Erasmus University Rotterdam

Erasmus School of Economics

Master Thesis in Econometrics and Management Science
Specialisation: Econometrics

Forecasting sales with micro-panels: Empirical Bayes approach Evidence from consumer goods sector

---

Author:

Michał Oleszak
Student number: 452451

Academic supervisor:
dr. Wendun Wang (EUR)

External supervisor:
Joeri Admiraal (Veneficus B.V.)

Co-reader:
Anoek Castelein (EUR)

Rotterdam, January 9, 2018

Abstract

Most of the existing non-Bayesian studies on dynamic panel models mainly focus on consistent estimation of the common auto-regressive parameters, while treating the heterogeneous intercept as a nuisance and removing it from the model, e.g. by first-differentiating. Bayesian literature proposes methods to obtain posterior distributions of the unit-specific intercept based on Markov Chain Monte Carlo sampling, which can be slow for large data sets. This paper investigates an empirical-Bayes procedure for obtaining the individual-specific intercept parameters without resorting to MCMC sampling. A set of simulation studies and empirical applications show that the proposed model outperforms benchmark models in some settings, especially for long-run forecasting of highly persistent data. Additionally, this paper proposes a bootstrap-based method for computing prediction intervals in the analysed modelling framework.

Key words: Forecasting Sales, Panel Data, Empirical Bayes, Tweedies Formula, Posterior Mean Predictor, Random Window Block Bootstrap

Contents

1 Introduction

Sales are one of the most common indicators of the performance of a company. Hence, knowing what future sales will amount to is vital for every firm, especially in the consumer goods sector, where the market is highly competitive. Accurate predictions of future sales allow the companies to make good investment decisions and to adjust the production process to the changing market conditions. Increasing future sales might also be regarded as a sign of high-quality management of the company, whereas decreasing forecasts may be read as an alert signifying that steps need to be undertaken in order to prevent the prediction from coming true. Moreover, potential shareholders make their investment decisions dependent on the performance indicators, such as sales. A reliable prediction of an increase in sales might, consequently, result in an easier access to capital for the company, as well as in an increase in its market value. Being able to accurately predict sales is therefore of significant importance.

Thanks to the rapid technological development taking place nowadays, companies have increasing possibilities of collecting and processing data. In the 'big data' era large data sets on companies' own activity, as well as on consumers' behaviour patterns and economic or demographic indicators are at the disposal of the forecasters. In case of sales data, they typically have a panel structure with information about some sales measure of various products in subsequent time periods. Employing these large panels for computing forecasts allows for a significant increase in the efficiency of the estimators (compared to individual time series models) and hence might lead to better predictions.

The main goal of this paper is to construct a dynamic panel data model that would deliver as accurate sales predictions as possible. Models of this class have received considerable attention in the literature. Most of the non-Bayesian studies on dynamic panel modelling, however, exclusively focus on consistently estimating the common parameters in the model, while treating the heterogeneous ones as a nuisance. Bayesian approaches, on the contrary, enable obtaining the distribution of the heterogeneous parameters. However, Markov Chain Monte Carlo (MCMC) sampling schemes used for this purpose might be slow for large data sets. An alternative approach has recently been proposed by Liu et al. (2016), whose focal point is primarily the estimation of the individual-specific parameters which, as they show, leads to an increase in the performance of the model. They consider an empirical-Bayes approach to estimate the correlated random effects in a linear dynamic panel model which does not employ MCMC sampling. This research builds on their insights to propose a set of Posterior Mean Panel Predictor (PMPP) models and to analyse their forecasting performance. This is done by means of a set of simulation studies and through an application to empirical data.

The empirical-Bayes approach pursued in this paper can be viewed as an approximation to the fully Bayesian methods. The difference is in the way hyper-parameters are

specified. While a fully Bayesian approach would be to specify the priors before observing the data, based on researcher's beliefs or previous studies, the empirical-Bayes method estimates prior distributions from the data. One may argue that plugging-in the estimates of hyper-parameters in the prior density ignores the uncertainty in these estimates. If the data set is large, however, these estimates do not depend much on any individual observation and might be deemed a good approximation.

The contribution of this paper to the existing literature is mainly four-folds. Firstly, while Liu et al. (2016) focus on one-step-ahead predictions only, this research extends the theoretical framework they propose to multi-step forecasting, providing appropriate formulae and assessing the model's performance over a long-run horizon.

Secondly, forecast uncertainty is taken into account, an issue Liu et al. (2016) ignore altogether, focusing only on the ex-post regret. This research proposes a simulation-based algorithm to calculate prediction intervals, which is of high importance especially in the long run, when accuracy of point predictions decreases.

Thirdly, this paper extends the model discussed by Liu et al. (2016) by proposing three additional estimation methods, which results in obtaining ten different implementations of the model. All of them are tested in a set of simulation studies as well as applied to empirical data.

Ultimately, this paper investigates whether the method proved by Liu et al. (2016) to enhance forecast accuracy in case of now-casting revenues of banks under changing economic conditions will also provide reinforcement for predicting future sales.

Moreover, this paper is accompanied by an R-package `pmpp` (posterior mean panel predictor) which contains a set of functions allowing to easily apply various implementations of the analysed model to any panel data set. The `pmpp` package also provides some functionalities that can be used outside of the scope of the model. A detailed description of the package can be found in the appendix.

The research conducted in this paper is relevant from both scientific and practical perspective. Researchers might be interested in the performance of various implementations of the model, as well as in the construction of the algorithm proposed for calculating prediction intervals, while business will find it decisive whether proposed methods are able to produce more accurate forecasts than the commonly employed models.

The remainder of this paper is structured as follows. Section 2 provides an overview of the existing studies on sales forecasting and dynamic panel modelling. Section 3 explains the methods for computation of point forecasts. Consecutive subsections introduce the analysed setting and tackle the issue of estimating two different classes of parameters in the model. Subsequently, Section 4 provides a simulation-based algorithm for obtaining prediction intervals. Next, in Section 5, three simulation studies to evaluate the model's performance are conducted, while Section 6 presents the application to empirical data. Section 7 concludes.

2 Literature review

This section provides a review of the existing literature on topics related to this research. Subection 2.1 covers various approaches to predicting sales, while Subsection 2.2 tackles the issue of dynamic panel modelling, including the paper of Liu et al. (2016) which constitutes the foundation for this research. The two fields receive separate treatments, as there is hardly any literature on forecasting sales with a specific focus on dynamic panel models.

2.1 Forecasting sales

Predicting future sales has been at the core of marketing research for a long time. Early literature in this field mostly relied on static techniques, such as regression or correlation analysis (see e.g. Pindyck and Rubinfeld (1976)), ignoring the dynamics or memory of the process. One of the pioneering dynamic approaches was the one by Winters (1960), who, in search of a quick and cheap method that could be used widely, proposed the Exponentially Weighted Moving Averages (EWMA) approach. This technique allows to remove the most volatile elements from a time series, which usually is the noise, to analyse the underlying process. Another approach gaining popularity at that time was the exponential smoothing introduced by Brown (1963). This method, however, involves a problem of arbitrarily selecting an appropriate weighting factor. The focus of the sales-forecasting literature then turned towards time-series analysis involving ARMA and error-correction models, see e.g. Kapoor et al. (1981) and the references therein.

Introduction of bar-code scanners has led to the availability of rich data bases. Wittink et al. (1988) took advantage of the new possibilities by introducing the SCAN*PRO model which focused mainly on promotion effectiveness. This model allows to accurately forecast sales as a function on various promotion indicators. Its generalisations take into account competitive response by considering sales of all brands in the analysed product category.

One of the the more recent approaches is the CHAN4CAST model proposed by Divakar et al. (2005). It is a sales forecasting model for consumer packaged goods, predicting sales volume by pack size, category, channel, region and customer account. It captures the effects of past sales, trend, seasonality, own and competitor prices and promotional activity, as well as accounts for the effects of temperature, new product releases and trading day corrections. As shown by the authors, CHAN4CAST is easy to implement and predicts sales volume satisfactorily.

