) \cdot \pi(\sigma_u^2|\chi) \cdot \pi(\chi) \tag{14}$$

where

$$\pi\left(\theta|\mathfrak{F}_{-1}\right) = N\left(\bar{\theta}, \bar{\rho}\right) \tag{15}$$

$$\pi(\mu_i|\theta) = N\left(\delta_{\mu_i}, \Psi_{\mu_i}\right) \quad \texttt{with} \quad \delta_{\mu_i} \sim N\left(0, \zeta\right) \quad \texttt{and} \quad \Psi_{\mu_i} \sim IG\left(\frac{\varphi}{2}, \frac{\varepsilon}{2}\right) \tag{16}$$

$$\pi(y_{i0}|\mu_i) = N(0, \kappa) \tag{17}$$

$$\pi(\chi) = w_{|\chi|} \cdot \begin{pmatrix} h \\ |\chi| \end{pmatrix}^{-1} \tag{18}$$

$$\pi(\sigma_u^2) = IG\left(\frac{\bar{\omega}}{2}, \frac{\nu}{2}\right) \tag{19}$$

where $N(\cdot)$ and $IG(\cdot)$ stand for normal and inverse-gamma distribution, respectively, $\mathfrak{F}_{-1}$ refers to the cross-sectional information available at time -1, κ in (17) refers to the decay factor, and $w_{|\chi|}$ in (18) denotes the model prior choice related to the sum of the PMPs (or Prior Inclusion Probabilities) with respect to the model size $|\chi|$, through which the θ's will require a non-0 estimate or the χ's should be included in the model. The decay factor usually varies in the range [0.9–1.0] and controls the process of reducing past data by a constant rate over a period of time. In this way, one would weight more according to model size and—setting $w_{|\chi|}$ large for smaller $|\chi|$—assign more weight to parsimonious models.

All hyperparameters are known. More precisely, collecting them in a vector $\tilde{\omega}$, where $\tilde{\omega} = \left(\bar{\theta}, \bar{\rho}, \zeta, \varphi, \varepsilon, \kappa, w_{|\chi|}, \bar{\omega}, \nu\right)$, they are treated as fixed and are either obtained from the data to tune the prior to the specific applications (such as $\varphi, \kappa, w_{|\chi|}, \bar{\omega}$) or selected a priori to produce relatively loose priors (such as $\bar{\theta}$, $\bar{\rho}, \zeta, \varepsilon, \nu$). Here, $w_{|\chi|}$ is restricted to a benchmark prior $max\left(NT, |\chi|\right)$ according to the non-0 components of χ.

To accomodate the correlated random coefficients model where the individual–specific heterogeneity (μ_i) can be correlated with the conditioning variables c_{i0} and y_{i0}, I use an empirical Bayes procedure where the posterior mean of the μ_i's is expressed in terms of the marginal distribution of a sufficient statistic ($\hat{\mu}_i(\theta)$) estimated from the cross-sectional whole information (Tweedie's formula). The main difference between an empirical and fully Bayesian approach is that the former picks the μ_i distribution by maximizing the Maximum Likelihood (ML) of the data,[7] whereas a fully Bayesian method constructs a prior for the correlated random effects and then evaluates it in view of the observed panel data.[8] Even if the fully Bayesian approach tends to be more suitable for density forecasting and more easily extended to the non-linear case, it would be a lot more computationally intensive.

In this study, I implement the EB predictor used in Liu et al. (2020) by using non-parametric Bayesian statistics to model and quantify correlated random effects. The latter are addressed through finite mixture approximation of multivariate (FMM) distributions, evaluated via MCMC integrations in order to maximize the log likelihood function (expectation-maximization (EM)) and then use EB estimators to draw posteriors for $\hat{\mu}_i(\theta)$ from the joint distribution between the μ_i's sufficient statistic and individual outcomes (Metropolis–Hastings algorithm).

Given the CIPM priors in (15)–(19), I define the compound risk and loss functions—under which the forecasts will be evaluated—accounting for expectations over the observed trajectories $\mathcal{Y}_i = \left(y_1^{0:T}, \ldots, y_N^{0:T}\right)$, with $y_i^{0:T} = \left(y_{i0}, y_{i1}, \ldots, y_{iT}\right)$, the unobserved heterogeneity $(\mu_i = \mu_1, \ldots, \mu_N)$, and the future shocks $u_{i,T+k} = \left(u_{1,T+k}, \ldots, u_{N,T+k}\right)$:

[7] See, e.g., Chamberlain and Hirano (1999), Hirano (2002), Lancaster (2002), Jiang and Zhang (2009), and Gu and Koenker (2017a, b).

[8] See, for instance, Liu (2018) and Liu et al. (2020) (linear case); and Liu et al. (2019) (non-linear case).

$$\mathcal{R}\left(\hat{y}_{i,T+k}\right) = \mathbb{E}_{\theta,\phi,\pi(\cdot)}^{(\mathcal{Y}_N,\mu_i,u_{i,T+k})}\left[L_N\left(\hat{y}_{i,T+k}, y_{i,T+k}\right)\right] \tag{20}$$

where $L_N\left(\hat{y}_{i,T+k}, y_{i,T+k}\right) = \sum_{i=1}^{N}\left(\hat{y}_{i,T+k} - y_{i,T+k}\right)^2$ denotes the compound loss obtained by summing over the units i the forecast error losses $(\hat{y}_{i,T+k} - y_{i,T+k})$, with $\hat{y}_{i,T+k} = (\hat{y}_{1,T+k}, \dots, \hat{y}_{N,T+k})'$ is a vector of k-period-ahead forecasts.

In the compound decision theory, the infeasible oracle forecast (or benchmark forecast) implies that the ϕ_i's and the distribution of the unobserved heterogeneity $(\pi(\mu_i, y_{i0}))$ are known, the trajectories $(\mathcal{Y}_i)$ are observed, and the values of the μ_i's are unknown across units i. Moreover, the integrated risk in (20) is minimized performing individual–specific forecasting that minimizes the posterior risk for each $\mathcal{Y}_i$. Thus, according to the Liu et al. (2020)'s framework, the posterior risk can be defined as:

$$\begin{aligned}\mathbb{E}_{\theta,\phi,\pi(\cdot)}^{(\mathcal{Y}_N,\mu_i,u_{i,T+k})}\left[L_N\left(\hat{y}_{i,T+k}, y_{i,T+k}\right)\right] = \sum_{i=1}^{N}\Bigg\{\left(\hat{y}_{i,T+k} - \mathbb{E}_{\theta,\phi,\pi(\cdot)}^{(\mathcal{Y}_i,\mu_i,u_{i,T+k})}[y_{i,T+k}]\right)^2 + \\ + \mathbb{V}_{\theta,\phi,\pi(\cdot)}^{(\mathcal{Y}_i,\mu_i,u_{i,T+k})}[y_{i,T+k}]\Bigg\}\end{aligned} \tag{21}$$

where $\mathbb{V}_{\theta,\phi,\pi(\cdot)}^{(\mathcal{Y}_i,\mu_i,u_{i,T+k})}[y_{i,T+k}]$ is the posterior predictive variance of $y_{i,T+k}$. The optimal predictor would be the mean of the posterior predictive distribution:

$$\hat{y}_{i,T+k}^{op} = \mathbb{E}_{\theta,\phi,\pi(\cdot)}^{(\mathcal{Y}_i,\mu_i,u_{i,T+k})}[y_{i,T+k}] = \mathbb{E}_{\theta,\phi,\pi(\cdot)}^{(\mathcal{Y}_i,\mu_i)}[\mu_i] + (\theta \cdot c_{it}) \tag{22}$$

where the acronym *op* stands for 'optimal'. Then, the compound risk in (20) associated with the infeasible oracle forecast can be rewritten as:

$$\mathcal{R}^{op} = \mathbb{E}_{\theta,\phi,\pi(\cdot)}^{(\mathcal{Y}_i,\mu_i,u_{i,T+k})}\left\{\sum_{i=1}^{N}\left(\mathbb{V}_{\theta,\phi,\pi(\cdot)}^{(\mathcal{Y}_i,\mu_i)}[\mu_i] + \sigma_u^2\right)\right\} \tag{23}$$

The optimal compound risk in (23) consists of two components: uncertainty concerning the individual–specific heterogeneity on the observations i and uncertainty with respect to the error terms. Because of the infeasible benchmark forecast, the parameter vectors (θ, ϕ) and the CRE distribution $(\pi(\cdot))$ are unknown. Thus, the posterior mean $\mathbb{E}_{\theta,\phi,\pi(\cdot)}^{(\mathcal{Y}_i,\mu_i)}[\mu_i]$ in (22) is assessed through Tweedie's formula by evaluating the marginal distribution of a sufficient statistic of the heterogeneous effects. The likelihood function associated with the multivariate HDPB-CRE in (2) is:

$$\begin{aligned}&\pi\left(y_i^{1:T}|y_{i0}, \mu_i, \theta\right) \propto exp\Bigg\{-\frac{1}{2\sigma_u^2} \\ &\sum_{t=1}^{T}\left(y_{it} - (c_{it-l}|\chi)\theta_t - \mu_i(\theta)\right)^2\Bigg\} \propto \left\{-\frac{T}{2\sigma_u^2}\left(\hat{\mu}_i(\theta) - \mu_i\right)^2\right\}\end{aligned} \tag{24}$$

where $\hat{\mu}_i(\theta)$ denotes the sufficient statistic and equals:

$$\hat{\mu}_i(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left(y_{it} - (\theta \cdot c_{it-l})\right) \tag{25}$$

