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

Pacifico, Antonio

University of Macerata, LUISS Guido Carli University, University of Bruxelles

November 2022

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

Antonio Pacifico*

Abstract

A novel for multivariate dynamic panel data analysis with correlated random effects is proposed when estimating high dimensional parameter spaces. A semiparametric hierarchical Bayesian strategy is used to jointly deal with incidental parameters, endogeneity issues, and model misspecification problems. The underlying methodology involves addressing 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 to a large panel of economies is conducted 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.

*Corresponding author: Antonio Pacifico, Department of Economics and Law, University of Macerata (Marche), Italy. Email: antonio.pacifico@unimc.it. ORCID: https://orcid.org/0000-0003-0163-4956

1 Introduction

Dynamic Panel Data (DPD) models are widely used in empirical economics for forecasting individuals' future outcomes (see, e.g., Hirano (2002), Gu and Koenker (2017b), Liu (2018), and Liu et al. (2020)) and allowing the possibility of controlling for unobserved time-invariant individual heterogeneity (see, e.g., Chamberlain (1984) and Arellano and Bond (1991) (linear case); and Chamberlain (2010) and Arellano and Bonhomme (2011) (non-linear case)). Such heterogeneity is an important issue and failure to control for it results in misleading inferences. That 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 $y_{i,t-1}$ denote the outcomes and their first lags, μ_i refers to individual-specific intercept with $w_{i,t-1} = 1$, and $u_{it} \sim N(0, \sigma^2)$ is an independent and identically distributed ($i.i.d.$) shock.

In the dynamic panel literature, the focus is to find a consistent estimate of β in the presence of the incidental parameters μ_i to avoid the incidental parameters problem and then perform better forecasts of the outcomes in period $T+1$ (y_{T+1}). In the context of panel data, the incidental parameters problem typically arises from the presence of individual–specific factors. The challenges because of incidental parameters are highly severe in dynamic panels where behavioural effects over time are jointly measured with individual–specific effects. Whereas the incidental parameters to be estimated are consistent in least squares methods, maximum likelihood estimation leads to inconsistent estimates of them affecting the dynamics of data (see, for instance, Nickell (1981)). Both fixed and random effects have been used to evaluate these individual–specific factors. The former treats them as parameters to be estimated, leaving the distribution of unobserved heterogeneity relatively unrestricted at the cost of introducing a large number of nuisance parameters; random effects typically assume that their distributions belong to a known parametric family indexed by a finite dimensional parameter. Closely related studies addressing similar deconvolution problem and estimates of the μ_i's distribution are Anderson and Hsiao (1981), Arellano and Bond (1991), Arellano and Bover (1995), Blundell and Bond (1998), and Alvarez and Arellano (2003) (Instrumental Variables (IV) and Generalized Method of Moments (GMM) estimators); Hahn and Newey (2004), Carro (2007), Arellano and Hahn (2007, 2016), Bester and Hansen (2009), Fernandez-Val (2009), and Hahn and Kuersteiner (2011) (fixed effects approach in non-linear panel data); and Compiani and Kitamura (2016) (mixture models-based approach).

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. (2019, 2020) (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. It would compromise 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.

The methodology proposed in this study aims to overtake the aforementioned issues by developing a structural semiparametric hierarchical Bayesian approach to conduct inference in high dimensional dynamic panel data with cross-sectional heterogeneity, where 'structural' stands for designing a more conventional empirical procedure to provide reduced-form causal relationships. The model takes the name of Hierarchical Dynamic Panel Bayesian model with Correlated Random Effects (HDPB-CRE) and is achieved by combining an implemented version of the 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. 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 contributions of this paper are threefold. First, let the framework be hierarchical, multivariate Conjugate Informative Proper Mixture (mvCIPM) priors are used to select the `best` promising subset of

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

covariates according to their Posterior Model Probability (PMP), which denotes the probability to better explain and thus fit the data in high dimensional model classes. Here, `best` stands for the model providing the most accurate predictive performance over all candidate models. 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 misspecification), 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 and Hurlin (2012)).

Second, to accomodate the correlated random coefficients model, I involve in the previous shrinking process an Empirical Bayes (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 and Bond (1991)).

Third, better conditional forecasts are involved in HDPB-CRE because of three main features: `(i)` the use of a semiparametric Bayesian approach modelling 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 the 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

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

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 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 $N \cdot 1$ vector of outcomes, $y_{i,t-l}$ and $z_{i,t-l}$ are $N \cdot 1$ vectors of predetermined and directly observed (endogenous) variables 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 $\tilde{l} = 1, \ldots, \lambda$, $x_{i,t}$ is a $N \cdot 1$ vector of strictly exogenous factors for each i, with 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 $N \cdot 1$ 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 AutoRegressive (AR) processes involved in the model (2) are stationaries, 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) and Arellano and 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 in order 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 and Hsiao (1981), Arellano and Honore (2001), and Blundell and 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 from the shock u_{it}, leading in inaccurate forecasts. Last but not least, when dealing with time-varying and high dimensional data, variable selection problems such as overshrinkage/undershrinkage, model misspecification problems, endogeneity issues, and model uncertainty[3] also matter in DPD models involving inconsistent estimates. The multivariate ROB procedure involved in this study arises from 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 having statistically significant predictive capability (accurate forecast). Here, nonparametric Bayesian statistics are also addressed through MCMC implementations in order 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 in order to deal with overestimation of effect sizes (or

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

individual contributions). I recall that, 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 the Pacifico (2020)'s analysis.

Given the HDPB-CRE in (2), I decompose the vectors of the observed endogenous variables: $y_{i,t-l} = \left[y^{o'}_{i,t-l}, y^{c'}_{i,t-l}\right]'$, with $y^{o'}_{i,t-l}$ denoting lagged outcomes to capture the persistence and $y^{c'}_{i,t-l}$ including lagged control variables such as general economic conditions; and $z_{i,t-l} = \left[z^{s'}_{i,t-l}, z^{p'}_{i,t-l}\right]'$, referring to other lagged factors such as socioeconomic conditions ($z^{s'}_{i,t-l}$) and policy implications ($z^{p'}_{i,t-l}$). Then, I combine the (non-)homogeneous parameters into the vector $\theta = \left(\beta^{o'}_l, \beta^{c'}_l, \alpha', \gamma^{s'}_l, \gamma^{p'}_l\right)'$.

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:

$$c_{it} = \left(y^o_{i,0:t-l}, y^c_{i,0:t-l}, z^s_{i,0:t-l}, z^p_{i,0:t-l}, x_{i,0:t}\right) \tag{6}$$

and

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

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 $\phi_i = \left(\mu_i, \sigma^2_u\right)$ under cross-sectional homoskedasticity. Assuming correlated random coefficients model, ϕ_i and c_{i0} could be correlated with each other, with:

$$c_{i0} = \left(y^o_{i,0}, y^c_{i,0}, z^s_{i,0}, z^p_{i,0}, x_{i,0:T}\right) \tag{8}$$

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:

$$\pi\Big(y|\theta_h\Big) = \int_{\mathcal{B}} \pi\Big(y, \mu_i|\theta_h, M_h\Big) \cdot d\mu \tag{9}$$

where $\mathcal{B}$ denotes the multidimensional (natural) parameter space for θ_h, $M_h = (M_1, \ldots, M_m)$ denotes a countable collection of all (potential) model solutions given the data. The integrand in (9) is defined as:

$$\int_{\mathcal{B}} \pi\Big(y, \mu_i|\theta_h, M_h\Big) = \pi\Big(\theta_h, \mu_i, M_h|y\Big) \cdot \pi\Big(y|M_h\Big) \tag{10}$$

where $\pi(\theta_h, \mu_i, M_h|y)$ denotes the joint likelihood and $\pi(y|M_h) = \int \pi(y|M_h, \theta_h, \mu_i) \cdot \pi(\theta_h, \mu_i|M_h) d\theta_h$ is the marginal likelihood, with $\pi(\theta_h, \mu_i|M_h)$ referring to the conditional prior distribution of θ_h and μ_i. 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:

$$\mathcal{S} = \Big\{M_j : M_j \subset \mathcal{S}, \mathcal{S} \in \mathcal{F}, \Theta_j \subset \Theta_h, \sum_{j=1}^{\varpi} \pi\Big(M_j|y_i = y_i, \chi\Big) \geq \tau\Big\} \tag{11}$$

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 an enough posterior consistency[4]. In this study, I use $\tau = 0.5\%$ with N high dimensional (predictors ≥ 15). In this study, I am able to jointly manage all equations within the system (through the conditioning set c_{it}), their (potential) interactions (through AR coefficients), and their possible causal links (through Granger (Non-)Causality test).