All the methods discussed so far base on statistical models. The literature proposes also another approach to sales forecasting: machine learning techniques based on neural networks. The pioneering paper in this field was the one by Thiesing and Vornberger (1997), who point out the two major weaknesses of statistical forecasting: firstly, for each problem a separate model needs to be chosen that makes some assumptions about

the analysed series; secondly, the power of deterministic data analysis is low for multidimensional time series with mutual, nonlinear relations. They propose a neural network approach based on the idea of training a multi-layer network by a supervised training algorithm. The authors apply their model to empirical sales data to show that its forecasting accuracy outperforms commonly employed statistical models.

A more recent machine learning approach to sales forecasting was developed by Lu and Shao (2012). They propose a hybrid forecasting model that makes use of ensemble empirical model decomposition (EEMD) and extreme learning machine (ELM). The EEMD is a method for decomposing a signal into the so-called intrinsic mode functions (IMFs). The ELM is a recently developed learning algorithm for single-hidden-layer feed-forward networks. The authors propose to first apply the EEMD to the analysed time series to obtain a number of IMFs, which are expected to contain hidden useful information about the original data series. Second, they are integrated with use of the ELM algorithm to obtain the forecasting model. The authors show that this approach outperforms competitive forecasting methods.

2.2 Dynamic panel modelling

Dynamic panel data models have received considerable attention in the literature. It dates back to Goldberger (1962), who proposed a method based on the best linear unbiased prediction (BLUP) allowing to estimate random effects. A broad overview of various applications of BLUP can be found in Robinson (1991). He shows how BLUP can be used to derive the Kalman filter, remove noise from images or perform small-area estimation and points out other applications in insurance, credibility theory and geostatistics. For an excellent review of the panel forecasting literature, including BLUP-based mathods as well as models for heterogeneous panels and Bayesian approach, see the survey article by Baltagi (2008).

There is also a vast literature addressing the consistent estimation of the auto-regressive parameters in AR panel models with fixed effects when the time dimension is small. Different estimators based on the generalised method of moments (GMM) or instrumental variables (IV) have been proposed by e.g. Anderson and Hsiao (1981), Arellano and Bond (1991), Arellano and Bover (1995), Blundell and Bond (1998), Alvarez and Arellano (2003) or Youssef and Abonazel (2015). They will be now discussed in greater detail.

In case of a micro-panel, i.e. a panel with a large number of cross-sectional units and a small number of time periods, the asymptotic results of the static panel model are not generally valid for models that include lagged response in the set of explanatory variables. The usual GLS estimator for random effects and within estimator for fixed effects are inconsistent. The within estimator is additionally biased because of the correlation between the within-transformed regressor and residuals. This poses the need for a different ap-

proach in dynamic panel modelling. The earliest solution was proposed by Anderson and Hsiao (1981). They considered a first-difference regression

$$\Delta y_{i,t} = \gamma \Delta y_{i,t-1} + \Delta \varepsilon_{i,t}, \tag{1}$$

where Δ denotes the first-difference operator. The OLS applied to (1) would yield inconsistent estimates of γ, as $\Delta y_{i,t-1}$ is correlated with $\Delta \varepsilon_{i,t}$. Anderson and Hsiao (1981) proposed to use an IV approach, which relies on finding an instrument that is correlated with the regressor, but uncorrelated with the disturbance. They note that both $y_{i,t-2}$ and $\Delta y_{i,t-2}$ fulfil these conditions. This approach can just as well be viewed as GMM estimation: the fact that the instrument is uncorrelated with the residuals can be used as orthogonality condition in the form of e.g. $\mathrm{E}[\Delta \varepsilon_{i,t} \Delta y_{i,t-2}] = 0$. Thus, an estimator is obtained that is consistent even if T is fixed and finite, as long as $N \to \infty$.

The argumentation of Anderson and Hsiao (1981) was further extended by Arellano and Bond (1991), who noted that although consistent, the Anderson-Hsiao estimator is not efficient, as it only makes use of a limited number of time periods. They generalise the orthogonality condition of Anderson-Hsiao estimator to $\mathrm{E}[\Delta y_{i,t} y_{i,t-j} - \gamma \Delta y_{i,t-1} y_{i,t-j}] = 0$, where t = 2, ..., T and j = 2, ..., t. This results in 1/2(T-1)T orthogonality conditions compared to Anderson-Hsiao's T-1, which yields a substantial increase in efficiency. In the Arellano-Bond approach the number of restrictions in so longer equal to the number of parameters, which makes it impossible to obtain a single solution. They propose a standard GMM approach for such cases, i.e. to minimise a quadratic loss function with well-defined weights. Thus obtained estimator is consistent and asymptotically normally distributed.

A typical concern associated with IV estimation is about the validity and strength of the instruments. Although valid, the instruments used in Arellano-Bond estimation can be weak, as shown by Blundell and Bond (1998). They proved that this is the case if the auto-regressive process is very persistent or if the ratio of the variance of the panel-level effect to the variance of the idiosyncratic error becomes too large. Extending the framework of Arellano and Bover (1995), Blundell and Bond (1998) proposed a GMM-based estimator in which, in addition to instrumenting the first-differenced equation (1), lagged differences are used as instruments for the level equation. Validity of these additional moment conditions requires that $\mathrm{E}[\lambda_i \Delta y_{i,2}] = 0$, where λ_i denotes the panel-level effects, holds for all units i. For a chapter-length treatment of the topic see Blundell et al. (2000).

The discussion on dynamic panels was further developed by Alvarez and Arellano (2003), who investigated the panels with both time and unit dimension large, as they note that such panels are increasingly available nowadays. They allow both N and T go to infinity and study behaviour of various estimators for alternative relative rates of increase in both dimensions. Ultimately, they propose a random effects ML estimator

that leaves the moments of initial conditions unrestricted but enforces homoskedascticity in the time dimension. They show this estimator to be asymptotically unbiased for large N and T, unlike the GMM estimators that entail incidental parameters.

More recent literature in the field attempts to improve the efficiency of the Blundell-Bond estimator by proposing different weighting matrices used by the GMM estimator to combine all moment conditions. Examples of this literature trend include Youssef et al. (2014) or Youssef and Abonazel (2015), who propose a sub-optimal system-GMM estimator and provide Monte Carlo evidence for efficiency improvement.

One of the most recent approaches to modelling dynamic panels is the empirical Bayes method developed by Liu et al. (2016), who aim at forecasting a collection of short time series. They note that the existing literature almost exclusively studies the estimation of the homogeneous parameters, treating the individual-specific ones as a distress. Liu et al. (2016), however, view the accurate estimation of the heterogeneous parameters as an essential ingredient for producing forecasts.

They remark that due to the small time dimension, none of the analysed time series in isolation can provide sufficient information to make inference about the distribution of heterogeneous parameters in the model. Hence, they use cross-sectional information in the sample in order to approximate this distribution, which subsequently serves as a prior for the individual-specific coefficients, allowing to enhance posterior inference. They consider a linear dynamic panel model with the unobserved individual heterogeneity in the following form:

$$Y_{i,t} = \lambda_i' W_{i,t-1} + \rho' X_{i,t-1} + \alpha' Z_{i,t-1} + U_{i,t}, \tag{2}$$

where i = 1, ..., N and t = 1, ..., T denote the numbers of cross-sectional observations and time periods respectively. The individual heterogeneity parameter λ_i is allowed to interact with a set of cross-sectionally varying and strictly deterministic explanatory variables, W. Vector W might also contain deterministic time effects, such as trend or seasonality. In a special case when $W = 1$, λ_i denotes an individual-specific intercept. The predictors in X may include lagged values of the dependent variable and other sequentially exogenous variables. Vector Z contains a set of strictly exogenous explanatory variables and U is an unpredictable error term, assumed to be given by

$$U_{i,t} = \sigma_t V_{i,t}, \tag{3}$$

where $V_{i,t} \sim iid(0,1)$ are independent over time and σ_t is a function that depends on the data and on an unknown parameter γ_t.