According to Bayes's theorem, the posterior distribution of μ_i can be obtained as:

$$\pi\left(\mu_i|y_i^{0:T},\theta\right) = \pi\left(\mu_i|\hat{\mu}_i, y_{i0},\theta\right) = \frac{\pi\left(\hat{\mu}_i|\mu_i, y_{i0},\theta\right)\cdot\pi\left(\mu_i|y_{i0}\right)}{exp\left\{ln\left(\pi(\hat{\mu}_i|y_{i0})\right)\right\}} \tag{26}$$

The last step to obtain Tweedie's formula is to differentiate the equation $\pi\left(\mu_i|\hat{\mu}_i, y_{i0},\theta\right)$ in (26) with respect to $\hat{\mu}_i$ and solve the equation for the posterior mean $\mathbb{E}^{(y_i,\mu_i)}_{\theta,\phi,\pi(\cdot)}[\mu_i]$ in (22). Thus, Tweedie's formula equals:

$$\mathbb{E}^{(y_i,\mu_i)}_{\theta,\phi,\pi(\cdot)}[\mu_i] = \hat{\mu}_i(\theta) + \frac{\sigma_u^2}{T}\cdot\frac{\partial}{\partial\hat{\mu}_i(\theta)}ln\left(\hat{\mu}_i(\theta), y_{i0}\right) \tag{27}$$

where the second term in (27) denotes the correction term capturing heterogeneous effects of the prior of μ_i ($\pi(\cdot)$) on the posterior. It is expressed as a function of the marginal density of $\hat{\mu}_i(\theta)$ conditional on y_{i0} and θ; contrary to the full Bayesian approach, where one needs to avoid the deconvolution problem that disentangle the prior density $\pi(\mu_i|y_{i0})$ from the distribution of the error terms (u_{it}).

3.2 Tweedie Correction and MCMC Implementations

Tweedie correction entails substituting the unknow parameters θ and the joint distribution between the μ_i's sufficient statistic and individual outcome values $\pi\left(\hat{\mu}_i(\theta), y_{i0}\right)$ in (27) by estimates. According to the multivariate ROB procedure, the cross-sectional information uploaded within the system is set into $\mathcal{E}$. In dynamic panel data, consistent estimates of the unkown parameters θ can be obtaining through genelarized method of moments (GMM) estimators. In this study, they correspond to the AR(λ) coefficients related to predetermined and endogenous variables.[9] Let the stationarity and moment conditions in (3–5) hold in the system, the time-series regressions are valid and GMM estimators are feasible. Concerning the density $\pi\left(\hat{\mu}_i(\theta), y_{i0}\right)$, I estimate it by using FMM distributions:

$$\pi_{mix}\left(\hat{\mu}_i, y_{i0} \mid |\chi|, c_{i0}\right) = |\chi|\cdot\pi_{\xi}\left(\hat{\mu}_i, y_{i0} \mid c_{i0}\right) \quad \texttt{with} \quad |\chi| > 0 \tag{28}$$

where $\pi_{\xi}(\cdot)$ is the conditional density distribution of heterogeneous effects with sample size $|\chi|$. In this way, I can account for the whole cross-sectional information to obtain estimates of (non-)homogenous parameters θ (`first step`) and density $\pi_{\xi}(\cdot)$

[9] See, e.g., Arellano (2003), Arellano and Honore (2001), Arellano and Bover (1995), and Blundell and Bond (1998).

(`second step`). Here, I focus on only the `best` promising submodels achieved through the shrinking process, working with sufficiently high posterior consistency.

The FMM distributions and their moments (means and covariance matrices) are evaluated by maximizing the log-likelihood function via an expectation-maximization (EM) algorithm. More precisely, I suppose $\bar{m}$ regimes in which heterogeneous effects (ϕ_i) can vary in each submodel solution, where $\bar{m} = 0, 1, \dots, \varkappa$ is close to $|\chi|$, with 0 indicating the uninformative model where heterogeneous effects do not affect outcomes (e.g., DPD with fixed effects), and $\bar{m} \subset \mathcal{E}$. Then, I use Metropolis–Hastings algorithm[10] to draw posteriors for $\hat{\mu}_i$ from the (proposal) joint density distribution $\pi^{\bar{m}} = |\chi| \cdot \pi^*_{\xi}\left(\hat{\mu}_i^{\bar{m}}, y_{i0}^{\bar{m}} \mid c_{i0}^{\bar{m}}\right)$, with probability $\mathfrak{p}_{\bar{m}}$ equal to:

$$\mathfrak{p}_{\bar{m}} = \frac{\pi\left(\hat{\mu}_i^{\bar{m}}, y_{i0}^{\bar{m}} \mid \hat{\mu}_i^{\bar{m}-l}, \mathcal{Y}_i, \left\{\theta_t\right\}_{t=1}^{T}, c_{i0}^{\bar{m}}\right) \cdot \pi^{\bar{m}-l}}{\pi\left(\hat{\mu}_i^{\bar{m}-l}, y_{i0}^{\bar{m}} \mid \hat{\mu}_i^{\bar{m}-l}, \mathcal{Y}_i, \left\{\theta_t\right\}_{t=1}^{T}, c_{i0}^{\bar{m}}\right) \cdot \pi^{\bar{m}}} \tag{29}$$

where π^*_{ξ} stands for the conditional density distribution of heterogeneous effects involved in the final model solution (`third step`).

Let $|\chi|^*$ be the sample size according to the uninformative model in which neither (non-)homogeneous parameters nor unobserved effects achieve sufficient posterior consistency, and $\theta_t^* = \mathbf{1}^i$ be a vector of ones, the probability function takes the form:

$$\pi\left(\theta_t \mid \mathcal{Y}_i\right) \cdot \pi^*(\theta_t^* \mid \theta_t) \cdot \mathfrak{p}(\theta_t^*, \theta_t) = \pi\left(\theta_t^* \mid \mathcal{Y}_i\right) \cdot \pi^*(\theta_t \mid \theta_t^*) \tag{30}$$

where

$$\mathfrak{p}(\theta_t^*, \theta_t) = \min\left[\frac{\pi(\theta_t^* \mid \mathcal{Y}_i) \cdot \pi^*(\theta_t \mid \theta_t^*)}{\pi(\theta_t \mid \mathcal{Y}_i) \cdot \pi^*(\theta_t^* \mid \theta_t)}, 1\right] \cong \mathfrak{p}_{\bar{m}} \tag{31}$$

with $\mathfrak{p}(\theta_t^*, \theta_t)$ displaying the probability to accept or reject a draw[11] and $\pi^*(\cdot)$ denoting the density distribution according to sample size $|\chi|^*$. In this way, I obtain the same probability that each submodel M_{ξ} would be true. In addition, since posterior distributions correspond—by construction—to the FMM distributions, I define three possible intervals—displayed in (32)—in which the posterior predictive variance of $\mu_i\left(\mathbb{V}_{\theta,\phi,\pi(\bar{m})}^{(\mathcal{Y}_i,\mu_i)}[\mu_i]\right)$ can vary according to the model size ($|\chi|$). Thus, I am able to obtain exact posteriors on the predictive variance of the μ_i's, by taking into account both the model space and the parameter space. According to the shrinking process (Sect. 2.2), I ensure that lower variability will be associated with less relative regret during the estimating procedure, achieving more accurate forecasts.

[10] See, for instance, Levine and Casella (2014).

[11] See, for instance, Jacquier et al. (1994) and Pacifico (2021) for some applications to multicountry and multidimensional time-varying panel setups with stochastic and time-varying volatility, respectively.

$$\begin{cases} 0.5 < \underset{(high\ dimension)}{\mathbb{V}^{(\mathcal{Y}_i,\mu_i)}_{\theta,\phi,\pi(\bar{m})}}[\mu_i] \leq 1.0 & \text{with} \quad \xi^* > 10 \quad (\texttt{heterogeneity}) \\ 0.1 \leq \underset{(moderate\ dimension)}{\mathbb{V}^{(\mathcal{Y}_i,\mu_i)}_{\theta,\phi,\pi(\bar{m})}}[\mu_i] \leq 0.5 & \text{with} \quad 5 < \xi^* \leq 10 \quad (\texttt{sufficient}-\texttt{homogeneity}) \\ 0.0 \leq \underset{(small\ dimension)}{\mathbb{V}^{(\mathcal{Y}_i,\mu_i)}_{\theta,\phi,\pi(\bar{m})}}[\mu_i] < 0.1 & \text{with} \quad \xi^* \leq 5 \quad (\texttt{near}-\texttt{homogeneity}) \end{cases} \tag{32}$$

3.3 Ratio Optimality and Posterior Distributions

Ratio optimality is a necessary tool to be addressed for evaluating empirical forecast optimality. However, it tends to be very weak when dealing with large parameter spaces due to limited to restrictive classes of models. According to the multivarite ROB procedure involved in (2), I am able to work on a restricted set of submodels well specified in order to obtain better available forecast models. In this context, the optimal point forecasts' objective is predicting the outcomes (y_{it}) by minimizing the expected loss in (23). Methodologically, it means proving that the predictor $\hat{y}_{i,T+k}$ achieves ϑ_0-ratio-optimality uniformly for priors $\pi^{\bar{m}} \subset \mathcal{E}$, with $\vartheta_0 \geq 0$. Thus,

$$\limsup_{N\to\infty} \underset{\pi^{\bar{m}}\subset\mathcal{E}}{} \frac{\mathcal{R}\left(\hat{y}_{i,T+k}, \pi^{\bar{m}}\right) - \mathcal{R}^{op}\left(\pi^{\bar{m}}\right)}{\left\{ N\xi^* \cdot \mathbb{E}^{\mathcal{Y}_i,\mu_i}_{\theta,\phi,\pi^{\bar{m}}}\left(\mathbb{V}^{(\mathcal{Y}_i,\mu_i)}_{\theta,\phi,\pi^{\bar{m}}}[\mu_i]\right)\right\} + N(\xi^*)\vartheta_0} \leq 0 \tag{33}$$