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

$$\mathcal{E} = \Big\{M_\xi : M_\xi \subset \mathcal{E}, \mathcal{E} \in \mathcal{S}, \sum_{j=1}^{\varpi} \pi\Big(M_j|y_i = y_i, \dot{\chi}\Big) \geq \dot{\tau}\Big\} \tag{12}$$

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 a 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

[4]In Bayesian analysis, posterior concistency ensures that the posterior probability (PMP) concentrates on the true model.

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 misspecified 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):

$$lBF_{\xi^*,\xi} = log\left\{\frac{\pi(M_{\xi^*}|y_i = y_i)}{\pi(M_{\xi}|y_i = y_i)}\right\} \tag{13}$$

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

$$\begin{cases} 0.00 \leq lB_{\xi^*,\xi} \leq 4.99 & \text{no evidence for submodel } M_{\xi^*} \\ 5.00 \leq lB_{\xi^*,\xi} \leq 9.99 & \text{moderate evidence for submodel } M_{\xi^*} \\ 10.00 \leq lB_{\xi^*,\xi} \leq 14.99 & \text{strong evidence for submodel } M_{\xi^*} \\ lB_{\xi^*,\xi} \geq 15.00 & \text{very strong evidence for submodel } M_{\xi^*} \end{cases} \tag{14}$$

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 multivariate Conjugate Informative Proper Mixture priors:

$$\pi(\theta, \phi, \chi) = \pi(\theta|\chi) \cdot \pi(\mu_i|\chi, y_{i0}) \cdot \pi(\sigma_u^2|\chi) \cdot \pi(\chi) \tag{15}$$

where

$$\pi\Big(\theta|\mathfrak{F}_{-1}\Big) = N\Big(\bar{\theta}, \bar{\rho}\Big) \tag{16}$$

$$\pi(\mu_i|\theta) = N\Big(\delta_{\mu_i}, \Psi_{\mu_i}\Big) \quad \texttt{with} \quad \delta_{\mu_i} \sim N\Big(0, \zeta\Big) \quad \texttt{and} \quad \Psi_{\mu_i} \sim IG\Big(\frac{\varphi}{2}, \frac{\varepsilon}{2}\Big) \tag{17}$$

$$\pi(y_{i0}|\mu_i) = N(0, \kappa) \tag{18}$$

$$\pi(\chi) = w_{|\chi|} \cdot \binom{h}{|\chi|}^{-1} \tag{19}$$

$$\pi(\sigma_u^2) = IG\left(\frac{\bar{\omega}}{2}, \frac{\nu}{2}\right) \tag{20}$$

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 (18) refers to the decay factor, and $w_{|\chi|}$ in (19) 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 $\tilde{\omega} = \Big(\bar{\theta}, \bar{\rho}, \zeta, \varphi, \varepsilon, \kappa, w_{|\chi|}, \bar{\omega}, \nu\Big)$, 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 $max\Big(NT, |\chi|\Big)$ according to the non-0 components of χ.

Nevertheless, 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[5], whereas a fully Bayesian method constructs a prior for the correlated random effects and then evaluates it in view of the observed panel data[6]. Even if the fully Bayesian approach tends to be more suitable for density forecasting and more easily extended to 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 nonparametric 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 (16) - (20), I define the compound risk and loss functions – under which the forecasts will be evaluated – accounting for expectations over the observed trajectories $\mathcal{Y}_i = \Big(y_1^{0:T}, \ldots, y_N^{0:T}\Big)$, with $y_i^{0:T} = \Big(y_{i0}, y_{i1}, \ldots, y_{iT}\Big)$, the unobserved heterogeneity ($\mu_i = \mu_1, \ldots, \mu_N$), and the future shocks $u_{i,T+k} = \Big(u_{1,T+k}, \ldots, u_{N,T+k}\Big)$:

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

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

$$\mathcal{R}\Big(\hat{y}_{i,T+k}\Big) = \mathbb{E}_{\theta,\phi,\pi(\cdot)}^{(\mathcal{Y}_N,\mu_i,u_{i,T+k})}\Big[L_N\Big(\hat{y}_{i,T+k}, y_{i,T+k}\Big)\Big] \tag{21}$$

where $L_N\Big(\hat{y}_{i,T+k}, y_{i,T+k}\Big) = \sum_{i=1}^{N}\Big(\hat{y}_{i,T+k} - y_{i,T+k}\Big)^2$ 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 $\hat{y}_{i,T+k} = (\hat{y}_{1,T+k}, \ldots, \hat{y}_{N,T+k})'$ 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 (21) 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:

$$\begin{aligned}\mathbb{E}_{\theta,\phi,\pi(\cdot)}^{(\mathcal{Y}_N,\mu_i,u_{i,T+k})}\Big[L_N\Big(\hat{y}_{i,T+k}, y_{i,T+k}\Big)\Big] &= \sum_{i=1}^{N}\Bigg\{\Big(\hat{y}_{i,T+k} - \mathbb{E}_{\theta,\phi,\pi(\cdot)}^{(\mathcal{Y}_i,\mu_i,u_{i,T+k})}[y_{i,T+k}]\Big)^2 + \\ &\quad + \mathbb{V}_{\theta,\phi,\pi(\cdot)}^{(\mathcal{Y}_i,\mu_i,u_{i,T+k})}[y_{i,T+k}]\Bigg\}\end{aligned} \tag{22}$$

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:

$$\hat{y}_{i,T+k}^{op} = \mathbb{E}_{\theta,\phi,\pi(\cdot)}^{(\mathcal{Y}_i,\mu_i,u_{i,T+k})}[y_{i,T+k}] = \mathbb{E}_{\theta,\phi,\pi(\cdot)}^{(\mathcal{Y}_i,\mu_i)}[\mu_i] + (\theta \cdot c_{it}) \tag{23}$$

where the acronym *op* stands for 'optimal'. Then, the compound risk in (21) associated with the infeasible oracle forecast can be rewritten as:

$$\mathcal{R}^{op} = \mathbb{E}_{\theta,\phi,\pi(\cdot)}^{(\mathcal{Y}_i,\mu_i,u_{i,T+k})}\Bigg\{\sum_{i=1}^{N}\Big(\mathbb{V}_{\theta,\phi,\pi(\cdot)}^{(\mathcal{Y}_i,\mu_i)}[\mu_i] + \sigma_u^2\Big)\Bigg\} \tag{24}$$

The optimal compound risk in (24) consists of two components: uncertainty concerning the individual–specific heterogeneity on the observations i and uncertainty with respect to the error terms. Because of 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 (23) is assessed through the 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:

$$\pi\Big(y_i^{1:T}|y_{i0},\mu_i,\theta\Big) \propto exp\bigg\{-\frac{1}{2\sigma_u^2}\sum_{t=1}^{T}\Big(y_{it}-(c_{it-l}|\chi)\theta_t-\mu_i(\theta)\Big)^2\bigg\} \propto \bigg\{-\frac{T}{2\sigma_u^2}\Big(\hat{\mu}_i(\theta)-\mu_i\Big)^2\bigg\} \tag{25}$$