As the benchmark for evaluating their model, Liu et al. (2016) employ the so-called oracle forecast, an infeasible prediction assumed to know the common parameters as well as the distribution of the heterogeneous ones, but not their specific values. They show that the oracle replaces the value of the individual-specific parameter by its posterior

mean and, consequently, the optimal forecast can be written as a sum of this posterior mean and the auto-regressive and exogenous terms, i.e.

$$\hat{Y}_{i,T+1}^{opt} = \mathrm{E}[\lambda_i]' W_{i,T+1} + \rho' X_{i,T} + \alpha' Z_{i,T}, \tag{4}$$

where $\mathrm{E}[\lambda_i]$ is the posterior mean of λ_i.

To obtain a feasible predictor, they adjust the classical posterior mean formula, called the Tweedie's formula after the statistician Maurice Tweedie, to fit the dynamic panel setting. According to this formula, the posterior mean of λ_i can be expressed as a function of the cross-sectional density of a certain sufficient statistic. Liu et al. (2016) propose two ways of approximating the posterior mean: a parametric approach assuming Gaussian distribution of the random effects and a non-parametric method based on kernel density estimation. Similarly, they offer two ways of estimating the common parameters: a quasi-maximum likelihood (QMLE) approach assuming normality of the error terms and the GMM estimation employing the approach by Arellano and Bover (1995). Liu et al. (2016) show that the feasible predictor obtained this way can achieve the same compound risk (asymptotically) as the oracle forecast.

Subsequently, they conduct a Monte Carlo study to compare the performance of various implementations of their predictor as well as apply their technique to empirical data, attempting to now-cast pre-provision net-revenues of a panel of banks. They find that the forecasting method they propose is considerably more accurate than a predictor that does not employ any prior distribution. They also show that their posterior mean predictor shrinks the estimates of the heterogeneous parameters towards a common prior mean, thus reducing its sampling variation.

3 Methods for point forecast construction

This section discusses the methods for obtaining point forecasts with the analysed posterior mean predictor. Subsection 3.1 introduces the model and shows how the original proposal by Liu et al. (2016) can be extended to multi-step forecasting. Next, in Subsection 3.2, a brief review of the Tweedie's Formula is provided. Subsection 3.3 discusses both parametric and non-parametric estimation methods for the heterogeneous parameters, while Subsection 3.4 tackles the issue of estimating the homogeneous ones, considering a set of different QMLE and GMM-based methods.

3.1 Linear Dynamic Panel Model

The model considered in this paper is based on model (2). It is a linear dynamic panel data model with unobserved individual heterogeneity given by:

$$Y_{i,t} = \lambda_i' W_{i,t-1} + \rho' Y_{i,(t-1:t-p)} + \alpha' Z_{i,t-1} + U_{i,t}, \tag{5}$$

where the autoregressive vector Y of length k_y on the right-hand side of the equation contains lags of the dependent variable from time t-1 up to t-p, which makes p the order of the autoregressive polynomial. $W_{i,t}$ is a vector of length k_w, that may contain time-effects such as deterministic trends or seasonality, or strictly exogenous variables expected to interact with the heterogeneous coefficients λ_i. Additionally, let $W^*_{i,t}$ denote the sub-vector of $W_{i,t}$ of length k_{w^*} with the cross-sectionally varying exogenous variables only. All the remaining strictly exogenous variables are gathered in the vector $Z_{i,t}$ of length k_z. The error term $U_{i,t}$ is assumed to be given by equation (3), i.e. $U_{i,t} = \sigma_t V_{i,t}$, where $V_{i,t} \sim iid(0,1)$ are independent over time and σ_t is a function that depends on the data and on an unknown parameter γ_t. Thus, $U_{i,t}$ is allowed to be conditionally heteroskedastic both over time (through γ_t) and in the cross-section (through $V_{i,t}$).

Estimation of model (5) involves dealing with two classes of parameters. First, there are three vectors of homogeneous parameters that we shall collect into one vector $\theta = [\alpha', \rho', \gamma']'$, where γ denotes the volatility-controlling parameters stacked into one vector: $\gamma = [\gamma_1', ..., \gamma_T']'$. Second, there is one vector of heterogeneous parameters λ_i. For further reference, the conditioning exogenous variables shall be defined as $H_i = (Y_i, W_i^*, Z_i)$ and their realisations as $h_i = (y_i, w_i^*, z_i)$. Finally, let $\varphi(v)$ denote the density of V_i. Under a set of reasonable assumptions discussed in Liu et al. (2016), the crucial one being that conditional on the predictors the observations are cross-sectionally independent, the parameters α, ρ and γ as well as the correlated random effects distribution $\pi(\lambda_i|h_i)$ in model (5) can be shown to be identified.

Noting from equation (3) that $V_{i,t} = U_{i,t}/\sigma_t$, one can combine the likelihood function for the observables with the distribution of the heterogeneous parameters conditional

on the common ones. Next, as the observations are assumed to be cross-sectionally independent conditional on the predictors, the joint density of all observations can be obtained as a product across all observations. This yields

$$p(y, \lambda|h, \theta) \propto \prod_{i=1}^{N} \left(\left(\prod_{t=1}^{T} \frac{1}{\sigma_t(h_i, \gamma_t)} \varphi\Big(\frac{y_{i,t} - \lambda_i' w_{i,t-1} - \rho' y_{i,(t-1:t-p)} - \alpha' z_{i,t-1}}{\sigma_t(h_i, \gamma_t)} \Big) \right) \pi(\lambda_i|h_i) \right). \quad (6)$$

Forecasting with the above model is straightforward. Building on the argumentation of Liu et al. (2016) regarding forecast optimality and the compound risk it attains (see Section 2 of their paper), one can use the recursive formula to write the infeasible optimal forecast for autoregression order $p = 1$ as

$$\hat{Y}_{i,T+h}^{opt} = \mathrm{E}[\lambda_i] \left(\sum_{s=0}^{h-1} \rho^s W_{i,T+h-1-s} \right) + \rho^h Y_{i,T} + \alpha' \left(\sum_{s=0}^{h-1} \rho^s Z_{i,T+h-1-s} \right), \quad (7)$$

where $\mathrm{E}[\lambda_i]$ denotes the mean of the posterior distribution of λ_i. For each $p > 1$ the closed-form expression becomes different, as the parameters $\rho_1, ..., \rho_p$ interact with each other. However, it is still possible to analytically derive the formula for each forecast horizon h. Assuming that the values of the explanatory variables in W and Z at time $T+h$ are known (e.g. provided by external forecasts), the optimal forecast in (7) can be approximated by replacing the posterior mean of λ_i by its estimate based on the Tweedie's formula and replacing the common parameters by some consistent estimates.

The estimation procedure is a two-step one. In the first step, ρ and α are estimated using some standard dynamic panel approach, see Section 3.4. These methods typically do not yield an estimate of the heterogeneous intercept, which is being removed from the model, e.g. by first-differentiating. The analysed posterior mean model assumes, however, that if there exists true heterogeneity across observations, adding individual-specific fixed effects might enhance prediction accuracy. Therefore, the second step conditions on the first-step estimates to produce individual-specific intercept parameters λ_i, see Section 3.3.

3.2 Tweedie's Formula

Tweedie's Formula, first introduced by Robbins (1951) as a way to estimate a vector of means for a normal distribution, has been extended by Effron (2011) to the family of exponential distributions and more recently by Liu et al. (2016) to the estimation of correlated random effect parameters in the analysed panel setting. It allows to analytically derive a formula for posterior expectation of the heterogeneous parameters assuming normal distribution of the error terms, i.e. $V_{i,t} \sim N(0,1)$.