In (33), some considerations are in order. (i) The predictor $\hat{y}_{i,T+k}$ in (20) is constructed by replacing θ with a consistent estimator $\hat{\theta}$ (estimated AR(λ) coefficients) and individual outcome values $\pi\left(\hat{\mu}_i(\theta), y_{i0}\right)$ in (27) with estimates based on the FMM distributions. (ii) Taking expectations over $y^{0:T}$ in (23), it follows that optimal point forecasts aim to work well on average rather than on a particular value (or single draw) of the outcomes. More precisely, the individual–specific forecasts do not consist of estimating the realization of the outcomes themselves, but rather a function of their predictive conditional distributions. (iii) The prediction accuracy of optimal forecasts can be assessed through the mean squared errors $\left(MSE(\hat{\theta}) = E_{\hat{\theta}}\left[\sum_{i=1}^{N}(\hat{y}_{i,T+k} - y_{i,T+k})^2\right]\right)$, computed as the average of the squared forecast errors for all observations assigned to the model class M_{ξ^*}. For high $\mathbb{V}^{(\mathcal{Y}_i,\mu_i)}_{\theta,\phi,\pi(\bar{m})}$ (e.g., with $\xi^* > 10$), the further $\hat{\mu}_i$'s will be in the tails of their distribution, resulting in larger MSEs. Conversely, the MSEs will be smaller for less $\mathbb{V}^{(\mathcal{Y}_i,\mu_i)}_{\theta,\phi,\pi(\bar{m})}$ (e.g., with $\xi^* \leq 5$) and moderate for quite high $\mathbb{V}^{(\mathcal{Y}_i,\mu_i)}_{\theta,\phi,\pi(\bar{m})}$ (e.g., with $5 \leq \xi^* \leq 10$). (iv) In a semiparametric context, whether model classes in $\mathcal{E}$ are high dimensional (e.g., highly large heterogeneity among subgroups), the expected loss in (23) is minimized as $N \to \infty$ and $\pi^{\bar{m}}$ will converge to a limit that is optimal. Indeed, the final model

solution will correspond to the oracle forecast of the prior (or correlated random effect distribution) that most favors the true model (Tweedie correction). For a sufficiently large sample size, the EB method will give a solution close to the Bayesian oracle, by exploiting information more efficiently than a fixed choice of the μ_i's (e.g., full Bayesian solutions).[12] (v) The ratio-optimality in (33) allows for the presence of estimated parameters in the sufficient statistic $\hat{\mu}_i$ and uniformity with respect to the correlated random effect density $\pi^{\bar{m}}$, allowed to have unbounded support.

For $\bar{m} > 0$, the resulting predictor is:

$$\hat{y}_{i,T+k} = \left[\hat{\mu}_i^{\bar{m}}(\theta) + \frac{\hat{\sigma}_u^2}{T} \cdot \frac{\partial}{\partial \hat{\mu}_i^{\bar{m}}(\theta)} ln\left(\hat{\mu}_i^{\bar{m}}(\theta), y_{i0}^{\bar{m}}\right)\right]^{\bar{m}} + \hat{\theta} y_{it} \tag{34}$$

with $\bar{m} < \infty$ according to all possible submodel solutions $M_{\xi^*} \subset \mathcal{E}$.

The posterior distributions for $\tilde{\omega}$ ($\ddot{\tilde{\omega}}$) are calculated by combining the prior with the (conditional) likelihood for the initial conditions of the data. The resulting function is then proportional to

$$L\left(y_i^{0:T}|\ddot{\tilde{\omega}}\right) \propto exp\left\{ -\frac{1}{2}\left[\sum_{t=1}^{T}\left(y_{it} - (c_{it}^{\bar{m}}|\dot{\chi})\hat{\theta}_t - \hat{\mu}_i^{\bar{m}}(\hat{\theta})\right)'\right] \cdot (\hat{\sigma}_u^2)^{-1} \cdot \left[\sum_{t=1}^{T}\left(y_{it} - (c_{it}^{\bar{m}}|\dot{\chi})\hat{\theta}_t - \hat{\mu}_i^{\bar{m}}(\hat{\theta})\right)\right]\right\} \tag{35}$$

where $y_i^{0:T} = (y_{i0}, y_{i1}, \ldots, y_{iT})$ denotes the data and $\ddot{\tilde{\omega}}$ refers to the unknowns whose joint distributions need to be found.

Despite the dramatic parameter reduction implicit in the shrinking process, the analytical computation of posterior distributions $(\ddot{\tilde{\omega}}|\hat{y}_{i,T+k})$ is unfeasible, where $\hat{y}_{i,T+k}$ denotes the expectations of outcomes associated with the infeasible oracle forecast to be estimated (Eq. 22). Thus, I include a variant of the Gibbs sampler approach—the Kalman-Filter technique—to analytically draw conditional posterior distributions of $(\theta_1, \theta_2, \ldots, \theta_T|\hat{y}_{i,T+k}, \ddot{\tilde{\omega}}_{-\theta_t})$, with $\ddot{\tilde{\omega}}_{-\theta_t}$ referring to the vector $\ddot{\tilde{\omega}}$ but excluding the parameter θ_t. Starting from $\bar{\theta}_{T|T}$ and $\bar{\rho}_{T|T}$, the marginal distributions of θ_t can be then computed by averaging over draws in the nuisance dimensions, and the Kalman filter backwards can be run to compute posterior distributions for $\ddot{\tilde{\omega}}$:

$$\theta_t|\theta_{t-l}, \hat{y}_{i,T+k}, \ddot{\tilde{\omega}}_{-\theta_t} \sim N\left(\ddot{\bar{\theta}}_{t|T+k}, \ddot{\bar{\rho}}_{t|T+k}\right) \tag{36}$$

where

$$\ddot{\bar{\theta}}_{t|T+k} = \left[\left(\ddot{\bar{\rho}}_{t|T+k}^{-1} \cdot \bar{\theta}\right) + \sum_{t=1}^{T}\left((c_{it}^{\bar{m}}|\dot{\chi})' \cdot (\hat{\sigma}_u^2)^{-1} \cdot (c_{it}^{\bar{m}}|\dot{\chi})\right)\hat{\theta}_t\right] \tag{37}$$

[12] See, for instance, George and Foster (2000) and Scott and Berger (2010).

$$\ddot{\bar{\rho}}_{t|T+k} = \left[I_h - \left(\bar{\rho} \cdot \ddot{\bar{\rho}}_{T+k|t}^{-1} \right) \right] \cdot \bar{\rho} \tag{38}$$

with

$$\hat{\theta}_t = \left[(c_{it}^{\bar{m}} | \dot{\chi})' \cdot (\hat{\sigma}_u^2)^{-1} \cdot (c_{it}^{\bar{m}} | \dot{\chi}) \right]^{-1} \cdot \left[(c_{it}^{\bar{m}} | \dot{\chi})' \cdot (\hat{\sigma}_u^2)^{-1} \cdot y_{it} \right] \tag{39}$$

The Eqs. (38) and (39) denote the variance-covariance matrix of the conditional distribution of $\ddot{\bar{\theta}}_{t|T+k}$ and the GMM estimator, respectively. By rearranging the terms, Eq. (37) can be rewritten as

$$\ddot{\bar{\theta}}_{t|T+k} = \left[\left(\ddot{\bar{\rho}}_{t|T+k}^{-1} \cdot \bar{\theta} \right) + \left(\sum_{t=1}^{T} (c_{it}^{\bar{m}} | \dot{\chi})' \cdot (\hat{\sigma}_u^2)^{-1} \cdot y_{it} \right) \right] \tag{40}$$

where $\ddot{\bar{\theta}}_{t|T+k}$ and $\ddot{\bar{\rho}}_{t|T+k}$ stand for the smoothed k-period-ahead forecasts of θ_t and of the variance-covariance matrix of the forecast error, respectively.

Generated a random trajectory for $\left\{ \theta_t \right\}_{t=1}^{T}$ from $N\left(\ddot{\bar{\theta}}_{T|T}, \ddot{\bar{\rho}}_{T|T} \right)$[13] in (36), the other posterior distributions can be defined as:

$$\pi(\hat{\mu}_i | \hat{y}_{i,T+k}, \hat{\theta}_t) \sim N\left(\tilde{\delta}_{\mu_i}, \tilde{\Psi}_{\mu_i} \right) \tag{41}$$

$$\pi(\hat{y}_{i0} | \hat{\mu}_i^{\bar{m}}) = N(0, \hat{\kappa}) \tag{42}$$

$$\pi(\dot{\chi}) = \tilde{w}_{|\chi|} \cdot \begin{pmatrix} \xi^* \\ |\chi| \end{pmatrix}^{-1} \tag{43}$$

$$\pi(\hat{\sigma}_u^2 | \hat{y}_{i,T+k}) = IG\left(\frac{\ddot{\bar{\omega}}}{2}, \frac{\tilde{\nu}}{2} \right) \tag{44}$$

Here, some considerations are in order.

In Eq. (41), $\tilde{\delta}_{\mu_i} \sim N\left(0, \bar{\zeta}\right)$ and $\tilde{\Psi}_{\mu_i} \sim IG\left(\bar{\varphi}/2, \bar{\varepsilon}/2\right)$, where $\bar{\zeta} = \zeta + (u_{it}' u_{it})$, $\bar{\varphi} = \varphi \cdot \dot{\chi}$, and $\bar{\varepsilon} = \varepsilon \cdot \dot{\chi}$, with (ζ, ε) denoting the arbitrary scale parameters (sufficiently small) and φ referring to the arbitrary degree of freedom (chosen to be close to zero). In this analysis, $\tilde{\Psi}_{\mu_i}$ is obtained by using the (proposal) joint posterior density ($\pi^{\bar{m}}$) sampled via EM algorithm, $(\zeta, \varepsilon) \cong 0.001$, and $\varphi \cong 0.1$.