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

$$\hat{\mu}_i(\theta) = \frac{1}{T}\sum_{t=1}^{T}\Big(y_{it}-(\theta\cdot c_{it-l})\Big) \tag{26}$$

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

$$\pi\Big(\mu_i|y_i^{0:T},\theta\Big) = \pi\Big(\mu_i|\hat{\mu}_i,y_{i0},\theta\Big) = \frac{\pi\Big(\hat{\mu}_i|\mu_i,y_{i0},\theta\Big)\cdot\pi\Big(\mu_i|y_{i0}\Big)}{exp\Big\{ln\Big(\pi(\hat{\mu}_i|y_{i0})\Big)\Big\}} \tag{27}$$

The last step to obtain the Tweedie's formula is to differentiate the equation $\pi\Big(\mu_i|\hat{\mu}_i,y_{i0},\theta\Big)$ in (27) with respect to $\hat{\mu}_i$ and solve the equation for the posterior mean $\mathbb{E}^{(\mathcal{Y}_i,\mu_i)}_{\theta,\phi,\pi(\cdot)}[\mu_i]$ in (23). Thus, the Tweedie's formula equals to:

$$\mathbb{E}^{(\mathcal{Y}_i,\mu_i)}_{\theta,\phi,\pi(\cdot)}[\mu_i] = \hat{\mu}_i(\theta) + \frac{\sigma_u^2}{T}\cdot\frac{\partial}{\partial\hat{\mu}_i(\theta)}ln\Big(\hat{\mu}_i(\theta),y_{i0}\Big) \tag{28}$$

where the second term in (28) 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; contrarily 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 $\pi\Big(\hat{\mu}_i(\theta),y_{i0}\Big)$ in (28) by estimates. According to the multivariate ROB procedure, the cross-sectional information uploaded within the system set into $\mathcal{E}$

. In dynamic panel data, consistent estimates of the unkown parameters θ can be obtaining through Genelarized Method of Moments estimators. In this study, they correspond to the AR(λ) coefficients related to predetermined and endogenous variables[7]. Let the stationarity and moment conditions in (3)-(5) hold in the system, the time-series regressions are valid (or computational) and GMM estimators are feasible. Concerning the density $\pi\Big(\hat{\mu}_i(\theta),y_{i0}\Big)$, I estimate it by using Finite Mixture approximation of Multivariate (FMM) distributions:

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

$$\pi_{mix}\Big(\hat{\mu}_i, y_{i0} \mid |\chi|, c_{i0}\Big) = |\chi| \cdot \pi_{\xi}\Big(\hat{\mu}_i, y_{i0} \mid c_{i0}\Big) \quad \texttt{with} \quad |\chi| > 0 \tag{29}$$

where $\pi_{\xi}(\cdot)$ is the conditional density distribution of heterogeneous effects with sample size $|\chi|$. In this way, I would able to account for the whole cross-sectional information in order to obtain estimates of (non-)homogenous parameters θ (`first step`) and density $\pi_{\xi}(\cdot)$ (`second step`). Here, I focus on the only `best` promising submodels achieved through the shrinking process, working with sufficiently high posterior consistency.

The FMM distributions and their moments themselves (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 $\bar{m} = 0, 1, \ldots, \varkappa$ 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[8] to draw posteriors for $\hat{\mu}_i$ from the (proposal) joint density distribution $\pi^{\bar{m}} = |\chi| \cdot \pi^*_{\xi}\Big(\hat{\mu}^{\bar{m}}_i, y^{\bar{m}}_{i0} \mid c^{\bar{m}}_{i0}\Big)$, with probability $\mathfrak{p}_{\bar{m}}$ equals to:

$$\mathfrak{p}_{\bar{m}} = \frac{\pi\Big(\hat{\mu}^{\bar{m}}_i, y^{\bar{m}}_{i0} \mid \hat{\mu}^{\bar{m}-l}_i, \mathcal{Y}_i, \Big\{\theta_t\Big\}_{t=1}^{T}, c^{\bar{m}}_{i0}\Big) \cdot \pi^{\bar{m}-l}}{\pi\Big(\hat{\mu}^{\bar{m}-l}_i, y^{\bar{m}}_{i0} \mid \hat{\mu}^{\bar{m}-l}_i, \mathcal{Y}_i, \Big\{\theta_t\Big\}_{t=1}^{T}, c^{\bar{m}}_{i0}\Big) \cdot \pi^{\bar{m}}} \tag{30}$$

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:

$$\pi\Big(\theta_t \mid \mathcal{Y}_i\Big) \cdot \pi^*(\theta^*_t \mid \theta_t) \cdot \mathfrak{p}(\theta^*_t, \theta_t) = \pi\Big(\theta^*_t \mid \mathcal{Y}_i\Big) \cdot \pi^*(\theta_t \mid \theta^*_t) \tag{31}$$

where

$$\mathfrak{p}(\theta^*_t, \theta_t) = \min\left[\frac{\pi(\theta^*_t \mid \mathcal{Y}_i) \cdot \pi^*(\theta_t \mid \theta^*_t)}{\pi(\theta_t \mid \mathcal{Y}_i) \cdot \pi^*(\theta^*_t \mid \theta_t)}, 1\right] \cong \mathfrak{p}_{\bar{m}} \tag{32}$$

with $\mathfrak{p}(\theta^*_t, \theta_t)$ displaying the probability to accept or reject a draw[9] and $\pi^*(\cdot)$ denoting the density distribution according to sample size $|\chi|^*$. In this way, I am able to get the same probability that each submodel

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

[9] 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.

M_ξ would be true. In addition, since posterior distributions corresponds – by construction – to the FMM distributions, I define three possible intervals – displayed in (33) – in which the posterior predictive variance of $\mu_i\Big(\mathbb{V}^{(\mathcal{Y}_i,\mu_i)}_{\theta,\phi,\pi(\bar{m})}[\mu_i]\Big)$ 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 (Section 2.2), I ensure that lower variability will be associated to less relative regrets during the estimating procedure, achieving more accurate forecasts.

$$\begin{cases} \underset{(\textit{high dimension})}{0.5 < \mathbb{V}^{(\mathcal{Y}_i,\mu_i)}_{\theta,\phi,\pi(\bar{m})}[\mu_i] \leq 1.0} \quad \text{with} \quad \xi^* > 10 \quad (\ \texttt{heterogeneity}\ ) \\ \underset{(\textit{moderate dimension})}{0.1 \leq \mathbb{V}^{(\mathcal{Y}_i,\mu_i)}_{\theta,\phi,\pi(\bar{m})}[\mu_i] \leq 0.5} \quad \text{with} \quad 5 < \xi^* \leq 10 \quad (\ \texttt{sufficient-homogeneity}\ ) \\ \underset{(\textit{small dimension})}{0.0 \leq \mathbb{V}^{(\mathcal{Y}_i,\mu_i)}_{\theta,\phi,\pi(\bar{m})}[\mu_i] < 0.1} \quad \text{with} \quad \xi^* \leq 5 \quad (\ \texttt{near-homogeneity}\ ) \end{cases} \tag{33}$$

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 because of 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 (24). 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,

$$\limsup_{N\to\infty\ \pi^{\bar{m}}\subset\mathcal{E}} \frac{\mathcal{R}\Big(\hat{y}_{i,T+k},\pi^{\bar{m}}\Big) - \mathcal{R}^{op}\Big(\pi^{\bar{m}}\Big)}{\Big\{N\xi^*\cdot\mathbb{E}^{\mathcal{Y}_i,\mu_i}_{\theta,\phi,\pi^{\bar{m}}}\Big(\mathbb{V}^{(\mathcal{Y}_i,\mu_i)}_{\theta,\phi,\pi^{\bar{m}}}[\mu_i]\Big)\Big\} + N(\xi^*)\vartheta_0} \leq 0 \tag{34}$$