For the sake of clear notation, define

$$\tilde{y}_t(\theta) = y_t - \rho' y_{(t-1:t-p)} - \alpha' z_{t-1}, \tag{8}$$

$$\Sigma(\theta) = diag(\sigma_1^2, ..., \sigma_T^2), \tag{9}$$

and let $\tilde{y}$ and w be matrices with rows given by (8) and w'_{t-1} respectively. Substituting $\varphi(v)$ in the joint likelihood function (6) with a standard normal density function yields

$$\begin{aligned} & p(y, \lambda | h, \theta) \\ & \propto \quad exp\bigg(-\frac{1}{2}(\hat{\lambda} - \lambda)' w' \Sigma^{-1} w (\hat{\lambda} - \lambda) \bigg) exp\bigg(-\frac{1}{2}(\tilde{y} - w\hat{\lambda})' \Sigma^{-1} (\tilde{y} - w\hat{\lambda}) \bigg) \pi(\lambda | h). \end{aligned} \tag{10}$$

Liu et al. (2016) use the Fisher-Neyman factorisation theorem to show that

$$\hat{\lambda} = (w' \Sigma^{-1} w)^{-1} (w' \Sigma^{-1} \tilde{y}) \tag{11}$$

is a sufficient statistic for λ. As a consequence, the posterior distribution of λ can be expressed as

$$p(\lambda | \hat{\lambda}, h, \theta) = \frac{p(\hat{\lambda} | \lambda, h, \theta) \pi(\lambda | h)}{p(\hat{\lambda} | h, \theta)}, \tag{12}$$

where the likelihood function of the sufficient statistic is given by

$$p(\hat{\lambda} | \lambda, h, \theta) = (2\pi)^{-k_w/2} |w' \Sigma^{-1} w|^{1/2} exp\bigg(-\frac{1}{2}(\hat{\lambda} - \lambda)' w' \Sigma^{-1} w (\hat{\lambda} - \lambda) \bigg). \tag{13}$$

Because equation (12) represents a density function, it integrates to one by definition. As the mean of the normal distribution is attained at the maximum of its density function, one can simply differentiate the equation $\int p(\lambda | \hat{\lambda}, h, \theta) d\lambda - 1 = 0$ with respect to $\hat{\lambda}$ to obtain the expression for the mean. First order conditions imply the Tweedie's formula for the posterior mean, expressed as the sum of the sufficient statistic and the correction term. It is given by

$$\mathrm{E}[\lambda_i] = \hat{\lambda}_i(\theta) + \bigg(W_i' \Sigma^{-1}(\theta) W_i \bigg)^{-1} \frac{\partial}{\partial \hat{\lambda}_i(\theta)} ln \bigg(p(\hat{\lambda}_i(\theta), H_i | \theta) \bigg). \tag{14}$$

Substituting the above formula for $\mathrm{E}[\lambda_i]$ into equation (7) yields the feasible optimal forecast. First, however, the common parameters θ and the joint density $p(\hat{\lambda}_i(\theta), H_i | \theta)$ in the correction term in (14) need to be estimated.

3.3 Estimation of Tweedie's correction term

Following Liu et al. (2016), this paper considers two different approaches to approximating the density $p(\hat{\lambda}_i(\theta), H_i|\theta)$ in the Tweedie's formula: a parametric, Gaussian approach and a non-parametric, kernel-based method are discussed in subsections 3.3.1 and 3.3.2 respectively.

3.3.1 Parametric estimation

Assuming the random-effects distribution to be normal allows to analytically derive the marginal density of the sufficient statistic. Conditional on the common parameters, the moments of this distribution can then be estimated by maximizing the corresponding marginal likelihood function. Let

$$\lambda_i|(H_i, \theta) \sim N(\Phi H_i, \underline{\Omega}) \tag{15}$$

and let the moments of this distribution be gathered in a hyperparameters vector $\xi = (vec(\Phi), vech(\underline{\Omega}))'$. To obtain the marginal density of $\hat{\lambda}$ in this setting, the Gaussian prior density in (15) can be plugged-in to the likelihood in (13). Integrating-out λ, this yields

$$p(\hat{\lambda}|h, \theta, \xi) = (2\pi)^{-k_w/2}|\underline{\Omega}^{-1}|^{1/2}|w'\Sigma^{-1}w|^{1/2}|\bar{\Omega}|^{1/2} \times exp\bigg(-\frac{1}{2}\Big(\hat{\lambda}'w'\Sigma^{-1}w\hat{\lambda} + h'\Phi'\underline{\Omega}^{-1}\Phi h - \bar{\lambda}'\bar{\Omega}^{-1}\bar{\lambda}\Big)\bigg), \tag{16}$$

where $\bar{\lambda}$ and $\bar{\Omega}$ denote the posterior mean and variance of λ respectively and are given by

$$\bar{\Omega}^{-1} = \underline{\Omega}^{-1} + w'\Sigma^{-1}w, \tag{17}$$

$$\bar{\lambda} = \bar{\Omega}\big(\underline{\Omega}^{-1}\Phi h + w'\Sigma^{-1}w\hat{\lambda}\big). \tag{18}$$

The hyperparameters vector ξ can be estimated by maximising the marginal likelihood

$$\hat{\xi} = \arg\max_{\xi} \prod_{i=1}^{N} p(\hat{\lambda}_i|h_i, \theta, \xi). \tag{19}$$

Thus obtained estimates can then be plugged-in to the density in (16). This density can then be used to evaluate the Tweedie's formula.

3.3.2 Kernel density estimation

Another approach to evaluating the marginal density of the individual-specific intercept is a non-parametric one, employing kernel density estimation. The kernel density estimator, first introduced by Rosenblatt (1956), is a method to obtain an estimate of the density

function of a random variable. It can be viewed as a smoother, continuous version of a histogram. If one considers a histogram-based density estimate as giving the same weight to all observations within the same bin, then a kernel-based estimate generalises it by using some alternative weighing function, called a kernel function. A choice akin to choosing the number of bins in a histogram is to choose the bandwidth - a parameter constituting a trade-off between the level of smoothness and the estimation bias. In case of assuming a Gaussian kernel, the optimal bandwidth is typically chosen as to minimise the integrated mean-squared error, following the so called Silverman's rule of thumb (Silverman (1986)).

In order to obtain the density of the individual-specific intercept conditional on the common parameters, a two-dimensional density needs to be estimated. This paper considers the one proposed by Botev et al. (2010). It is an adaptive method based on linear diffusion processes. It assumes a Gaussian kernel and uses a plug-in bandwidth selection method that is free from any arbitrary normal reference rules, which results in greater accuracy and reliablility compared to other methods, as proved by the authors.

3.4 Estimation of homogeneous parameters

This paper considers five different methods for estimating the common parameters θ. First, under the assumption that the error terms U_{it} in (5) are normally distributed, one can derive the density of the dependent variable conditional on the common parameters. The parameters can be estimated by maximizing the corresponding quasi-likelihood function, yielding the QMLE estimator, as shown in Subsection 3.4.1. Second, in case one does not want to make distributional assumptions about the model's error terms, it is possible to estimate the common parameters using a GMM estimator based on the sample analogues of the appropriate moment conditions. Four different variants of this approach are discussed in Subsections 3.4.2 through 3.4.5.

3.4.1 Quasi-Maximum Likelihood Estimator

Assuming the error terms in the model are normally distributed, the joint density of the data y and the heterogeneous coefficients λ has the representation given by (10), where $\pi(\lambda|h)$ is the random effect density. Integrating-out λ under a parametric random effect distribution yields (omitting the subscripts i and the function arguments θ)

$$\begin{aligned} p(y|h,\theta,\xi) = & \int p(y|h,\theta,\lambda)\pi(\lambda|h,\hat{\xi})d\lambda \\ \propto \quad & |\Sigma|^{-1/2} exp\bigg(-\frac{1}{2}(\tilde{y}-w\hat{\lambda})'\Sigma^{-1}(\tilde{y}-w\hat{\lambda})\bigg) \\ \times \quad & |w'\Sigma^{-1}w|^{-1/2} p(\hat{\lambda}|h,\theta,\xi). \end{aligned} \tag{20}$$

Because the density on the right-hand side of the bottom row in (20) is identical to the objective function that is maximised to obtain ξ in (19), θ and ξ can be estimated jointly by maximising the integrated likelihood as

$$(\hat{\theta}, \hat{\xi}) = \arg\max_{\xi,\theta} \prod_{i=1}^{N} p(y|h_i, \theta, \xi). \tag{21}$$

Because of the use of proportionality in (20), the function maximised in (21) is not exactly the likelihood function, but rather its simplification. For this reason the maximum-likelihood estimator applied to this function is referred to as a *quasi*-ML estimator.