In Eq. (42), $\hat{\kappa} = \kappa \cdot \mathbb{V}_{\theta,\phi,\pi(\bar{m})}^{(\mathcal{Y}_i,\mu_i)}[\mu_i]$, with κ and $\mathbb{V}_{\theta,\phi,\pi(\bar{m})}^{(\mathcal{Y}_i,\mu_i)}[\cdot]$ denoting the arbitrary scale parameter and the posterior predictive variance of μ_i, respectively. In this analysis, $\kappa \cong 1.0$ and $\mathbb{V}_{\theta,\phi,\pi(\bar{m})}^{(\mathcal{Y}_i,\mu_i)}[\mu_i]$ is obtained according to the sample size $|\chi|$ as described in (32).

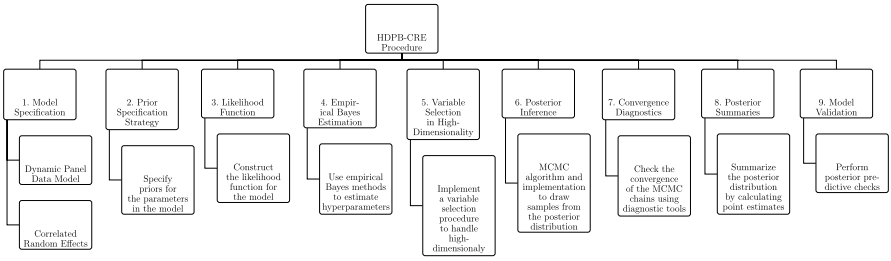
[13] See, for instance, Carro (2007).

In Eq. (43), $\tilde{w}_{|\chi|}$ refers to the model posterior choice according to the sum of the PMPs, with $\tilde{w}_{|\chi|} = max^*(NT, |\chi|)$ accounting for the non-0 components of $\dot{\chi}$.

In Eq. (44), $\ddot{\bar{\omega}} = \bar{\omega} + \hat{\bar{\omega}}$ and $\tilde{\nu} = \nu + \hat{\nu}$, with $\bar{\omega}$ and ν denoting the arbitrary degrees of freedom (sufficiently small) and the arbitrary scale parameter, respectively, $\hat{\bar{\omega}} = \left(\sum_{t=1}^{T} log(\tau_t)/t \right) + log\left(\sum_{t=1}^{T} (1/\tau_t) \right) - log(t)$ and $\hat{\nu} = (t \cdot \hat{\bar{\omega}})/\left(\sum_{t=1}^{T} (1/\tau_t) \right)$ referring to the Maximum Likelihood Estimates (MLEs). This latter is obtained by numerically computing $\hat{\bar{\omega}}$. In this analysis, $\tau_t = \{\tau_1, \ldots, \tau_T\}$ is the random sample from the data $\{0, T\}$, $\bar{\omega} \cong 0.1$, and $\nu \cong 0.001$.

Finally, the last two hyperparameters to be defined are $\bar{\theta} = \hat{\theta}_0$, with $\hat{\theta}_0$ denoting the GMM estimators of Eq. (2) related to the posteriors $\hat{y}_{i0}$ in (42), and $\bar{\rho} = I_{\xi^*}$.

The diagram below summarizes step-by-step the algorithm involved in the proposed computational approach.



4 Empirical Evidence

4.1 Data Description and Results

The HDPB-CRE in (2) contains 22 country-specific models, including 9 advanced economies[14], 7 emerging economies[15], and 6 non-European countries[16]. All advanced countries—except for *SV*—refer to Western Europe (WE) economies and all emerging countries—except for *GR*—refer to Central-Eastern Europe (CEE) economies, respectively. All European countries are Eurozone members, except for *CZ* and *PO*, allowing for an in-depth investigation of interdependencies and inter-country linkages. The estimation sample is expressed in years and covers the period from 1990 to 2021, with all data sourced from the World Bank database. Given the hierarchical structural conformation of the model and a sufficiently large number of years describing economic–financial and policy issues, it is able to investigate: (i)

[14] Austria (*AU*), Finland (*FI*), France (*FR*), Germany (*DE*), Ireland (*IR*), Italy (*IT*), Netherlands (*NL*), Slovenia (*SV*), and Spain (*ES*).

[15] Czech Republic (*CZ*), Poland (*PO*), Slovak Republic (*SV*), Estonia (*ES*), Latvia (*LV*), Lithuania (*LT*), and Greece (*GR*).

[16] United States (*US*), China (*CH*), Korea (*KO*), Japan (*JP*), United Kingdom (*GB*), and Chile (*CH*).

endogeneity issues; (ii) interdependency, commonality, and homogeneity; (iii) relevant monetary and fiscal policy interactions; and (iv) mis-specified dynamics.

The panel set contains 92 observable variables dealing with all potential determinants and policy tools described through the vectors $y_{i,t-l}$ and $z_{i,t-l}$. In this study, I split them into four groups: (*i*) `Economic status`, including 41 determinants combining information on education, income, economic development, and labor market; (*ii*) `Healthcare statistics`, addressing 11 determinants combining information on health coverage and expenditures on health; (*iii*) `Demographic and environment statistics`, accounting for 28 determinants combining information on population and sources of electricity; and (*iv*) `Economic–financial issues`, referring to 12 determinants dealing with real–financial economy and financial markets.

By running the shrinking process, in the `first step`, I find that 53 `best` covariates,[17] better fit the data with PIPs $\geq \tau$ and $\chi = 1$ (Table 1). Thus, I obtain 2^{53} submodel solutions ($M_j \subset \mathcal{S}$). Because of the curse of dimensionality, I further shrink the data performing the `second step` involved in the multivariate ROB procedure. Overall, 31 `best` promising covariates are found, obtaining 2^{31} submodel solutions ($M_\xi \subset \mathcal{E}$) with $\dot{\chi} = 1$. Here, some preliminary results can be addressed. (i) Most of the model uncertainty and overfitting are avoided. Indeed, dealing with the sign certainty, the Conditional Posterior Sign (CPS)[18] tends to be close to 0 (such as predictors 7, 25) and 1 (such as predictors 2, 9, 11, 26, 31). (ii) Socioeconomic factors matter more than economic status due to the ongoing outbreak of the epidemiology. (iii) The main policy tools correspond to some of the core variables of real and financial business cycles affecting the spreading and transmission of spillover effects (such as current account balance, gross fixed capital formation, credit, and inflation rate). (iv) The final solution includes 20 `best` promising predictors that better fit the data with PIPs $\geq \dot{\tau}$ (in bold in Table 1).

According to empirical evidence, a performance comparison with respect to the Stuctural Panel Bayesian VAR model with time-varying volatilities (SPBVAR-MTV) developed in Pacifico (2021) is addressed to emphasize the computional improvement of the proposed method. Firstly, the variable selection procedure is involved in the full panel set. Then, density forecasts on the results achieved are also performed.

By running the shrinking process, three main differences are in order (Fig. 1). (i) Even if only by a little, the performance accuracy in HDPB-CRE is higher than in SPBVAR-MTV, displaying a lower Posterior Model Size distribution. (ii) Overfitting tends to be larger. Indeed, 27 potential predictors have been selected with PIPs $\leq \dot{\tau}$ in the final step, obtaining 2^{27} submodel solutions. (iii) Uncertain effects matter more in SPBVAR-MTV because of higher PMPs and CPS values not strictly close to 0 and 1 for more than 10 predictors. For instance, focusing on plot (a), 8 on the

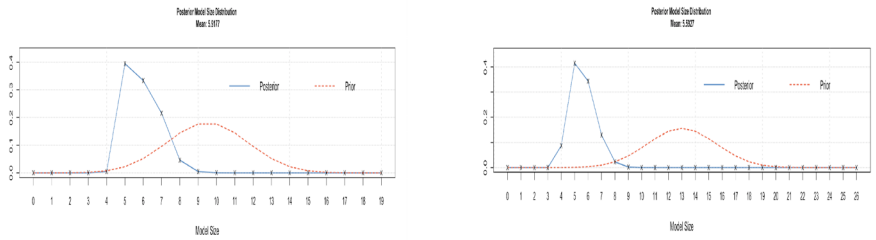
[17] More precisely, 19 predictors refer to `economic status` 8 predictors account for `healthcare statistics`, 16 predictors account for `demographic and environment statistics`, and 10 predictors refer to `Economic–financial issues`.

[18] The CPS takes values close to 1 or 0 if a covariate in c_{it} has a positive or negative effect on the outcomes, respectively.