In (34), some considerations are in order. `(i)` The predictor $\hat{y}_{i,T+k}$ in (21) is constructed by replacing θ with a consistent estimator $\hat{\theta}$ (estimated AR(λ) coefficients) and individual outcome values $\pi\Big(\hat{\mu}_i(\theta), y_{i0}\Big)$ in (28) with estimates based on the FMM distributions. `(ii)` Taking expectations over $y^{0:T}$ in (24), 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 $\Big(MSE(\hat{\theta}) = E_{\hat{\theta}}\Big[\sum_{i=1}^{N}(\hat{y}_{i,T+k} - y_{i,T+k})^2\Big]\Big)$, 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, the larger the 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 (24) 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 favours 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)[10]. (v) The ratio-optimality in (34) 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:

$$\hat{y}_{i,T+k} = \left[\hat{\mu}_i^{\bar{m}}(\theta) + \frac{\hat{\sigma}_u^2}{T} \cdot \frac{\partial}{\partial \hat{\mu}_i^{\bar{m}}(\theta)} ln\Big(\hat{\mu}_i^{\bar{m}}(\theta), y_{i0}^{\bar{m}}\Big)\right]^{\bar{m}} + \hat{\theta} y_{it} \tag{35}$$

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

$$L\Big(y_i^{0:T} | \ddot{\tilde{\omega}}\Big) \propto exp\Bigg\{ -\frac{1}{2}\Bigg[\sum_{t=1}^{T} \Big(y_{it} - (c_{it}^{\bar{m}}|\dot{\chi})\hat{\theta}_t - \hat{\mu}_i^{\bar{m}}(\hat{\theta})\Big)'\Bigg] \cdot (\hat{\sigma}_u^2)^{-1} \cdot \Bigg[\sum_{t=1}^{T} \Big(y_{it} - (c_{it}^{\bar{m}}|\dot{\chi})\hat{\theta}_t - \hat{\mu}_i^{\bar{m}}(\hat{\theta})\Big)\Bigg]\Bigg\} \tag{36}$$

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 (equation (23)). 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 $\ddot{\tilde{\omega}}$:

$$\theta_t | \theta_{t-l}, \hat{y}_{i,T+k}, \ddot{\tilde{\omega}}_{-\theta_t} \sim N\Big(\ddot{\bar{\theta}}_{t|T+k}, \ddot{\bar{\rho}}_{t|T+k}\Big) \tag{37}$$

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

where

$$\ddot{\bar{\theta}}_{t|T+k} = \Big[\Big(\ddot{\bar{\rho}}_{t|T+k}^{-1} \cdot \bar{\theta}\Big) + \sum_{t=1}^{T}\Big((c_{it}^{\bar{m}}|\dot{\chi})' \cdot (\hat{\sigma}_u^2)^{-1} \cdot (c_{it}^{\bar{m}}|\dot{\chi})\Big)\hat{\theta}_t\Big] \tag{38}$$

$$\ddot{\bar{\rho}}_{t|T+k} = \Big[I_h - \Big(\bar{\rho} \cdot \ddot{\bar{\rho}}_{T+k|t}^{-1}\Big)\Big] \cdot \bar{\rho} \tag{39}$$

with

$$\hat{\theta}_t = \Big[(c_{it}^{\bar{m}}|\dot{\chi})' \cdot (\hat{\sigma}_u^2)^{-1} \cdot (c_{it}^{\bar{m}}|\dot{\chi})\Big]^{-1} \cdot \Big[(c_{it}^{\bar{m}}|\dot{\chi})' \cdot (\hat{\sigma}_u^2)^{-1} \cdot y_{it}\Big] \tag{40}$$

The equations (39) and (40) 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, equation (38) can be rewritten as

$$\ddot{\bar{\theta}}_{t|T+k} = \Big[\Big(\ddot{\bar{\rho}}_{t|T+k}^{-1} \cdot \bar{\theta}\Big) + \Big(\sum_{t=1}^{T}(c_{it}^{\bar{m}}|\dot{\chi})' \cdot (\hat{\sigma}_u^2)^{-1} \cdot y_{it}\Big)\Big] \tag{41}$$

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 $\Big\{\theta_t\Big\}_{t=1}^{T}$ from $N\Big(\ddot{\bar{\theta}}_{T|T}, \ddot{\bar{\rho}}_{T|T}\Big)$[11] in (37), the other posterior distributions can be defined as:

$$\pi(\hat{\mu}_i|\hat{y}_{i,T+k}, \hat{\theta}_t) \sim N\Big(\tilde{\delta}_{\mu_i}, \tilde{\Psi}_{\mu_i}\Big) \tag{42}$$

$$\pi(\hat{y}_{i0}|\hat{\mu}_i^{\bar{m}}) = N(0, \hat{\kappa}) \tag{43}$$

$$\pi(\dot{\chi}) = \tilde{w}_{|\chi|} \cdot \begin{pmatrix} \xi^* \\ |\chi| \end{pmatrix}^{-1} \tag{44}$$

$$\pi(\hat{\sigma}_u^2|\hat{y}_{i,T+k}) = IG\Big(\frac{\ddot{\omega}}{2}, \frac{\tilde{\nu}}{2}\Big) \tag{45}$$

Here, some considerations are in order.

In equation (42), $\tilde{\delta}_{\mu_i} \sim N\Big(0, \bar{\zeta}\Big)$ and $\tilde{\Psi}_{\mu_i} \sim IG\Big(\bar{\varphi}/2, \bar{\varepsilon}/2\Big)$, 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 equation (43), $\hat{\kappa} = \kappa \cdot \mathbb{V}_{\theta,\phi,\pi(\bar{m})}^{(\mathcal{Y}_i,\mu_i)}[\mu_i]$, 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

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

according to the sample size $|\chi|$ as described in (33).

In equation (44), $\tilde{w}_{|\chi|}$ 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 equation (45), $\ddot{\omega} = \bar{\omega} + \hat{\bar{\omega}}$ 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, $\hat{\bar{\omega}} = \Big(\sum_{t=1}^{T} log(\tau_t)/t \Big) + log\Big(\sum_{t=1}^{T} (1/\tau_t) \Big) - log(t)$ and $\hat{\nu} = (t \cdot \hat{\bar{\omega}})/\Big(\sum_{t=1}^{T} (1/\tau_t) \Big)$ 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 equation (2) related to the posteriors $\hat{y}_{i0}$ in (43), and $\bar{\rho} = I_{\xi^*}$.

4 Theoretical Properties

The proof of the ratio-optimality results and posterior consistency are as follow.

Assumption 4.1 (Identification: General Model)**.** *Consider the HDPD model in (2):*

1. *Model Setup*

   a. *$(c_{i0}, \mu_i, \sigma_u^2)$ are i.i.d. across i.*

   b. *For all t, conditional on (y_{it}, c_{it-l}), $y_{i,t}^c$ is independent of $\phi = (\mu_i, \sigma_u^2)$.*

   c. *$x_{i,0:T}$ is independent of $\phi = (\mu_i, \sigma_u^2)$.*

   d. *Conditioning on (c_{i0}, μ_i) and σ_u^2, they are independent of each other.*

   e. *Let $u_{it} \sim N(0, \sigma_u^2)$ is i.i.d. across i and independent of (c_{it-l}, μ_i).*

2. *Indentification*

   a. *The characteristic functions for $\mu_i|c_{i0}$ and $\sigma_u^2|c_{i0}$ do not steadily disappearing altogether into high-shrinkage processing method.*

   b. *Let $v_{it} = \mu_i + u_{it}$ be the composite error at time t, the sequence $\{v_{it} : t = 1, 2, \ldots, T\}$ is almost certainly serially correlated, and definitely is if $\{u_{it}\}$ is serially uncorrelated.*

   c. *With the panel setup in (2) – with large N and sufficiently large T – x_{it} includes interactions of variables with time periods dummies and general non-linear functions and interactions, so the model is quite flexible.*

   d. *For the CRE approach, each kind of covariates in c_{it} is separated out and the heterogeneous factors in μ_i are correlated with them.*