3.4.2 Arellano-Bond Estimator (Difference-GMM)

Arellano and Bond (1991) propose a GMM setting for IV-estimation of the parameters in dynamic panel data models. For the sake of notation simplicity, we inspect the case when $p = 1$. Consider the first differenced model:

$$\Delta Y_{i,t} = \rho \Delta Y_{i,t-1} + \alpha \Delta Z_{i,t-1} + \Delta U_{i,t}. \tag{22}$$

The moment conditions proposed by Anderson and Hsiao (1981) can be generalised to $\mathrm{E}[\Delta Y_{i,t} Y_{i,t-j} - \rho \Delta Y_{i,t-1} Y_{i,t-j}] = 0$, where $t = 2, ..., T$ and $j = 2, ..., t$ (for a more detailed explanation see the literature review in Section 2 of this paper). This estimator uses the fact that the number of valid instruments increases with time. Due to the fact that the moment conditions are based on the estimated first-differences of the error terms, it is sometimes referred to as Difference-GMM. Variables used to instrument $Y_{i,t-1}$ in (22) can be written in the matrix form as

$$Z_{DIFF,i} = \begin{pmatrix} Y_{i,0} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & Y_{i,0} & Y_{i,1} & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & Y_{i,0} & Y_{i,1} & Y_{i,2} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & Y_{i,0} & \cdots & Y_{i,T-2} \end{pmatrix}, \tag{23}$$

where the first row of Z_{DIFF} contains instruments for time $t = 2$, the second row for $t = 3$ and so on. This notation allows to summarise the moment conditions as

$$\mathrm{E}\left[\begin{pmatrix} Z'_{DIFF,i} \\ \Delta Z'_{i,-1} \end{pmatrix} (\Delta Y_i - \rho \Delta Y_{i,-1} - \alpha \Delta Z_{i,-1}) \right] = 0, \tag{24}$$

where $\Delta Y_i = [\Delta Y_{i,2}, ..., \Delta Y_{i,T}]'$ and $\Delta Y_{i,-1} = [\Delta Y_{i,1}, ..., \Delta Y_{i,T-1}]'$, with $\Delta Z_{i,-1}$ defined accordingly. The estimates of the parameters ρ and α are found by minimising a quadratic

criterion in the sample analogue of (24):

$$g(\rho,\alpha) = \frac{1}{N}\sum_{i=1}^{N}\left(\begin{pmatrix} Z'_{DIFF,i} \\ \Delta Z'_{i,-1} \end{pmatrix}(\Delta Y_i - \rho\Delta Y_{i,-1} - \alpha\Delta Z_{i,-1})\right),$$
$$(\hat{\rho},\hat{\alpha}) = \arg\min_{\rho,\alpha} g(\rho,\alpha)' W g(\rho,\alpha), \tag{25}$$

where W is some positive semi-definite weighting matrix, specifying relative importance of each condition. The optimal choice of W, yielding an efficient estimate, is the inverse of the asymptotic covariance matrix of the sample analogue of the moment conditions:

$$W^{opt} = \left(\lim_{N\to\infty} N\cdot \mathrm{E}[g(\rho,\alpha)g(\rho,\alpha)']\right)^{-1}. \tag{26}$$

Based on the estimates $\hat{\rho}$ and $\hat{\alpha}$, variance of the error terms can be estimated based on the in-sample variance of fitted residuals.

3.4.3 Arellano-Bover Estimator (Level-GMM)

Another approach to estimating dynamic panel models is the one of Arellano and Bover (1995). It is based on the Orthogonal Forward Deviations (FOD) transformation, which subtracts all available future observations from the current value. This estimator is sometimes referred to as Level-GMM, as, contrary to Difference-GMM, its moment conditions are based on level-values of the error terms. First, all variables from the level model undergo the FOD transformation, which yields

$$Y^*_{i,t} = Y_{i,t} - \left(\sum_{s=t+1}^{T} Y_{i,s}W'_{i,s-1}\right)\left(\sum_{s=t+1}^{T} W_{i,s-1}W'_{i,s-1}\right)^{-1} W_{i,t-1}, \tag{27}$$

where $t = 1,...,T-1$ with $Y^*_{i,t-1}$ and $Z^*_{i,t-1}$ defined accordingly. Next, the moment conditions are set up as

$$\mathrm{E}\left[(Y^*_{i,t} - \rho' Y^*_{i,t-1} - \alpha' Z^*_{i,t-1})\begin{pmatrix} Y_{i,(0:t-1)} \\ Z_{i,(0:T)} \end{pmatrix}\right] = 0. \tag{28}$$

The estimation procedure is analogous with the one employed for Difference-GMM.

3.4.4 Blundell-Bond Estimator (System-GMM)

An extension of Difference-GMM has been proposed by Blundell and Bond (1998), who noted that the instruments it employs can become weak in some circumstances. They showed this is the case if the data generating process is close to being a random walk or

if the ratio of the variance of the panel-level effect to the variance of the idiosyncratic error becomes too large. As a solution, they propose to extend the list of instruments by adding lagged differences of the response variable as instruments for the level equation. In this setting, the matrix of instruments is given by

$$Z_{SYS,i} = \begin{pmatrix} Z_{DIFF,i} & 0 & 0 & \cdots & 0 \\ 0 & \Delta Y_{i,2} & 0 & \cdots & 0 \\ 0 & 0 & \Delta Y_{i,3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \Delta Y_{i,T-1} \end{pmatrix} \tag{29}$$

and the moment conditions can be written as

$$\mathrm{E}\left[Z'_{SYS,i} \begin{pmatrix} \Delta Y_i - \rho \Delta Y_{i,-1} - \alpha \Delta Z_{i,-1} \\ Y_i - \rho Y_{i,-1} - \alpha Z_{i,-1} \end{pmatrix} \right] = 0. \tag{30}$$

Because the original level model includes unit-specific effects λ_i, the additional instruments are only valid under the initial condition $\mathrm{E}[\lambda_i \Delta Y_{i,2}] = 0$.

The estimation procedure is analogous with the Difference-GMM case. This estimator is sometimes referred to as System-GMM, as it combines difference- and level-instruments into one system. It has been shown to be more efficient than Difference-GMM, as long as the initial condition holds.

3.4.5 Sub-optimal System-GMM Estimator

The Blundell-Bond approach simply removes the heterogeneous intercept from the model by first-differentiating. For this reason, the optimal weighting matrix commonly applied in the GMM estimation is truly optimal only under the assumption of no individual effects, i.e. $\sigma^2_\lambda = 0$. However, if the ratio of the variance of the individual effects to the variance of the idiosyncratic error is large, i.e. $\sigma^2_\lambda / \sigma^2_U >> 0$, it is no longer the case.

As a solution, Youssef and Abonazel (2015) propose an improvement to the System-GMM estimator. They use the same set of instruments as Blundell and Bond, but a different weighting matrix in the GMM estimation. The sub-optimal weighting matrix they propose is given by

$$W_{SSYS} = \left(\frac{1}{N} \sum_{i=1}^{N} Z'_{SYS,i} G Z_{SYS,i} \right)^{-1}, \tag{31}$$

where $G = \begin{pmatrix} FF' & 0 \\ 0 & A_{T-2} \end{pmatrix}$, $F = \begin{pmatrix} -1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}$ and $A_{T-2} = I_{T-2} + \frac{\sigma_\lambda^2}{\sigma_U^2} l_{T-2} l'_{T-2}$

with I_{T-2} denoting an identity matrix of size $T-2$ and l_{T-2} a vector of ones of length $T-2$.