Table 1 `Best` candidate predictors—`second stage`

Idx.	Predictor	Label	Unit	PIP (%)	CPS
Economic status					
1	Current education expenditure, secondary	edusec	Total exp. (%)	0.43	0.63
2	Employers, total	emto	Total emp. (%)	**46.75**	1.00
3	Employment to population ratio, 15+	empo	Total pop. (%)	0.17	0.33
4	Foreign direct investment, net inflows	fdinet	% GDP	**16.41**	0.96
5	Labor force partecipation rate, 15+	labpar	Total pop. (%)	0.22	0.27
6	Labor force, total	labtot	Logarithm (thousands)	**33.65**	0.68
7	Unemployment change	unem	Total labor force (%)	**65.51**	0.00
8	Wage and salaried workers	wage	Total emp. (%)	**27.40**	0.91
Healthcare statistics					
9	Capital health expenditure	cahe	% GDP	**31.56**	1.00
10	Current health expenditure	cuhe	% GDP	**44.02**	0.37
11	Dom. gen. gov. health expenditure	gghe	% GDP	**41.04**	0.95
12	Dom. gen. gov. health expenditure	hegg	% gen. gov. exp	**28.13**	1.00
13	Current tobacco use	tobuse	% adults (15+)	**17.37**	0.61
14	Alcohol consumption per capita	alcuse	logarithm (adults, 15+)	0.36	0.33
Demographic and environment statistics					
15	CO2 emissions, total	co2tot	total (%)	**23.06**	0.16
16	Age dependency ratio	arat	working-age pop. (%)	**48.12**	0.44
17	Fertility rate, total	frat	births per woman	**35.43**	0.10
18	Death race, crude	death	per 1,000 people	0.15	0.06
19	Energy imports, net	eneim	energy use (%)	**28.31**	0.71
20	Population, total	pop	logarithm (thousands)	0.23	0.47
21	Rural population	rural	total pop. (%)	0.18	0.35
22	Urban population	urban	total pop. (%)	**21.33**	0.51
23	School enrollment, secondary	school	total pop. (% net)	0.36	0.68
24	Human Capital Index	hci	working-age pop. [0–1]	0.32	0.81
Economic–financial issues					
25	Central government debt, total	debt	% GDP	**37.87**	0.00
26	Current account balance	cab	% GDP	**67.31**	1.00
27	Domestic credit, financial sector	crefin	% GDP	0.41	0.83
28	Gen. gov. final consumption exp	ggfce	% GDP	0.24	0.75
29	Gross fixed capital formation	gfcf	% GDP	**61.50**	0.92
30	Inflation, consumer prices	inf	% GDP	**63.24**	0.04
31	GDP growth per capita	gdpg	annual %	**74.45**	1.00
–	GDP per capita	gdp	PPP	-	-

The Table is organised as follows: the first column denotes the predictor number; the second and third columns describe the predictors and the corresponding labels, respectively; the fourth column refers to the measurement unit; and the last two columns display the PIPs (in %) and the CPS, respectively. The last row refers to the outcomes of interest. All contractions stand for: *exp*., 'expenditure'; *emp*., 'employment'; *pop*., 'population'; and *dom. gen. gov*., 'domestic general government'. All data refer to the World Bank database



(a) HDPB-CRE model (b) SPBVAR-MTV

Fig. 1 Posterior model size and probability distributions performing the proposed multivariate ROB procedure in HDPB-CRE (plot **a**) and SPBVAR-MTV (plot **b**) models

20 final predictors display higher divergences in terms of prior and posterior distributions: `labtot` (predictor 6); `cuhe` (predictor 10); `tobuse` (predictor 13); `co2tot` (predictor 15); `arat` (predictor 16); `frat` (predictor 17); `eneim` (predictor 19); `urban` (predictor 22).

Finally, the final model solution better performing the data—with $lBF = 13.49$—consists of 10 final `best` covariates included in $M_{\xi^*} \subset \mathcal{E}$ and so split: predictors 2, 7 for $y^{c'}_{i,t-l}$; predictors 10, 11, 16, 17 for $z^{s'}_{i,t-l}$; and predictors 26, 29, 30, 31 for $z^{p'}_{i,t-l}$. These results emphasize the performance of the HDPB-CRE procedure compared to multivariate dynamic models. Indeed, only 3 out of the 10 predictors need to be interpreted with care due to a CPS not strictly close to 0 or 1 (greater uncertainty). All their available lags, including lagged outcomes ($y^{o'}_{i,t-l}$), are then included as `external` instruments. The third step concludes by performing a reverse causality, finding high significance in both directions (Table 2).

In Table 3, the main diagnostic tests highlighting the performance of the HDPB-CRE model are displayed. In the estimating procedure, two time-invariant effects (x_{1t} and x_{2t}) are also included, denoting the presence of structural breaks in 2008 (due to the global financial crisis) and in 2020 (due to the COVID-19 pandemic). Here, some considerations are in order. (i) The best optimal lag chosen according to Arellano (2003)'s test is 3. (ii) All estimates are consistent and valid, showing no autocorrelation among residuals and highly strong linear dependencies; thus, variable selection problems are dealt with. (iii) The posterior predictive variance of the μ_i's reenters in the range displayed in (32), dealing with high dimensional data carefully ($\mathbb{V}^{\mathcal{Y}_i,\mu_i}_{\theta,\phi,\pi(\tilde{m})}[\mu_i] = 0.74$ with $\xi^* > 10$). (iv) In Table 2, highly strong causal links confirm the presence of heterogeneity across units. (v) The Posterior model size distribution (PMSD) is close to 10 and then to the `best` candidate predictors better explaining the data ($\dot{\chi}$). (vi) The estimating procedure is robust dealing, with most of the explained variability of the outcomes ($R^2_{adj.} = 0.78$).

Table 2 Granger (non-)causality test—`third step`

	emto	unem	cuhe	gghe	arat	frat	cab	gfcf	inf	gdpg
From c_{it}^{*} to y_{it}										
Z-tilde	5.40(0.00)	4.56(0.00)	5.85(0.00)	4.08(0.00)	5.17(0.00)	3.66(0.00)	5.39(0.00)	4.04(0.00)	3.42(0.00)	5.98(0.00)
From y_{it} to c_{it}^{*}										
Z-tilde	7.59(0.00)	3.45(0.00)	3.10(0.00)	2.50(0.01)	4.21(0.00)	3.23(0.00)	5.03(0.00)	2.48(0.01)	6.05(0.00)	3.44(0.00)

The Table displays all Z-tilde test statistics and p-values (in parentheses) based on the panel Granger (non-)causality test. Here, c_{it}^{*} stands for the `best` final candidate predictors in $M_{g_S^*} \subset \mathcal{E}$ with higher lBF (equation (12))

Table 3 Diagnostic tests

Main statistics	Results
$AR(l^*)$	3
ξ^*	10
LGB_s	0.00
LGB_r	0.91
$\mathbb{V}^{\mathcal{Y}_i,\mu_i}_{\theta,\phi,\pi(\bar{m})}[\mu_i]$	0.74
$PMSD$	9.92
lBF	13.49
$R^2_{adj.}$	0.78

Here, l^* denotes the optimal lag; LGB_s and LGB_r stand for Ljung–Box test statistics of series and residuals (p-values), respectively; PMSD refers to the posterior model size distribution; and $R^2_{adj.}$ denotes the adjusted R^2

4.2 Implications for Data Analysis and Country Sample

According to the results achieved in the empirical evidence, some main implications on the data and country samples are discussed and graphically displayed.

Firstly, by running a correlation analysis among the final 10 `best` predictors (Table 4), sufficiently high relationships matter without incurring in potential multicollinearity problems (correlation functions $\leq 60\%$), ensuring more reliable and stable parameter estimates. Thus, the methodology can be possibly involved to study further applications and decision-making processes dealing with endogeneity issues and model mis-specification problems.

In Table 5, the predictors are then estimated using the dynamic panel model described in (2). Every covariate is significant at least at 10%, fitting well the outcome of interest ($\bar{R}^2 = 79\%$). Furthermore, two additional estimates assuming individual– and time–specific effects have been developed, obtaining an adjusted R^2 close to 100% (94% and 96%, respectively). This emphasizes the need to deal

Table 4 Correlation analysis

Idx		emto	unem	cuhe	gghe	arat	frat	cab	gfcf	inf	gdpg
2	emto	1.00	0.48	0.39	0.54	−0.60	−0.31	0.33	−0.30	−0.26	−0.53
7	unem	0.48	1.00	−0.58	−0.44	0.21	0.15	−0.45	−0.42	0.47	−0.58
10	cuhe	0.39	−0.58	1.00	0.52	0.37	0.12	0.17	−0.46	−0.39	−0.42
11	gghe	0.54	−0.44	0.52	1.00	0.40	−0.26	0.22	−0.47	−0.30	−0.48
16	arat	−0.60	0.21	0.37	0.40	1.00	0.17	0.10	−0.50	0.12	−0.16
17	frat	−0.31	0.15	0.12	−0.26	0.17	1.00	−0.37	−0.07	0.17	0.11
26	cab	0.33	−0.45	0.17	0.22	0.10	−0.37	1.00	−0.10	−0.06	−0.17
29	gfcf	−0.30	−0.42	−0.46	−0.47	−0.50	−0.07	−0.10	1.00	−0.32	0.39
30	inf	−0.26	0.47	−0.39	−0.30	0.12	0.17	−0.06	−0.32	1.00	0.15
31	gdpg	−0.53	−0.58	−0.42	−0.48	−0.16	0.11	−0.17	0.39	0.15	1.00

Correlation matrix among the final `best` covariates included in $M_{\xi^*} \subset \mathcal{E}$

with correlated random effects when studying dynamic multidimensional data in high-dimensionality.