Assumption 4.2 (Identification: Unbalanced Panel)**.** *For all i:*

1. *Indentification*

   a. c_{i0} *is observed.*

   b. (y_{iT}, z_{iT}, x_{iT}) *are observed.*

   c. *With the panel setup in (2) – with large N and sufficiently large T – x_{it} includes interactions of variables with time periods dummies and general non-linear functions and interactions, so the model is quite flexible.*

   d. *For the CRE approach, each kind of covariates in c_{it-l} is separated out and the heterogeneous factors in μ_i are correlated with them.*

2. *Sequential Exogeneity (Conditional on the Unobserved Effects)*

   a. $E(y_{it}|c_{it}, c_{it-1}, \ldots, c_{i1}, \mu_i) = E(y_{it}|c_{it}, \mu_i) = \theta c_{it} + \mu_i$ *for* $i = 1, 2, \ldots, N$.

   b. *Sequential exogeneity is a middle ground between contemporaneous and strict exogeneity. It allows lagged dependent variables and other variables that change in reaction to past shocks.*

   c. *Because including contemporaneous exogeneity, standard kinds of endogeneity – where some elements of c_{it} are correlated with u_{it} – are ruled out such as measurement error, simultaneity, and time-varying omitted variables.*

   d. *Sequential exogeneity is less restrictive than strict exogeneity imposing restrictions on economic conditions.*

3. *Model Setup*

   a. *The term "correlated random effects" is used to denote situations where one models the relationship between $\{\mu_i\}$ and $\{c_{it}\}$.*

   b. *The CRE approach allows to unify fixed and random effects estimation approaches.*

   c. *With the CRE approach, time-constant variables can be included within the system.*

   d. *GMM estimators can be used to consistently estimate all time-varying parameters in θ.*

   e. $\hat{\theta}_{GMM} \xrightarrow{d} N\left(\theta, \frac{1}{N}V_\theta\right)$ *where V_θ denotes the covariance matrix estimated via posteriors on $\{\bar{\theta}, \bar{\rho}\}$.*

Assumption 4.3 (Identification: Conjugate Proper Informative Priors)**.** *Let $\mathbb{E}^{\mathcal{Y}_i,\mu_i}_{\theta,\phi,\pi(\cdot)}$ be the expectations over the observed trajectories ($\mathcal{Y}_i$) and the unobserved heterogeneity (μ_i), the CIPM priors in (16) - (20) can be rewritten as:*

i.

$$\theta_t|\theta_{t-l}, \mathbb{E}^{\mathcal{Y}_i,\mu_i}_{\theta,\phi,\pi(\cdot)}[y_{i,T+k}] \sim N\Big(\bar{\theta}, \bar{\rho}\Big)$$

ii.

$$\mu_i|\mathbb{E}^{\mathcal{Y}_i,\mu_i}_{\theta,\phi,\pi(\cdot)}[y_{i,T+k}] \sim N\Big(\delta_{\mu_i}, \Psi_{\mu_i}\Big) \quad \textbf{\textit{with}} \quad \delta_{\mu_i} \sim N\Big(0, \zeta\Big) \quad \textbf{\textit{and}} \quad \Psi_{\mu_i} \sim IG\Big(\frac{\varphi}{2}, \frac{\varepsilon}{2}\Big)$$

iii.

$$y_{i0}|\mathbb{E}^{\mathcal{Y}_i,\mu_i}_{\theta,\phi,\pi(\cdot)}[y_{i,T+k}] = N(0, \kappa)$$

iv.

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

v.

$$\sigma^2_u|\mathbb{E}^{\mathcal{Y}_i,\mu_i}_{\theta,\phi,\pi(\cdot)}[y_{i,T+k}] \sim IG\Big(\frac{\bar{\omega}}{2}, \frac{\nu}{2}\Big)$$

Statement 4.3.1 (Posterior Distributions)**.** *Under Assumption (4.3), all posterior distributions in (41)-(45) hold and are estimable through MCMC algorithms and implementations.*

Statement 4.3.2 (Moment Conditions and GMM Estimator)**.** *Let all moment conditions in (3)-(5) hold for all i. Then, under Assumption (4.3) and Statement (4.3.1), the GMM estimator is consistent and equals to:*

$$\hat{\theta}_t = \bigg[\Big(c^{\bar{m}}_{it}|\dot{\chi}\Big)' \cdot \hat{\sigma}^2_u \cdot \Big(c^{\bar{m}}_{it}|\dot{\chi}\Big)\bigg]^{-1} \cdot \ \bigg[\Big(c^{\bar{m}}_{it}|\dot{\chi}\Big)' \cdot \hat{\sigma}^2_u \cdot y_{it}\bigg]$$

Theorem 4.4 (Correlated Random Coefficients: Cross-sectional Homoskedasticity)**.** *Let the Tweedie's formula in (28) hold.*

1. *The proof of the Tweedie correction for (28) – when dealing with correlated random coefficients – builds on a finite mixture approximation of multivariate distributions as defined in (29).*

2. *With correlated random coefficients homoskedastic case, one would work with the following space:*

$$\mathcal{Q}^{\bar{m}} = \bigg\{\bar{m} \subset \mathcal{E} : \iint \parallel \mu_i \parallel^2_2 \cdot \pi^{\bar{m}} \cdot q^*_{i0}(c_{i0}) \ d\mu_i \ dc_{i0} \leq M_{\xi^*}\bigg\} \quad \textit{for} \quad M_{\xi^*} > 0$$

where $\pi^{\bar{m}}$ denotes the (proposal) joint density distribution to draw posteriors for $\hat{\mu}_i$ through Metropolis-Hastings algorithm, and q^*_{i0} denotes the true marginal density of c_{i0}.

3. *The space for common parameters* $\upsilon = (\theta, \sigma_u^2)$ *is* $\Upsilon = \mathbb{R}_{|\chi|} \cdot \sigma_u^2$.

4. *The conditional individual–specific likelihood function is described as:*

$$\mathcal{L}(y_{it}|\upsilon, \bar{m}) = \pi\Big(c_{it}|c_{it-l}, c_{i,0:T}\Big) \iint \prod_t \pi\Big(y_{it}; \theta' c_{it-l} + \mu_i, \sigma_u^2\Big) \cdot \pi^{\bar{m}} \cdot q_{i0}^*(c_{i0})\ d\mu_i\ dc_{i0}$$

Theorem 4.5 (Posterior Consistency: Correlated Random Coefficients)**.** *Given the DPB-CRE in (2):*

1. `Model` $\rightarrow$ *Assumptions (4.1) and (4.2).*

2. `Covariates` $\rightarrow$ $\Big(y_{i,0}^o, y_{i,0}^c, z_{i,0}^s, z_{i,0}^p, x_{i,0:T}\Big)$ *satisfy Assumptions (4.1) and (4.2).*

3. `Common Parameters`

   a. υ *is unknown and estimable.*

   b. *The domain of* σ_u^2 *is finite and estimable (homoskedastic case).*

   Thus, the posterior would be highly consistent at υ *given* $\pi^{\bar{m}}$ *with regimes* $\bar{m}$.

Theorem 4.6 (Ratio-Optimality: Tweedie Correction)**.** *Suppose that Assumptions (4.1) - (4.3) hold: then, in the DPB-CRE in (2), the EB predictor* $(\hat{y}_{i,T+k})$ *defined in (35) achieves* ϑ_0*-ratio-optimality uniformly for priors* $\pi^{\bar{m}} \subset \mathcal{E}$, *with* $\vartheta_0 \geq 0$.