Because the true variance ratio $\sigma_\lambda^2/\sigma_U^2$ is unknown, one has to use some estimates. This research considers the ones proposed by Jung and Kwon (2007), which are given by

$$\hat{\sigma}_U^2 = \frac{\sum_{i=1}^{N}(\Delta\hat{U}_i'\Delta\hat{U}_i)}{2N(T-2)}, \qquad \hat{\sigma}_\lambda^2 = \frac{\sum_{i=1}^{N}(\tilde{U}_i'\tilde{U}_i - \Delta\tilde{U}_i'\Delta\tilde{U}_i/2)}{N(T-2)}, \tag{32}$$

where $\Delta\hat{U}_i$ are the residuals from the Difference-GMM estimator, while $\tilde{U}_i$ and $\Delta\tilde{U}_i$ are the residuals from, respectively, the first-difference and level equations in the System-GMM estimator.

Applying the above formulas yields a one-step estimator. Youssef and Abonazel (2015) suggest to use it to obtain fitted residuals $\hat{U}_i$ and then replace G in (31) by $\hat{U}_i\hat{U}_i'$, which yields a two-steps estimator. Repeating the process once more produces a three-steps estimator. In practice it is sometimes infeasible to obtain the three- or even two-steps estimator, as the matrix within the brackets in (31) is computationally singular. If this is the case in the simulation study in Section 5 or the empirical application in Section 6, the Moore-Penrose generalised inverse is used (Penrose (1955)).

Youssef and Abonazel (2015) show that the larger σ_λ^2, the larger the efficiency gains from using the Sub-optimal System-GMM estimator. For $\sigma_\lambda^2 = 0$ it can be shown to be numerically identical to the standard Blundell-Bond appraoch.

4 Methods for forecast uncertainty assessment

The previous section provides methods, whose main objective is to produce accurate point forecasts of a panel-structured data set. However, for the decision-makers employing forecasts to reach strategic business decisions it is equally important to know to what extent can they rely on the predictions they obtain. This section will, therefore, focus on computing prediction intervals.

The common approach to calculating prediction intervals in a regression setting makes use of the standard errors of the estimates of model's parameters and the correlations between them to asses the interval in which a future observation is likely to fall with a given probability. For the PMPP model analysed in this paper, however, this approach would be cumbersome, as one would need to take into account the relation between the two classes of parameters - the homogeneous ones, calculated following the rules of frequentist statistics, and the heterogeneous ones, computed in the Bayesian way conditional on the common parameters. Moreover, different implementations of the model would require separate calculation procedures, taking into account the particular estimators employed in each case.

A more generic approach is bootstrapping. It allows to compute prediction intervals by means of random sampling with replacement from a distribution of the model's residuals. This paper proposes a 2-Steps Random Window Block Bootstrap for prediction intervals. A theoretical introduction to the principles of block bootstraps in Section 4.1 is followed by the presentation of the algorithm itself in Section 4.2.

4.1 Block bootstrap theory

Originally proposed by Efron (1979), the bootstrap is a statistical tool allowing to approximate the distribution of a statistic by means of drawing with replacement from its observed sample. It can be used to obtain not only the point estimates, but also approximated distributions of the statistics of interest, allowing to analyse their variability. In the context of regression analysis, bootstrap methods can be used to sample from the model's residuals to obtain prediction intervals, see e.g. Davison and Hinkley (1997). This approach, however, can only be relied on under the assumption that the error terms are independently and identically distributed, which is hard to maintain in case of time-series data.

The block bootstrap is a method to improve bootstrap accuracy when handling time-series data. Its main idea is to divide data into several blocks and sample entire blocks of adjacent residuals instead of individual values. This allows to preserve the original serial correlation within each block. Under the assumption that data is weakly stationary and with a sufficiently large number of repeated draws, block bootstrap yields correct inference regarding prediction intervals.

Two main issues associated with employing a block bootstrap are the construction of the window and the choice of block's length. The former typically boils down to choosing between an overlapping and non-overlapping window. In the overlapping window setting, the same residual is part of multiple blocks. For instance, when sampling from residuals $\varepsilon_1, \varepsilon_2, ..., \varepsilon_T$, the overlapping blocks of length 3 are: $B_1 = [\varepsilon_1, \varepsilon_2, \varepsilon_3]$, $B_2 = [\varepsilon_2, \varepsilon_3, \varepsilon_4]$ and so on. In case of a non-overlapping window, each residual belongs to one block only. Continuing the above example, the blocks would now be: $B_1 = [\varepsilon_1, \varepsilon_2, \varepsilon_3]$, $B_2 = [\varepsilon_4, \varepsilon_5, \varepsilon_6]$ and so on. As shown by Lahiri (1999) or Andrews (2002), the two approaches perform similarly.

Far more important than window construction is the choice of block's length, to which the performance of a block bootstrap is highly sensitive. However, there are no rules to follow and the literature has seen multiple approaches. For instance, Hall and Horowitz (1996) use two lengths, 5 and 10, for sample sizes of $T = 50$ and $T = 100$ respectively, while Inoue and Shintani (2001) employ an automated procedure resulting in average block length of 3.5 and 6 for sample sizes of $T = 64$ and $T = 128$ respectively.

4.2 Random-window block bootstrap for prediction intervals

The bootstrap method proposed here draws to some extent from Li (2013), who develops a similar approach for vector-autoregression models. The main distinction is that his analysis is performed conditional on the last observed data point. The algorithm proposed below is free of such conditionality. It is also different from the standard approaches in that it makes use of a kind of an overlapping window, but with window location sampled randomly. Moreover, its two-step procedure allows to account for the sampling variability of parameter estimates (bullets (a) through (e) in the algorithm below), as well as for the one of the residuals ((f) through (g)).

Throughout this paper $R = 1000$ repetitions are used, while the block size b is selected as the integer closest to $T^{1/3}$, so that it is similar to the average block size of Inoue and Shintani (2001).

Algorithm (2-Steps Random Window Block Bootstrap):

1. Estimate model (5) to obtain fitted residuals $\hat{u}_{it}$ and parameter estimates: $\hat{\lambda}_i$, $\hat{\rho}$, $\hat{\alpha}$;
2. Define block length b;
3. For the number of repetitions $r = 1, ..., R$:
   (a) Sample j from a discrete uniform distribution over the interval $[1, N - b + 1]$, where N is the cross-sectional dimension of the data;
   (b) Build the first block, B_1, as a matrix of size $N \times b$, with the consecutive columns being residual vectors for all cross-sectional units: the first column at

time $t = j$, the second at $t = j + 1$ and so on:

$$B_1 = [\hat{u}_j, \hat{u}_{j+1}, ..., \hat{u}_{j+b-1}] = \begin{pmatrix} \hat{u}_{1,j} & \cdots & \hat{u}_{1,j+b-1} \\ \vdots & \ddots & \vdots \\ \hat{u}_{N,j} & \cdots & \hat{u}_{N,j+b-1} \end{pmatrix}; \tag{33}$$

(c) Repeat (a) and (b) to create consecutive blocks, until the total number of columns in all blocks equals or exceeds the time dimension of the data, T. Create the grand block B by stacking up the consecutive blocks column by column. If the number of columns of B is larger than T, remove the last columns to make them equal:

$$B = [B_1, B_2, ...] = [\hat{u}_{j_1}, ..., \hat{u}_{j_1+b-1}, \hat{u}_{j_2}, ..., \hat{u}_{j_2+b-1}, ...], \tag{34}$$

with $j_1, j_2, ...$ being consecutive draws of j and B is of size $N \times T$;

(d) Generate bootstrap replicates of the response variable, Y_{it}^B, according to

$$Y_{it}^B = \hat{\lambda}_i' W_{i,t-1} + \hat{\rho}' Y_{i,(t-1:t-p)} + \hat{\alpha}' Z_{i,t-1} + B_{i,t}, \tag{35}$$

with the residual $B_{i,t}$ being the (i, t) element of the grand block matrix B;

(e) Use bootstrap replicates Y_{it}^B to re-fit the model and obtain bootstrap coefficients: $\lambda_i^B, \rho^B, \alpha^B$;

(f) Obtain random blocks again: repeat (a) and (b), but this time only until the total number of columns in all blocks equals or exceeds the forecast horizon H. The result is the grand block forecast matrix B^F of size $N \times H$ (again remove last columns if necessary);

(g) Compute the r^{th} bootstrap forecast for each $h \in [1, H]$ according to:

$$\hat{Y}_{i,T+h}^{B,r} = \lambda_i^{B\prime} W_{i,T+h-1} + \rho^{B\prime} Y_{i,(T+h-1:T+h-p)} + \alpha^{B\prime} Z_{i,T+h-1} + B_{i,h}^F, \tag{36}$$

with the residual $B_{i,h}^F$ being the (i, h) element of the grand block forecast matrix B^F;

4. Choose the confidence level $\alpha \in (0, 1)$. Compute the upper and lower bounds for each h as the maximum and the minimum, respectively, of the set of R bootstrap forecasts with $R(1 - \alpha)/2$ largest and $R(1 - \alpha)/2$ lowest values removed.