Based on economic theory and empirical evidence, the positive effects of `emto`, `gghe`, `cab`, `gfcf`, and `gdpg` on the outcome of interest are consistent and theoretically justified. Indeed, (i) the total number of employers is often correlated with business activity and economic dynamism; (ii) higher government health expenditure can improve the overall health of the population, leading to healthy workers being generally more productive and contributing more effectively to economic output; (iii) a positive current account balance (surplus) indicates that a country is exporting more than it is importing, which can be a sign of a strong economy;

Table 5 Estimation output in HDPB-CRE Model

Dependent variable: GDP per capita in PPP	
lag(gdp, 3)	0.773^{***}
	(0.038)
emto	0.025^{**}
	(0.012)
unem	-0.004^{*}
	(0.002)
cuhe	-0.064^{***}
	(0.020)
gghe	0.078^{***}
	(0.024)
arat	-0.004^{*}
	(0.002)
frat	-0.005^{***}
	(0.001)
cab	0.008^{***}
	(0.002)
gfcf	0.016^{***}
	(0.002)
inf	-0.017^{***}
	(0.001)
gdpg	0.018^{***}
	(0.002)
TSS	40.688
RSS	7.802
R^2	0.810
Adjusted R^2 ($\bar{R}^2$)	0.780
Chi-squared (p-value)	0.000

The estimation output obtained by running the HDPB-CRE model described in (2) is displayed, with standard errors in parentheses. Here, TSS and RSS stand for 'Total Sum of Squares' and 'Residual Sum of Squares', respectively. The significant codes are: $***$, significance at 1%; $**$, significance at 5%; and $*$, significance at 10%

investments in physical assets are crucial for enhancing productive capacity and fostering economic growth; and (iv) the positive relationship with GDP growth per capita is almost tautological and expected in dynamic panel regressions. Similarly, (i) a higher unemployment rate typically leads to lower overall production and economic output; (ii) while one might initially expect health expenditures to positively impact GDP per capita due to improved health and productivity, high current health expenditures can indicate inefficiencies or high costs that do not translate into proportional health improvements, mainly in the short term; (iii) a higher dependency ratio implies that a larger segment of the population is not working, reducing the amount of income available for investment and consumption by the working population; (iv) a higher total fertility rate can result in greater expenditures on child care and education, potentially at the expense of savings and investment, and slow economic growth by limiting the resources available for enhancing productivity; (v) the negative relationship between the inflation rate and GDP per capita is consistent with economic theory.

Finally, the framework being hierarchical, a clustering analysis is esaliy performed among the final predictors, where their PMPs are used as weights to compute the distance function and then build the membership values (Fig. 2). Setting the number of clusters to 4, the expected result would be to group each factor according to the four macro-groups specified in Table 1. Nevertheless, this does not occur since—by construction—the predictors are grouped based on their individual–specific characteristics. More precisely, clustering economic variables based on the countries' characteristics can reveal important insights into how these variables are related and how they jointly influence economic outcomes. Moreover, it allows for in-depth investigation of cross-country interdependency and heterogeneity and the performance of relevant policy strategies in a multidimensional context.

Based on economic theory and empirical evidence, the results achieved in Fig. 2 can be evaluated as follows. (i) The first cluster (plot *a*) groups `arat` and `gfcf` predictors. Their connection can be interpreted in terms of investment dynamics and resource allocation. In economies with a high age dependency ratio, a significant portion of resources might be directed towards consumption and social services to support dependents, potentially at the expense of investment in fixed capital. Conversely, lower dependency ratios can free up resources for higher levels of physical investments, supporting long-term economic growth. Thus, these two variables can be closely linked in how they shape economic outcomes. (ii) The second cluster (plot *b*) includes the most predictors: `gdpg`, `emto`; `cuhe`; `gghe`; and `cab`. These variables are interrelated through their collective impact on economic growth and stability. A growing economy (high GDP growth rate) typically exhibits a robust job market (total employers), balanced trade and financial interactions (current account balance), and adequate health investments. Health expenditures ensure a productive workforce, supporting sustainable growth. Hence, these variables form a cluster representing the dynamics of economic growth, employment, trade balance, and health investment. (iii) In the third cluster (plot *c*), the unemployment rate (`unem`) and total fertility (`frat`) rate are connected through demographic and labor market dynamics. High fertility rates can lead to a growing labor force, which, if not matched with sufficient employment opportunities, can result in high unemployment. Conversely,

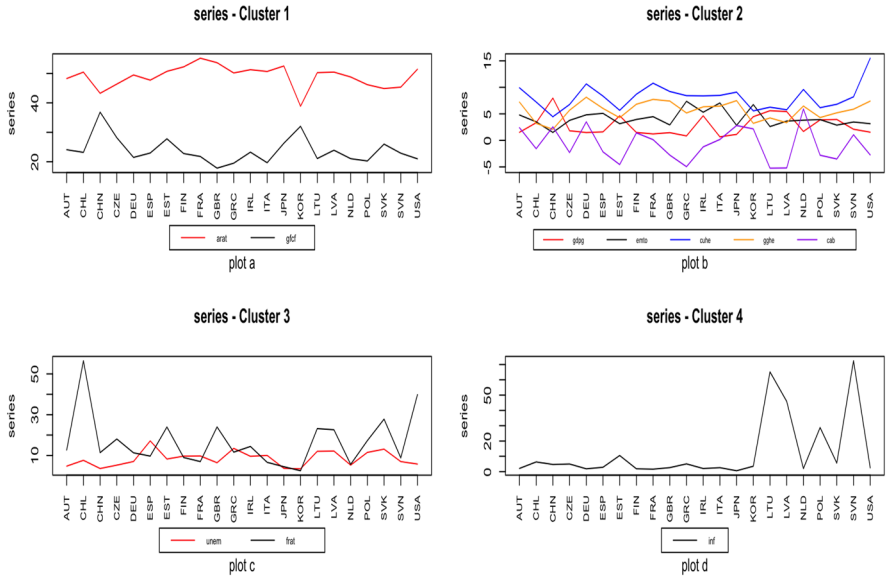


Fig. 2 The panel shows the clustering procedure performed on the HDPB-CRE estimates displayed in Table 5. The x- and y-axes refer to the country samples and series, respectively. The weighted distance measure and membership values have been computed using the PMPs as weights

high unemployment might discourage higher fertility rates due to economic uncertainty and reduced family income. These interactions show why these variables might cluster together, reflecting the interplay between demographic trends and labor market conditions. (iv) The distinct cluster for the inflation rate (`inf` in plot *d*) underscores its unique role in economic dynamics, influenced by and influencing a broad array of economic activities and policy decisions. Generally, because of its pervasive influence and the specific policy tools aimed at controlling it, inflation is often considered separately in economic analysis. This theoretical basis justifies why the inflation rate is grouped in a distinct cluster, reflecting its unique and broad-reaching impact on the economy (e.g., price stability, monetary policy, broad economic impact).

Baes on country sample implications, the high heterogeneity (variation) in `arat` and `gfcf` across countries suggests differences in demographic structures and investment levels (cluster 1, top-left plot). For instance, while some countries would prioritize infrastructure to boost productivity, others focus on healthcare and social services. The divergence (opposite dynamics) between age dependency ratios and gross fixed capital formation is due to several economic and demographic factors (e.g., economic growth and productivity, government spending, savings and investment rates). In cluster 2 (top-right plot), co-movements across countries (common trends) include a positive correlation between GDP growth and health expenditures, as wealthier nations can afford better healthcare. Similarly, countries with robust economies might exhibit better current account balances and more employers. Nevertheless, persistent heterogeneity matters due to varying economic policies,

healthcare systems, and trade practices. Cluster 3 (bottom-left plot) displays both common patterns, including developing countries with higher fertility rates and higher unemployment due to rapid population growth outstripping job creation, and heterogeneity, accounting for developed countries showing lower fertility and unemployment rates due to better economic conditions and family planning policies. In cluster 4 (bottom-right plot), including the inflation rate only, common trends matter among groups of countries with low inflation (e.g., due to stable economic policies) and those with high inflation (e.g., due to economic instability or external shocks). The latter displays wide variation in inflation rates, probably reflecting differences in economic policies, market conditions, and external shocks.

4.3 Forecasting and Policy Purposes

Concerning dynamic analyses, the total number of draws is $2000 + 3000 = 5000$, which corresponds to the sum of the final number of draws to discard and save, respectively. Convergence is achieved at approximately 1000 draws,[19] which are used to conduct posterior inference at each t.

The natural conjugate prior refers to three subsamples: x(i) 2007–2009 to evaluate the impact of the Great Recession; (ii) 2010–2018 to adress how fiscal consolidation periods affected the dynamics of the productivity among countries; and (iii) 2019–2023 to investigate the evolution of productivity in light of the ongoing pandemic crisis and the Russia-Ukraine war (predictors 15, 19). The time frame $2022 - 2023$ refers to outcomes incorporated in the forecasting analysis.

Without restrictions, the estimation sample comprises 726 regression parameters: each estimate of the HDPB-CRE in (2) account for 22 country indices and 33 time periods. Assuming hyperparameters in $\tilde{\omega}$ are all known and estimable, posterior distributions are computed according to Eqs. (36–38) for $\theta_t|\theta_{t-l}, \hat{y}_{i,T+k}$, (41–42) for moment distributions in $\mu_i|\hat{y}_{i,T+k}$ given initial values $(\hat{y}_{i0}|\hat{\mu}_i^{\bar{m}})$, (43) for the final `best` parameter space, and (44) for $u_t|\hat{y}_{i,T+k}$. All data are expressed in standard deviations.

In Fig. 3, conditional forecasts for outcomes $\hat{y}_{i,T+k}$ are shown for advanced (top plot) and emerging (bottom plot) economies. The yellow and red curves denote the 95% confidence bands, and the blue and purple curves represent the conditional[20] and unconditional[21] projections of outcomes $\hat{y}_{i,T+k}$ for each N country indexes and T time periods.

From a modelling perspective, three main findings are addressed. (i) Even if there has been evidence of significant co-movements and interdependencies among countries, consistent heterogeneities matter in both the spreading and the intensity of countries' dynamics. Thus, forecasters and policymakers need to account for

[19] Convergence is found by averaging about 1.2 draws per regression parameter.

[20] Generally, the conditional projection in forecasting models is the one that the model would have obtained over the same period conditionally on the actual path of unexpected dynamics for that period (μ_i dependent on y_{i0}).