Statement 4.6.1 (Properties of the Common Parameters Estimation)**.** *The estimators of the common parameters* (υ) *have the following properties:*

i

$$\mathbb{E}_{\theta,\phi,\pi^{\bar{m}}}^{\mathcal{Y}_i,\mu_i}[\hat{\sigma}_u]^4 = o_{u.\pi^{\bar{m}}}(N^+)$$

ii

$$\mathbb{E}_{\theta,\phi,\pi^{\bar{m}}}^{\mathcal{Y}_i,\mu_i}\Big[|\sqrt{N}(\hat{\theta} - \theta)|^4\Big] = o_{u.\pi^{\bar{m}}}(N^+)$$

iii

$$\mathbb{E}_{\theta,\phi,\pi^{\bar{m}}}^{\mathcal{Y}_i,\mu_i}\Big[|\sqrt{N}(\hat{\sigma}_u^2 - \sigma_u^2)|^2\Big] = o_{u.\pi^{\bar{m}}}(N^+)$$

iv

$$N \iint \Big[|\chi| \cdot \pi_\xi^*\Big(\hat{\mu}_i^{\bar{m}}, y_{i0}^{\bar{m}} \mid c_{i0}^{\bar{m}}\Big)\Big]^2 \cdot \Big[|\chi| \cdot \pi_\xi\Big(\hat{\mu}_i, y_{i0} \mid c_{i0}\Big)\Big] d\hat{\mu}\ dy_{i0} = o_{u.\pi^{\bar{m}}}(N^+) \quad \textit{with} \quad \bar{m} < \infty$$

where $\pi^{\bar{m}} = |\chi| \cdot \pi_\xi^*\Big(\hat{\mu}_i^{\bar{m}}, y_{i0}^{\bar{m}} \mid c_{i0}^{\bar{m}}\Big)$ *stands for the (designed) joint density distribution – under probability*

$(\mathfrak{p}_{\bar{m}})$ – *to get draw samples from posteriors of the empirical distribution of* $\hat{\mu}_i$*.*

v

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

where the further $\mathfrak{p}_{\bar{m}}$ *will be close to zero, the lower the PMP and then the lower the possibility that individual outcome values would perform well the data in the cross-sectional information subset* $\mathcal{E}$*. Whether* $\mathfrak{p}_{\bar{m}} = 1$*, the associated model solution* (M_{ξ^*}) *would contain best estimates for the* μ_i*'s sufficient statistic and individual outcome values in (28). Thus, one would expect to find the final model solution with higher lBF.*

Statement 4.6.2 (Posterior Mean Functions in DPB-CRE with Empirical Bayes Approach)**.** *The Theorem (4.6) can be proved by showing that the below inequality holds:*

$$\limsup_{N\to\infty} \sup_{\pi^{\bar{m}}\subset\mathcal{E}} \frac{N\mathbb{E}_{\theta,\phi,\pi^{\bar{m}}}^{\mathcal{Y}_i,\mu_i}\left[\left(\hat{y}_{i,T+k} - (\mu_i + \theta y_{it})\right)^2\right]}{N\mathbb{E}_{\theta,\phi,\pi^{\bar{m}}}^{\mathcal{Y}_i,\mu_i}\left[\left(\mu_i - \mathbb{E}_{\theta,\phi,\pi^{\bar{m}}}^{\mathcal{Y}_i,\mu_i}[\mu_i]\right)^2\right] + N^{\vartheta_0}} \leq 1$$

Thus, according to a generalization of the Brown and Greenshtein (2009)'s insight[12]*, the sufficient condition (4.6.2) is proved by decomposing the discrepancy between the predictor* $\hat{y}_{i,T+k}$ *and the unknown parameters* $(\mu_i + \theta y_{it})$ *into three terms:*

i

$$N\mathbb{E}_{\theta,\phi,\pi^{\bar{m}}}^{\mathcal{Y}_i,\mu_i}[\hat{y}_{i,T+k}]^2 = o_{u.\pi^{\bar{m}}}(N^{\vartheta_0})$$

It displays the difference between the posterior mean of μ_i *according to the Tweedie correction defined in (30) and the (proposal) joint density distribution* $\pi^{\bar{m}}$ *with probability* $\mathfrak{p}_{\bar{m}}$ *in (29).*

ii

$$\limsup_{N\to\infty} \sup_{\pi^{\bar{m}}\subset\mathcal{E}} \frac{N\mathbb{E}_{\theta,\phi,\pi^{\bar{m}}}^{\mathcal{Y}_i,\mu_i}\left[\mu_i + \theta y_{it})\right]^2}{N\mathbb{E}_{\theta,\phi,\pi^{\bar{m}}}^{\mathcal{Y}_i,\mu_i}\left[\left(\mu_i - \mathbb{E}_{\theta,\phi,\pi^{\bar{m}}}^{\mathcal{Y}_i,\mu_i}[\mu_i]\right)^2\right] + N^{\vartheta_0}} \leq 1$$

It displays the difference between the posterior density estimated in (29) and the (proposal) joint density distribution $\pi^{\bar{m}}$ *with probability* $\mathfrak{p}_{\bar{m}}$ *in (30).*

iii

$$N\mathbb{E}_{\theta,\phi,\pi^{\bar{m}}}^{\mathcal{Y}_i,\mu_i}[\pi^{\bar{m}}] = o_{u.\pi^{\bar{m}}}(N^{+})$$

[12] See, for instance, Liu et al. (2020) using the insight of Brown and Greenshtein (2009) to prove ratio-optimality by replacing the μ_i's distribution with a kernel density estimator.

It displays the structural model uncertainty dealt with replacing the common parameters (θ) by estimates.

The results (i), (ii), and (iii) are relatively straightforward under Statement (4.6.1) and can be handled according to Statements (4.3.1) and (4.3.2).

Theorem 4.7 (Finite Mixtures of Multivariate Distributions)**.** *Suppose that Assumptions (4.1)-(4.3) hold: then, the density $\pi\Big(\hat{\mu}_i(\theta), y_{i0}\Big)$ can be estiamted using FMM distributions as defined in (29).*

Statement 4.7.1 (Mixture Components and Model Classes)**.** *The finite mixture of multivariate distributions in (29) is able to approximate a large set of distributions as the number of mixture (potential) combination of predictors ($|\chi|$) – set into $\mathcal{E}$ – increases. In this study, I use finite mixtures of multivariate normal-inverse-gamma distributions dealing with the common parameters υ (Theorem (4.4)). According to posterior distributions (41)-(45) – under Assumption (4.3) – the finite mixtures of multivariate distributions will be:*

$$\pi^*_{mix}\Big(\hat{\mu}_i^{\bar{m}}, y_{i0}^{\bar{m}} \mid |\chi|, c_{i0}^{\bar{m}}\Big) = |\chi| \cdot \pi^*_{\xi}\Big(\hat{\mu}_i^{\bar{m}}, y_{i0}^{\bar{m}} \mid c_{i0}^{\bar{m}}\Big) \quad \textit{with} \quad |\chi| > 0 \quad \textit{and} \quad \mathfrak{p}_{\bar{m}} \leq 1$$

Theorem 4.8 (Normal-Inverse-Gamma-Distribution)**.** *It is the conjugate prior of a normal distribution with unknown mean and variance. For instance, suppose that the distribution of the (non-)homogeneous parameters θ would be affected by heterogeneous effects ϕ_i. According to CIPM in (15):*

$$\theta|\sigma_u^2 \sim N\Big(\bar{\theta}, \bar{\rho} \otimes \sigma_u^2\Big) \quad \textit{and} \quad \sigma_u^2|\bar{\omega}, \nu \sim IG\Big(\frac{\bar{\omega}}{2}, \frac{\nu}{2}\Big)$$

Then,

$$\Big(\theta, \sigma_u^2\Big) \sim NIG\Big(\bar{\theta}, \bar{\rho}, \dot{\bar{\omega}}, \dot{\bar{\nu}}\Big)$$

with $\dot{\bar{\omega}} = \frac{\bar{\omega}}{2}$ and $\dot{\nu} = \frac{\nu}{2}$.