5 Monte Carlo simulation studies

In this section, three simulation studies are conducted to assess the performance of the Posterior Mean Panel Predictor. The study in Section 5.1 assumes that the parameters λ_i are normally distributed and uncorrelated with the initial observation Y_{i0}. These assumptions are dropped in the second design in Section 5.2. Simulation in Section 5.3 is concerned with testing the accuracy of the block-bootstrapped prediction intervals.

5.1 Gaussian random effects design

The first Monte Carlo experiment is based on a simple panel model with individual-specific intercept, autoregressive order of $p = 1$ and no exogenous explanatory variables. Further, the intercept parameters are normally distributed and uncorrelated with the initial observation Y_{i0}. The innovations U_{it} and the intercept parameters λ_i have unit variances. Two values for the autoregressive coefficient are considered: $\rho \in \{0.50, 0.95\}$. The panel consists of 1000 cross-sectional observations and for the time dimension two values are considered: $T \in \{5, 15\}$. This experimental design is summarised in Table 1.

Table 1: Simulation Study Design 1: Gaussian Random Effects

Data Generating Process:	$Y_{it} = \lambda_i + \rho Y_{i,t-1} + U_{it}$
	where $U_{it} \sim N(0,1)$ and $\rho \in \{0.50, 0.95\}$
Initial observations:	$Y_{i0} \sim N(0,1)$
Random Effects:	$\lambda_i \vert Y_{i0} \sim N(\phi_0 + \phi_1 Y_{i0}, 1)$
	where $\phi_0 = \phi_1 = 0$
Forecast horizon:	$T + h$ where $h \in \{1, 2, 3, 4, 5\}$
Sample size:	$(N, T) \in \{(1000, 5), (1000, 15)\}$
Number of repetitions:	$N_{sim} = 1000$

This study compares the models' performance by means of the Root Mean Squared Error (RMSE) values when forecasting from one up to five time periods ahead. We compare 12 different implementations of the Posterior Mean Panel Predictor with 3 benchmark models. The PMPP variants are obtained by taking all combinations of methods for estimating common parameters (QMLE, Difference-GMM, Level-GMM, System-GMM and one- or three-steps Sub-optimal System-GMM) and heterogeneous parameters (parametric or kernel-based). The benchmark models are Difference-GMM, System-GMM and Pooled OLS. Table 2 shows a selection of interesting results. Detailed simulation outcomes have been relegated to the Appendix - see Table 8 in Appendix A.

Table 2: RMSE of the forecasts under Simulation Study Design 1: Gaussian Random Effects - a selection of results. Values lower than the lowest benchmark in each column are bolded. For detailed results see Table 8 in Appendix A.

	$T = 5$				$T = 15$			
	$\rho = 0.50$		$\rho = 0.95$		$\rho = 0.50$		$\rho = 0.95$	
Horizon:	+1	+5	+1	+5	+1	+5	+1	+5
PMPP:								
(QMLE, par)	**1.09**	**1.44**	**1.09**	**2.84**	**1.03**	**1.25**	**1.03**	**2.33**
(QMLE, kde)	**1.09**	**1.44**	**1.10**	**2.89**	**1.03**	**1.25**	**1.03**	**2.34**
(DIFF, par)	1.18	1.62	1.17	**3.40**	**1.06**	**1.32**	3.85	9.00
(DIFF, kde)	**1.12**	**1.49**	**1.12**	**3.02**	**1.04**	**1.26**	2.44	6.06
(SYS, par)	**1.11**	**1.55**	1.15	5.45	**1.06**	**1.33**	**1.09**	**2.81**
(SYS, kde)	1.13	1.62	1.20	6.26	**1.06**	**1.29**	**1.05**	**2.47**
(LEV, par)	1.16	1.67	**1.11**	**3.66**	1.16	**1.64**	**1.09**	**3.36**
(LEV, kde)	1.18	2.99	**1.14**	**4.26**	1.32	4.74	**1.11**	**3.35**
(SSYS3, par)	1.24	1.96	1.41	5.39	**1.12**	**1.47**	**1.07**	**2.88**
(SSYS3, kde)	1.18	2.05	1.31	5.31	**1.12**	1.68	**1.08**	**3.03**
(SSYS1, par)	1.19	1.91	1.27	6.27	1.14	**1.53**	**1.14**	**4.02**
(SSYS1, kde)	1.15	2.03	1.24	6.26	1.13	2.01	**1.15**	**4.13**
DIFF	1.45	2.24	1.42	4.97	1.47	2.28	5.56	12.66
SYS	1.13	1.75	1.15	5.45	1.18	2.00	1.61	5.90
Pooled OLS	1.12	1.57	1.14	5.10	1.12	1.66	1.53	5.55

Among the three benchmark models, Pooled OLS performs best in most of the settings, i.e. it produces forecasts with the lowest RMSE. It is only beaten by Difference-GMM in the case of short sample, strong correlation and long forecast horizon. This is not surprising, considering that the intercept parameters are drawn from a standard normal distribution and thus most of their values lay in a narrow range. For a panel in which the true intercept values are equal for all observations, the OLS would estimate them most efficiently. Apparently the variation in these parameters in this simulation design in not large enough to result in forecast accuracy deterioration caused by OLS's inconsistency.

For each setting (i.e. a combination of sample length, auto-correlation strength and forecast horizon) there were at least four PMPP models able to beat the best of the benchmarks. The number of different PMPP implementations that outperform the benchmark increases with sample length and auto-correlation strength. Parametric QMLE is the most accurate model in each setting, closely followed by its kernel-based version.

In general, there seems to be no significant advantage of either parametric estimation of the Tweedie correction nor of the kernel-based approach over the other method. For QMLE estimation both perform similarly. In case of Difference-GMM, kernel-based implementation is more accurate, especially for high value of ρ. On the contrary, Level-GMM works best with parametric approach, especially for low ρ. System-GMM estimation

yields most accurate forecasts with the parametric approach for short sample and the kernel-based one for long sample. The differences become significant for high value of ρ. Both SSYS-GMM estimators show the opposite relation: kernel-based estimation prevails for small T and the parametric approach works better for large T, although the differences are smaller than for other estimators.

Finally, the outperformance of the best PMPP model over the best benchmark increases with h and ρ, suggesting that the gains from employing PMPP are largest for long-run forecasting, especially when the process is highly persistent. In all settings the advantage of PMPP over the benchmark is stronger for the longer sample and slightly weaker when T is small. This finding stays in contradiction to Liu et al. (2016), who argue that, in theory, the smaller T, the larger the gain from using a prior distribution.

5.2 Non-Gaussian correlated random effects design

The second Monte Carlo experiment uses the same data generating process (but now with only one value for the autoregressive parameter, $\rho = 0.50$), forecast horizon, sample size and number of repetitions as the first one. The major difference is that now random effects have a bimodal distribution and are allowed to be correlated with the initial observation. The design is summarised in Table 3.