[21] Generally, the unconditional projection in forecasting models is the one that the model would obtain for output growth for that period only on historical information, consistent with a model-based forecast path for the other variables (μ_i independent of y_{i0}).

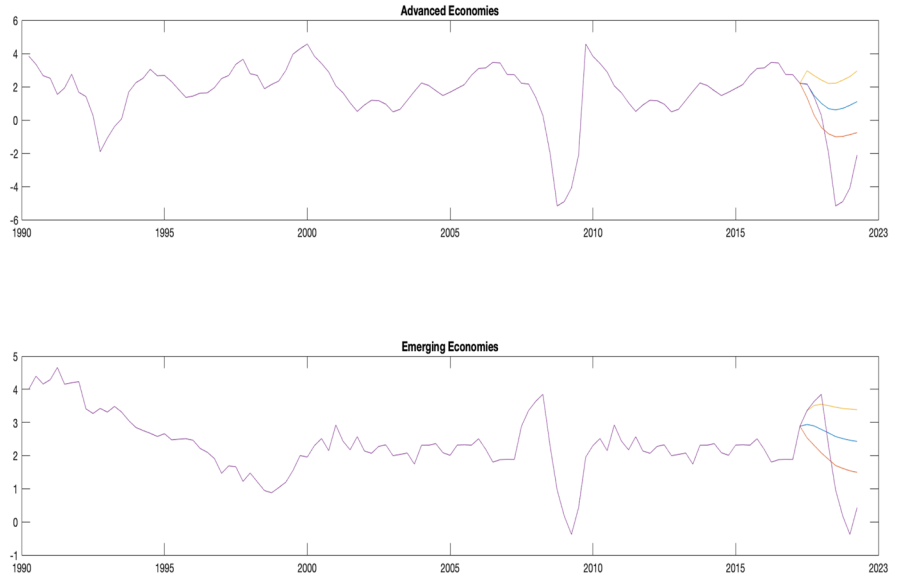


Fig. 3 The plot draws conditional forecasts for outcomes $\hat{y}_{i,T+k}$ given individual–specific (μ_i) and time-fixed (α) effects given a pool of socioeconomic–demographic, real–financial, and policy determinants. All time-varying parameters are posterior means

heterogeneous effects (correlated random coefficients) when formulating policy strategies and forecasting in multivariate dynamic panel data. (ii) Conditional projections lie within the confidence interval; conversely, unconditional projections tend to diverge over T. Therefore, when studying large time-varying panel data, cross-unit lagged interdependencies, dynamic feedback, and interactions must be assessed to address with endogeneity issues and mis-specified dynamics. (iii) A hint of boosting productivity to potential growth (2022–2023) can be observed among countries, particularly in advanced economies. Although recent dynamics suggest significant improvements in fiscal sustainability (e.g., during post-crisis periods), the risk of a cascade of policy errors, adverse political economy incentives, and divergence in financial integration become relevant issues for early and coordinated fiscal consolidation.

From a policy perspective, three main results are highlighted. (i) Empirical forecasts show that most European emerging economies are strongly exposed to financial interlinkages and are highly dependent on other European countries (e.g., Western European countries). However, the presence of persistent heterogeneities among countries' responses emphasize the need to accelerate financial development in developing countries, stimulate domestic resource mobilization, and support consistent reforms of the international financial system to boost investment and growth. (ii) Even though several measures have already been taken at the international and European Union levels, most countries have been limited in effectively using monetary and fiscal tools due to stringent economic–institutional interdependencies, preventing them from deploying conventional consolidation measures during triggering

events. Moreover, most countries have failed to control the extent of COVID-19 due to people's attitudes of denial and misunderstanding of social distancing for outbreak control. Therefore, in a context of radical uncertainty and heterogeneous territorial effects, appropriate policy measures need to be addressed locally rather than globally. (iii) Heterogeneity among countries' responses is significant because of different policy adjustments applied by governments during a recession. Indeed, governments have tended to follow distinct national rather than consensual international standards (as seen in the current outbreak and previous pandemic crises). Overall, policy tools should be implemented by closely monitoring the economic status of each country. More coordinated country-specific European and international measures, along with a participatory government approach, are needed to ensure robust health systems and more resilient economic development to safeguard against sudden outbreaks in the global economy.

5 Relative Regrets for Tweedie Correction: MCMC-Based Experiments

In this example, the performance of the estimation method is investigated by summarizing the regrets for the Tweedie correction in (28) relative to the posterior predictive variance of the μ_i's. More precisely, according to (32), I consider three sequences of $\left(N, \xi^*, \mathbb{V}^{y_i,\mu_i}_{\theta,\phi,\pi(\bar{m})}[\mu_i]\right)$ with correlated random coefficients in the homoskedastic case to evaluate different improvements in forecasting performance: (i) (10000, 15, 1.0), heterogeneity with high dimension; (ii) (10000, 10, 0.5), sufficient-homogeneity with moderate dimension; (iii) (10000, 5, 0.0), near-homogeneity with small dimension. I suppose a basic HDPD-CRE model with $\alpha = \gamma = 0$, homoskedastic variance $\sigma^2 = 1$, and regimes $\bar{m} = 1$ (e.g., a unique common individual–specific effect across units).

I include two additional empirical Bayes estimators dealing with alternative Tweedie corrections (Table 6): the Kernel Density (KD) estimator (see, for instance, Liu et al., 2020) and the NonParametric Maximum Likelihood (NPML) estimation (see, for instance, Gu and Koenker, 2017b). Concerning the former, the problem of forecasting a collection of short time-series processes is addressed using cross-sectional information in a dynamic panel data. A nonparametric kernel estimate of the Tweedie correction is then constructed, showing its asymptotic equivalence to the risk of an empirical predictor treating the CREs' distribution as known. As regards

Table 6 MCMC-based designs

Law of motion	$y_{it} = \mu_i + \beta y_{i,t-1} + u_{it}$ where $\beta = 0.5$, $u_{it} \sim i.i.d.N(0,1)$
Initial observations	$y_{i0} \sim N(0,1)$
Correlated random effects	$\mu_i \vert y_{i0} \sim N(0, \Psi_{\mu_i})$ where $\Psi_{\mu_i} \sim IG\left(\frac{0.1}{2}, \frac{0.01}{2}\right)$

The Table shows the three sequences of $\left(N, \xi^*, \mathbb{V}^{y_i,\mu_i}_{\theta,\phi,\pi(\bar{m})}[\mu_i]\right)$ with correlated random coefficients homoskedastic case conducted in the simulated example according to (32)

the NPML estimation, the EB estimators are constructed by specifying appropriate bounds for the domain of CREs and then partitioning them into a predetermined set of bins.

Table 7 provides the relative regrets for Tweedie corrections according to the three proposed MCMC-based designs. The best choice of ϑ_0 improving forecasting performance in terms of ratio-optimality was set at 0.5 (middle point in an arbitrary range [0.1–0.9]). The findings underscore the effectiveness of the multivariate ROB procedure in significantly reducing model size when dealing with high-dimensional data. Additionally, the use of semiparametric Bayesian statistics with FMM distributions (Tweedie correction) leads to more accurate forecasts. Notably, lower posterior predictive variances of the μ_i values correspond to reduced relative regret. When compared to KD and NPLM estimates, FMM distributions consistently exhibit lower regret. In experiments with a much larger sample size (e.g., $N = 100,000$) and lower ratio-optimality (e.g., $\vartheta_0 \cong 0.1$), the results show that relative regret decreases as the number of cross-sectional units N increases. However, although lower ratio-optimality reduces computational costs, it is associated with higher regret.

Summarizing, the study demonstrates that the multivariate ROB procedure is efficient in shrinking model size for high-dimensional data. In the simulated experiment, the use of semiparametric Bayesian statistics with FMM distributions enhances forecast precision, leading to lower posterior predictive variances and reduced relative regret. Computationally, experiments with larger sample sizes indicate that increasing the number of cross-sectional units N decreases regret. However, while lower ratio-optimality reduces computational costs, it results in higher associated regret. The main implication of the trade-off highlighted is the need for a balanced approach in model development and application. While computational efficiency is important, it should not come at the expense of significantly increased regret, which could undermine the model's utility and reliability. For instance, decision-makers should carefully evaluate the acceptable levels of computational cost and model performance to ensure that the models serve their intended purpose effectively.