Concerning $\Big(\theta, \mu_i\Big)$:

$$\theta|\mu_i \sim N\Big(\bar{\theta}, \bar{\rho} \otimes \Psi_{\mu_i}\Big) \quad \textit{and} \quad \mu_i|\delta_{\mu_i}, \Psi_{\mu_i} \sim IG\Big(\delta_{\mu_i}, \Psi_{\mu_i}\Big)$$

Then,

$$\left(\theta, \mu_i\right) \sim NIG\left(\bar{\theta}, \bar{\rho}, \delta_{\mu_i}, \Psi_{\mu_i}\right)$$

Theorem 4.9 (Time-varying Parameters and GMM Estimators). *Let the stationarity and moment conditions in (3)-(5) hold in the system, then the time-series regressions are valid (or computational) and GMM estimators are feasible.*

5 Empirical Evidence

5.1 Data Description and Results

The HDPB-CRE in (2) contains 22 country-specific models, including 9 advanced economies[13], 7 emerging economies[14], and 6 non European countries[15]. 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, and thus interdependencies and inter-country linkages can be investigated in depth. The estimation sample is expressed in years and covers the period from $1990 - 2021$, and all data comes from 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)` endogeneity issues; `(ii)` interdependency, commonality, and homogeneity; `(iii)` relevant monetary and fiscal policy interactions; and `(iv)` misspecified 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 in four groups: (i) `Economic Status`, including 41 determinants combining information on education, income, economic development, and labour 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[16] better fit the data

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

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

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

[16] 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`.

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 model uncertainty and overfitting are avoided. Indeed, dealing with the sign certainty, the Conditional Posterior Sign (CPS)[17] tends to be close to 0 (such as predictors 7, 25) and 1 (such as predictors 2, 9, 11, 26, 31). `(ii)` Uncertain effects tend to persist for predictors 3, 5, 10, 16, 20, 22: thus, they should be interpreted with care. `(iii)` Socioeconomic factors matter more than economic status because of the ongoing outbreak of the epidemiology. `(iv)` The main policy tools correspond to some of the core variables of real and financial business cycles affecting the spreading and the transmission of spillover effects (such as current account balance, gross fixed capital formation, credit, and inflation rate). And `(v)` all predictors with PIPs $\geq \dot{\tau}$ (in bold in Table 1) will correspond to the ones to be accounted for the final solution.

Nevertheless, although the intense shrinkage in the parameter space, the final solution would still require some effort: indeed, there are 20 `best` promising that would better fit the data. Thus, according to the `third step`, I test for panel Granger (Non-)Causality among all selected predictors. In Table 2, I display the only covariates with p-value $< \dot{\tau}$ to be included in the submodels $M_{\xi^*} \subset \mathcal{E}$.

[17] 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 so split: the first column denotes the predictor number; the second and the third column describe the predictors and the corresponding labels, respectively; the fourth column refers to the measurement unit; and the last two columns displays 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 World Bank database.

Finally, the final model solution better performing the data – with $lBF = 13.49$ – consists of 10 final `best` covariates 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 $z^{p'}_{i,t-l}$. All their available lags, including lagged outcomes ($y^{o'}_{i,t-l}$), are then included as `external` instruments. In the estimation method, I also include two time-invariant effects (x_{1t} and x_{2t}) denoting the presence of structural breaks in 2008 (due to the global financial crisis) and in 2020 (due to the COVID-19 pandemic).

Table 2: Granger (Non-)Causality Test – `third step`

From c^*_{it} to y_{it}	emto	unem	cuhe	gghe	arat	frat	cab	gfcf	inf	gdpg
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}	emto	unem	cuhe	gghe	arat	frat	cab	gfcf	inf	gdpg
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 parenthesis) based on the panel Granger (Non-)Causality test. Here, c^*_{it} stands for the `best` final candidate predictors in $M_{\xi^*} \subset \mathcal{E}$ with higher lBF (equation (13)).

In Table 3, I display the main estimation outputs and diagnostic tests highlighting the performance of the HDPB-CRE model. 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 (33), dealing with high dimensional data carefully ($\mathbb{V}^{\mathcal{Y}_i,\mu_i}_{\theta,\phi,\pi(\bar{m})}[\mu_i] = 0.74$ with $\xi^* > 10$). `(iv)` In Table 2, highly strong causal links confirms 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 ($\dot{\chi}$). And `(vi)` the estimating procedure is robust dealing with the most of the explained variability of the outcomes ($R^2_{adj.} = 0.78$).

Table 3: Estimation Outputs and 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.

5.2 Forecasting and Policy Purposes

Concerning dynamic analyses, the total number of draws has been $2,000+3,000 = 5,000$, which corresponds to the sum of the final number of draws to discard and save, respectively. The convergence is obtained at about $1,000$ draws[18], used to conduct posterior inference at each t.

The natural conjugate prior refers to three subsamples: `(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 the productivity according to an ongoing pandemic crisis and the Russia-Ukraine war (predictors 15, 19). The time frame $2022 - 2023$ refers to outcomes absorbed in the forecasting analysis.

Without restrictions, the estimation sample amounts to 726 regression parameters: every estimate of the HDPB-CRE in (2) account for 22 country indices and 33 time periods. Let hyperparameters in $\tilde{\omega}$ be all known and estimable, posterior distributions are computed according to equations (37)-(39) for $\theta_t|\theta_{t-l}, \hat{y}_{i,T+k}$, (42)-(43) for moment distributions in $\mu_i|\hat{y}_{i,T+k}$ given initial values $(\hat{y}_{i0}|\hat{\mu}_i^{\bar{m}})$, (44) for the final `best` parameter space, and (45) for $u_t|\hat{y}_{i,T+k}$. All data are expressed in standard deviations.

In Figure 1, conditional forecasts for outcomes $\hat{y}_{i,T+k}$ are drawn for advanced (top plot) and emerging

[18] The convergence has been found by amounting to about 1.2 draws per regression parameter.

(bottom plot) economies. The yellow and red curves denote the 95% confidence bands, and the blue and purple curves denote the conditional[19] and unconditional[20] projections of outcomes $\hat{y}_{i,T+k}$ for each N country indexes and T time periods.

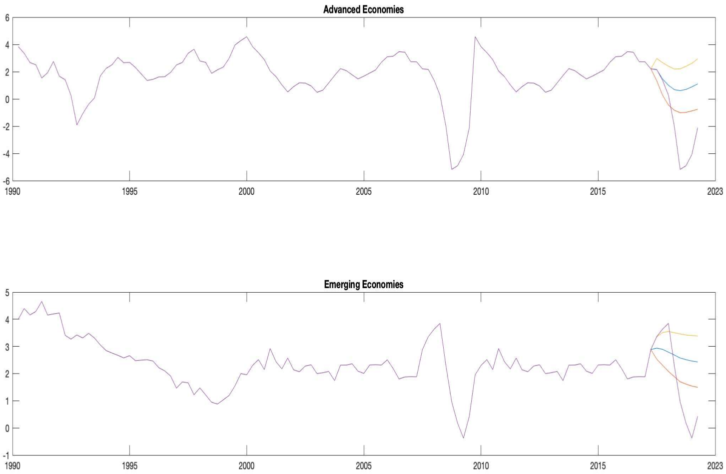

Figure 1: The plot draws conditional forecasts for outcomes $\hat{y}_{i,T+k}$ given individual–specific (μ_i) and time-fixed (α) effects given a pool of socioeconomic–demographic, real–financial, and policy determinants. All time-varying parameters are posterior means.