All results in Table 7 find confirmation in Fig. 4. More precisely, lower posterior predictive variances of the μ_i's are associated to less mean squared errors (MSEs)

Table 7 Relative regrets for Tweedie corrections by MCMC-based designs

	Design I	Design II	Design III
N	10000	10000	10000
$\mathbb{V}_{\theta,\phi,\pi^{\vartheta_0}}^{\mathcal{Y}_i,\mu_i}[\mu_i]$	1.0	0.5	0.0
ξ^*	15	10	5
N_{sim}	10000	10000	10000
KD	0.026	0.051	0.074
FMM	0.014	0.010	0.007
NPML	0.021	0.019	0.013

Relative regrets for Tweedie corrections according to the three supposed MCMC-based designs. The regret is standardized by the average posterior predictive variance of the μ_i's, with $\vartheta_0 = 0.5$

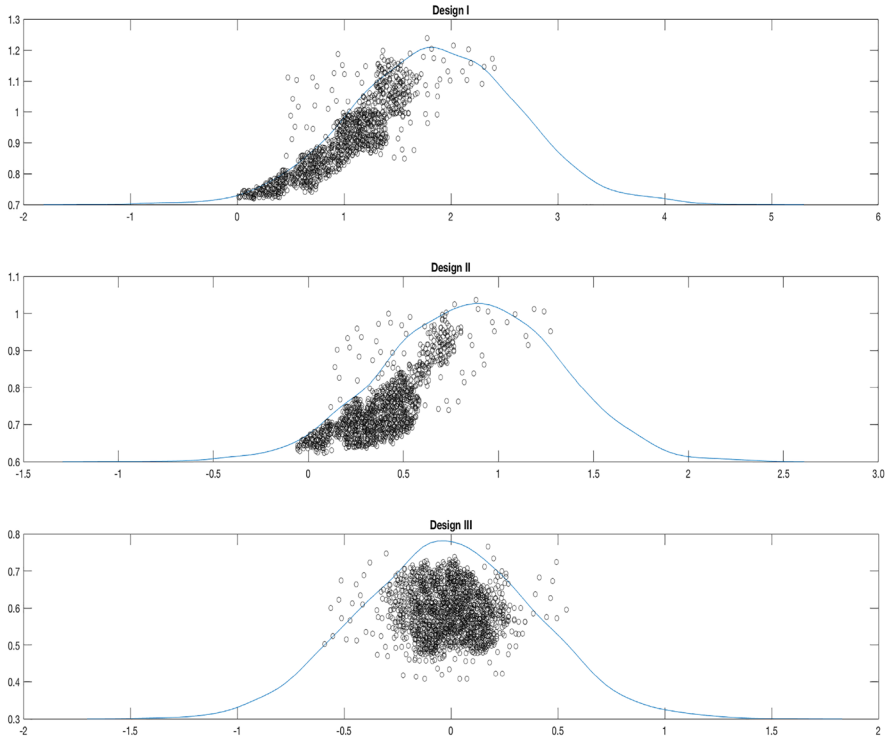


Fig. 4 The panels show the MSEs associated to the three supposed MCMC-based designs. The solid lines display the posterior draw samples of the empirical distribution of $\hat{\mu}_i$ according to the (designed) joint density distribution $\pi^{\bar{m}}$ and the FMM-based Tweedie correction

and then better accuracy forecasts (associated with less relative regrets). Moreover, the (designed) joint density distribution of $\pi^{\bar{m}}$—depicting posterior draw samples of the empirical distribution of $\hat{\mu}_i$—asymptotically converges to a Normal distribution, and then the FMM-based Tweedie correction— in Theorem (1.6)—approaches a linear distribution function. Furthermore, in the second and third designs, the empirical realizations of $\hat{\mu}_i$ are greater and lie in the distribution, highlighting lower MSEs and less sampling variance in the estimated posterior means.

6 Conclusion

This study aims to construct and develop a methodology to improve the recent literature on DPD models when dealing with (i) individual–specific forecasts, (ii) ratio-optimality and posterior consistency in dynamic panel setups, (iii) empirical Bayes approaches and alternative Tweedie corrections for nonparametric priors, and (iv)

the curse of dimensionality and variable selection problems when estimating time-varying data.

The contributions of this study are threefold. First, a multivariate shrinking procedure is used to select the `best` promising subset of covariates according to their Posterior Model Probability, which denotes the probability to better explain and thus fit the data in high-dimensional model classes. Second, the correlated random effects are addressed by involving in the shrinking process an empirical Bayes procedure, where the posterior mean of the unobserved heterogeneity is expressed in terms of the marginal distribution of sufficient statistics estimated from the cross-sectional whole information (Tweedie's formula). Third, better conditional forecasts can be involved in the estimation model because of the use of a semiparametric Bayesian approach modelling either time-varying and fixed effects, and the observation of incidental parameters possibly correlated with some of the predictors within the system.

An empirical application on a pool of advanced and emerging economies is assessed, describing the functioning and the performance of the methodology. The estimation sample refers to the period 1990–2021, covering a sufficiently large sample to address potential causal links and interdependencies between outcomes and a set of time-varying factors, including heterogeneous individual–specific and time-fixed effects. A simulated experiment using MCMC-based designs is also addressed to highlight the performance of the estimating procedure in comparison with related works.

Funding No fund was used to conduct this study.

Data availability Data available on request.

Declarations

Conflict of interest The author declares no Conflict of interest.

Ethical 'Not applicable'.

Consent to participate 'Not applicable'.

Code availability Code available on request.

References

Alvarez, J., & Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica, 71*(4), 1121–1159.

Anderson, T. W., & Hsiao, C. (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association, 76*(375), 598–606.

Arellano, M. (2003). *Panel data econometrics*. Oxford University Press.

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies, 58*(2), 277–297.

Arellano, M., & Bonhomme, S. (2011). Nonlinear panel data analysis. *Annual Review of Economics, 3*, 395–424.

Arellano, M., & Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics, 68*(1), 29–51.

Arellano, M., & Hahn, J. (2007). Understanding bias in nonlinear panel models: Some recent developments. In W. N. R. Blundell & T. Persson (Eds.), *Advances in Economics and Econonometrics: Theory and Applications, Ninth World Congress.* Cambridge: Cambridge University Press.

Arellano, M., & Hahn, J. (2016). A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects. *Global Economic Review, 45*(3), 251–274.

Arellano, M., & Honore, B. (2001). Panel data models: Some recent developments. In Heckman, J. & Leamer E. (Eds.) *Handbook of Econometrics* (Vol. 5, pp. 3229–3296).

Bester, C. A., & Hansen, C. (2009). A penalty function approach to bias reduction in non-linear panel models with fixed effects. *Journal of Business and Economic Statistics, 27*(2), 131–148.

Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics, 87*(1), 115–143.

Brown, L. D., & Greenshtein, E. (2009). Nonparametric empirical Bayes and compound decision approaches to estimation of a high-dimensional vector of normal means. *The Annals of Statistics, 37*, 1685–1704.

Carro, J. (2007). Estimating dynamic panel data discrete choice models with fixed effects. *Journal of Econometrics, 140*(2), 503–528.

Chamberlain, G. (1984). Panel data. In Griliches, Z. & Intriligator M. D. (Eds.) *Handbook of Econometrics* (Vol. 2, pp. 3847–4605).

Chamberlain, G. (2010). Binary response models for panel data: Identification and information. *Econometrica, 78*, 159–168.

Chamberlain, G., & Hirano, K. (1999). Predictive distributions based on longitudinal earnings data. *Annales d'Economie et de Statistique, 55–56*, 211–242.

Dumitrescu, E., & Hurlin, C. (2012). Testing for granger non-causality in heterogeneous panels. *Economic Modelling, 29*(4), 1450–1460.

Fernandez-Val, I. (2009). Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics, 150*(1), 71–85.

Forni, M., Hallin, M., Lippi, M. & Reichlin, L. (2000). The generalized dynamic factor model: Identification and estimation. *The Review of Economics and Statistics, 82*(4), 540–554.

Gelman, A., & Hill, J. (2012). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790942

George, E. I., & Foster, D. P. (2000). Calibration and empirical Bayes variable selection. *Biometrika, 87*, 731–747.

Gu, J., & Koenker, R. (2017). Empirical Bayesball remixed: Empirical Bayes methods for longitudinal data. *Journal of Applied Economics, 32*(3), 575–599.

Gu, J., & Koenker, R. (2017). Unobserved heterogeneity in income dynamics: An empirical Bayes methods for longitudinal data. *Journal of Business and Economic Statistics, 35*(1), 1–16.

Hahn, J., & Kuersteiner, G. (2011). Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory, 72*, 1295–1319.

Hirano, K. (2002). Semiparametric Bayesian inference in autoregressive panel data models. *Econometrica, 70*(2), 781–799.

Jacquier, E., Polson, N., & Rossi, P. (1994). Bayesian analysis of stochastic volatility. *Journal of Business and Economic Statistics, 12*, 371–417.

Jiang, W., & Zhang, C.-I.H. (2009). General maximum likelihood empirical Bayes estimation of normal means. *The Annals of Statistics, 37*(4), 1647–1684.

Kiefer, J., & Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *The Annals of Mathematical Statistics, 27*(4), 887–906.

Lancaster, T. (2002). Orthogonal parameters and panel data. *The Review of Economic Studies, 69*(3), 647–666.

Levine, R. A., & Casella, G. (2014). Implementations of the Monte Carlo EM algorithm. *Journal of Computational and Graphical Statistics, 10*(3), 422–439.

Liu, L. (2018). Density forecasts in panel data models: A semiparametric Bayesian perspective. *Working Paper*. Cornell University, pp. 1–32. Available at: arXiv:1805.04178

Liu, L., Moon, H.R., & Schorfheide, F. (2019). Forecasting with a panel Tobit model. *NBER Working Papers*, *26569*

Liu, L., Moon, H. R., & Schorfheide, F. (2020). Forecasting with dynamic panel data models. *Econometrica, 88*(1), 171–201.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica, 49*(6), 1417–1426.

Norets, A., & Pelenis, J. (2012). Bayesian modeling of joint and conditional distributions. *Journal of Econometrics, 168*, 332–346.

Pacifico, A. (2020). Robust open Bayesian analysis: Overfitting, model uncertainty, and endogeneity issues in multiple regression models. *Econometric Reviews, 40*(2), 148–176. https://doi.org/10.1080/07474938.2020.1770996

Pacifico, A. (2021). Structural panel Bayesian VAR with multivariate time-varying volatility to jointly deal with structural changes, policy regime shifts, and endogeneity issues. *Econometrics, 9*(2), 1–36. https://doi.org/10.3390/econometrics9020020

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *The Annals of Mathematical Statistics, 35*, 1–20.

Robert, C. (1994). *The Bayesian choice*. Springer.

Scott, J. G., & Berger, J. O. (2010). Bayes and empirical-Bayes multiplicity adjustment in problem. *The Annals of Statistics, 38*, 2587–2619.