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, the need for forecasters and policymakers to account for heterogeneous effects (correlated random coefficients) when formulating policy strategies and forecasting in multivariate dynamic panel data. (ii) Conditional projections lie in the confidence interval; conversely, unconditional projections tend to diverge over T. Thus, when studying large time-varying panel data, cross-unit lagged interdependencies, dynamic feedback, and interactions have to be assessed in order to deal with endogeneity issues and misspecified dynamics. (iii) A hint of boosting productivity to potential growth $(2022 - 2023)$ can be observed among countries, mainly among advanced economies. Thus, although recent dynamics would 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 an early and coordinated fiscal consolidation.

From a policy perspective, three main results are highlighted. (i) Empirical forecasts show that most

[19]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}).

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

European emerging economies are strongly exposed to financial interlinkages and then highly dependent on other European countries (e.g., Western European countries). Nevertheless, the presence of persistent heterogeneities among countries' responses emphasize the need for accelerating financial development in developing countries, stimulating domestic resource mobilization, and supporting consistent reforms of the international financial system in order to boost investment and growth. (ii) Even if several measures have already been taken at the international and European Union levels, most countries have been limited to use monetary and fiscal tools effectively due to stringent economic–institutional interdependencies, and then they have not been able to deploy 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 the control of the outbreak. Thus, in a context of radical uncertainty and heterogeneous territorial effects, appropriate policy measures need to be addressed at the local level rather than globally. (iii) Heterogeneity among countries' responses would matter because of different policy adjustments applied by governments during a recession. Indeed, they have been led to follow distinct national rather than consensual international standards (such as in the current outbreak and previous pandemic crises). Overall, policy tools should be implemented by closely monitoring the evolution of the economic status for every country. More coordinated country-specific European and international measures and a participatory government are needed for ensuring robust health systems and more resilient economic development so as to safeguard against sudden outbreak on the global economy.

6 Relative Regrets for Tweedie Correction: MCMC-based Experiments

In this example, the performance of the estimation method is investigated by summarizing the regrets for Tweedie correction in (29) relative to the posterior predictive variance of the μ_i's. More precisely, according to (33), I consider three sequences of $\left(N, \xi^*, \mathbb{V}^{\mathcal{Y}_i,\mu_i}_{\theta,\phi,\pi(\bar{m})}[\mu_i]\right)$ with correlated random coefficients homoskedastic case to evaluate different improvements in the 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).

Table 4: 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 $\left(N, \xi^*, \mathbb{V}^{\mathcal{Y}_i,\mu_i}_{\theta,\phi,\pi(\bar{m})}[\mu_i]\right)$ with correlated random coefficients homoskedastic case conducted in the simulated example according to (33).

I include two additional empirical Bayes estimators dealing with alternative Tweedie corrections (Table 4): Kernel Density (KD) estimator (see, for instance, Liu et al. (2020)) and 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 dealt with the 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 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 5 provides the relative regrets for Tweedie corrections according to the three supposed MCMC-based designs. The best choice of ϑ_0 improving the forecasting performance in terms of ratio-optimality was set 0.5 (middle point in an arbitrary range [0.1 - 0.9]). The findings highlight the usefulness of the multivariate ROB procedure for dramatically shrinking the model size with high dimensional data, and the performance of the semiparametric Bayesian statistics based on the FMM distributions (Tweedie correction) for performing better forecasts. For instance, lower posterior predictive variances of the μ_i's are associated to less relative regrets. Compared to KD and NPLM estimates, FMM distributions show lower regrets. Replicating the experiment for highly larger sample size (e.g., $N = 100,000$) and lower ratio-optimality (e.g., $\vartheta_0 \cong 0.1$), I find that the relative regrets are negatively correlated with the number of cross-sectional units N and that less ratio-optimality – even if reduces computational costs – would suffer to higher associated regrets.

Table 5: Relative Regrets for Tweedie Corrections by MCMC-based Designs

	Design I	Design II	Design III
N	10000	10000	10000
$\mathbb{V}^{\mathcal{Y}_i,\mu_i}_{\theta,\phi,\pi^{\bar{m}}}[\mu_i]$	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 $\vartheta_0 = 0.5$.

All results in Table 5 find confirmation in Figure 2. More precisely, lower posterior predictive variances of the μ_i's are associated to less Mean Squared Errors (MSEs) 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 and then the FMM-based Tweedie Correction – in Theorem (1.6) – approaches 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.

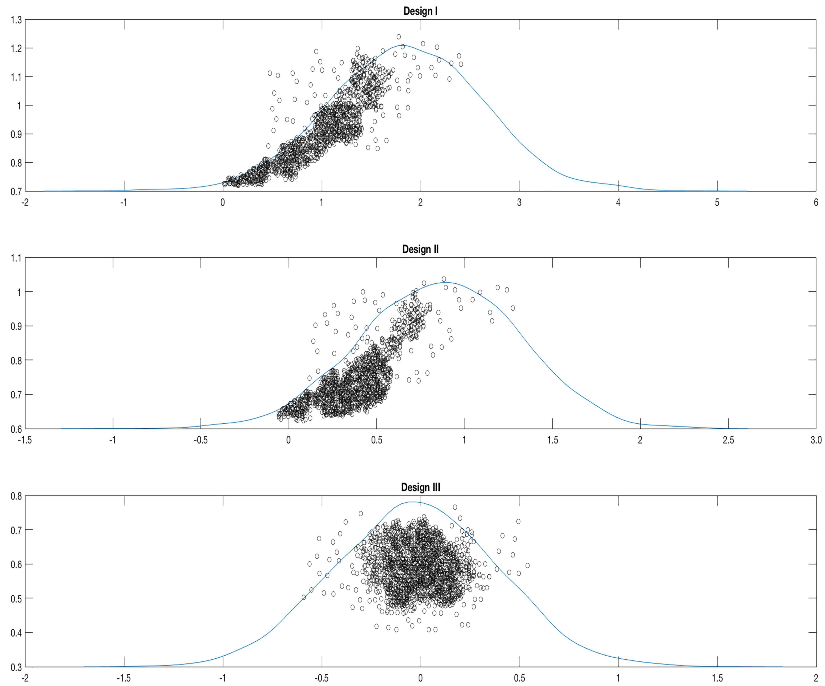

Figure 2: The panels show the MSEs associated to the three supposed MCMC-based designs. The solid lines display the posterior draw samples of the empirical distribution of $\hat{\mu}_i$ according to the (designed) joint density distribution $\pi^{\bar{m}}$ and the FMM-based Tweedie Correction.

7 Concluding remarks

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 non-parametric 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 with related works.

Compliance with Ethical Standards

Funding: No fund was used to conduct this study.

Conflict of Interest: The author declares no conflict of interest.

References

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

Anderson, T. W. and 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, New York*.

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

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

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

Arellano, M. and 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. and Honore, B. (2001). Panel data models: Some recent developments. *Handbook of Econometrics*, ed. by J. Heckman, and E. Leamer, 5:3229–3296.

Bester, C. A. and 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. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143.

Brown, L. D. and 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. *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, 2:3847–4605.

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

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

Compiani, G. and Kitamura, Y. (2016). Using mixtures in econometric models: a brief review and some new results. *The Econometrics Journal*, 19(3):C95–C127.

Dumitrescu, E. and 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.

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

Gu, J. and Koenker, R. (2017a). Empirical bayesball remixed: Empirical bayes methods for longitudinal data. *Journal of Applied Economics*, 32(3):575–599.

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

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

Hahn, J. and Newey, W. (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica*, 72:1295–1319.

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

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

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

Kiefer, J. and 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. and 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:1–32. Available at `https://arxiv.org/abs/1805.04178`.

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

Liu, L., Moon, H. R., and 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.

Pacifico, A. (2020). Robust open bayesian analysis: Overfitting, model uncertainty, and endogeneity issues in multiple regression models. *Econometric Reviews*, 40(2):148–176. DOI: `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. DOI: `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. *New York: Springer Verlag*.

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