Table 3: Simulation Study Design 2: Non-Gaussian Correlated Random Effects

Data Generating Process:	$Y_{it} = \lambda_i + \rho Y_{i,t-1} + U_{it}$
	where $U_{it} \sim N(0,1)$ and $\rho = 0.50$
Initial observations:	$Y_{i0} \sim N\big(\frac{1}{1-\rho}, (\frac{1}{1-\rho^2} + \frac{1}{(1-\rho)^2})\big)$
Random Effects:	$\lambda_i \vert Y_{i0} \sim N(\phi_0 + \phi_1 Y_{i0}, 1)$
	where $(\phi_0, \phi_1) = \begin{cases} \big(2, \frac{1-\rho^2}{1-\rho} \cdot (\frac{1-\rho^2}{(1-\rho)^2} + 1)^{-1} + 1\big), & p_1 = 0.5 \\ \big(0, \frac{1-\rho^2}{1-\rho} \cdot (\frac{1-\rho^2}{(1-\rho)^2} + 1)^{-1} - 1\big), & p_2 = 0.5 \end{cases}$
Forecast horizon:	$T + h$ with $h \in \{1, 2, 3, 4, 5\}$
Sample size:	$(N, T) \in \{(1000, 5), (1000, 15)\}$
Number of repetitions:	$N_{sim} = 1000$

The set of compared models and the comparison criteria are the same as in the first simulation study in Section 5.1. Similarly, detailed results have been relegated to the Appendix - see Table 9 in Appendix A - while Table 4 presents a selection of outcomes.

Table 4: RMSE of the forecasts under Simulation Study Design 2: Non-Gaussian Correlated Random Effects - a selection of results. Values lower than the lowest benchmark in each column are bolded. For detailed results see Table 9 in Appendix A.

	$T = 5$		$T = 15$	
Horizon:	+1	+5	+1	+5
PMPP:				
(QMLE, par)	1.91	**6.18**	1.68	3.92
(QMLE, kde)	8.71	42.22	4.14	18.50
(DIFF, par)	2.48	**4.55**	3.88	4.66
(DIFF, kde)	4.76	9.08	2.69	3.20
(SYS, par)	1.63	7.46	1.20	2.25
(SYS, kde)	1.88	9.70	1.24	3.09
(LEV, par)	**1.52**	**5.35**	1.20	2.26
(LEV, kde)	1.67	**6.46**	1.25	3.31
(SSYS3, par)	**1.50**	**5.49**	1.20	2.26
(SSYS3, kde)	1.82	7.88	1.24	3.11
(SSYS1, par)	**1.50**	**5.47**	1.20	2.27
(SSYS1, kde)	1.79	7.73	1.24	3.15
DIFF	6.82	13.11	11.19	13.53
SYS	1.55	6.83	1.15	1.60
Pooled OLS	1.56	6.91	1.17	1.75

In the setting of non-Gaussian and correlated random effects Pooled OLS is no longer the most accurate benchmark. It is beaten by System-GMM, which was expected, as OLS is highly inconsistent when the true parameter is far from being a constant. It is worth noting that the advantage of System-GMM over Pooled OLS increases with forecast horizon.

In contradiction to the Gaussian and uncorrelated random effects design, here PMPP models are able to beat the benchmark only in the settings with a short data dimension. In this case three of the models (parametric Level-GMM, parametric SSYS3 and parametric SSYS1) outperform the best benchmark for all forecast horizons, while three others (parametric QMLE, parametric Difference-GMM and kernel-based Level-GMM) do so only for a long horizon. Similarly to the previous simulation study, when T is small, the outperformance of the PMPP models over the benchmarks increases rapidly with forecast horizon. In the setting with long data dimension PMPP's underperformance also increases with forecast horizon, albeit at a slower pace.

To sum up, the second simulation experiment leads one to conclude that when the random effects are non-normal and correlated with the data, PMPP framework provides an improvement in forecast accuracy only if the analysed data has a short time dimension. This improvement, measured against the best benchmark model, increases sharply with forecast horizon.

5.3 Test for prediction intervals

A natural approach to evaluating the performance of prediction intervals in a simulation study would have been to perform a Monte Carlo experiment similar to the one from Section 5.1 and to calculate the percentage of times the true value falls in between the predicted bounds. This approach, however, turns out to be computationally infeasible in the analysed setting - the large number of bootstrap replications needed to obtain trustworthy approximation, combined with the large number of Monte Carlo replications immensely increase the computation time. Therefore, this section proposes an alternative approach to assess the performance of prediction intervals, which does not rely on repetitively simulated data.

Consider again the Gaussian Random Effects design from Section 5.1. The properties of the data generating process together with the lack of correlation between λ_i, Y_{i0} and the error term ensure that for all $t \neq 0$ it holds that $Y_{i,t} \sim N(\lambda_i + \rho Y_{i,t-1}, 1)$, with λ_i and ρ known. Thus, for each unit i and time period t, the 'true' confidence bounds can be calculated based on the quantiles of the normal distribution. Consequently, one can simluate the data only once and then compare these 'true' bounds with the bootstrapped bounds. The obvious drawback of this approach is that it is vulnerable to unrepresentativeness of the once simulated data. It can still, however, yield valuable insights into the performance of prediction intervals.

Table 5: Accuracy of bootstrapped prediction intervals for the parametric QMLE model. Data was simulated according to the Gaussian Random Effects design with $T = 5$ and 1000 bootstraps were used. The top panel present the Mean Absolute Error of approximation. The three bottom panels show percentages of specific cases occurring.

	$\rho = 0.50$					$\rho = 0.95$				
Horizon:	+1	+2	+3	+4	+5	+1	+2	+3	+4	+5
MAE, upper bound	1.51	0.95	1.30	1.02	1.21	1.12	1.83	2.64	3.22	4.10
MAE, lower bound	1.49	0.96	1.27	1.05	1.23	1.33	1.86	2.43	2.93	3.50
MAE, span	2.11	1.14	1.57	1.15	1.42	1.70	2.47	3.32	4.35	5.59
Lower bound too low	0.60	0.46	0.58	0.53	0.54	0.58	0.65	0.67	0.70	0.71
Lower bound too high	0.40	0.54	0.42	0.47	0.46	0.42	0.35	0.33	0.30	0.30
Upper bound too low	0.39	0.55	0.43	0.52	0.46	0.39	0.28	0.27	0.23	0.20
Upper bound too high	0.61	0.45	0.57	0.48	0.54	0.61	0.72	0.73	0.77	0.80
Interval too wide	0.76	0.45	0.64	0.51	0.58	0.67	0.85	0.90	0.96	0.96
Interval too narrow	0.24	0.55	0.36	0.49	0.42	0.33	0.15	0.10	0.04	0.04

Table 5 shows various accuracy measures of the bootstrapped prediction intervals, computed as discussed above. For the less persistent data generating process with $\rho = 0.5$

the mean absolute error of the approximations of lower and upper bounds is almost the same, which is desirable, as it indicates that the bootstrap is not systematically biased towards either upper or lower direction. There is no clear correlation between the absolute error value and the forecast horizon, which is also satisfying, as the bootstrap should take the horizon into account and it should not influence its accuracy. For both upper and lower bounds, the percentage of cases when they were over-predicted is, on average, close to the percentage of cases when they were under-predicted. This results in the estimated interval being too wide only slightly more often than too narrow. A significant increase in the number of bootstrap replications is, therefore, expected shrink all the pairs of probabilities to the desired $(0.5, 0.5)$ values.

Bootstrap accuracy is less satisfying for the highly persistent process with $\rho = 0.95$. In this case the mean absolute approximation error increases with the forecast horizon. The prediction interval tends to be over-estimated, with the prevailance of too wide intervals over the too narrow ones increasing sharply with the forecast horizon. The analysed bootstrap method seems not to be appropriate for highly persistent data series.

6 Empirical application

This section tests PMPP models on empirical data. The analysed data set describes the demand for paint in Brazil. It consists of 60 observations, each being a different type of paint available on the Brazilian market, and 70 time periods. Time dimension is monthly, from March 2011 through December 2016. For each paint type and time period, the number of litres of paint sold is provided.

THe model considered in this section explains the volume of paint sold by an individual-specific intercept and the response itself, lagged one period. There are no other explanatory variables. For the sake of computational feasibility, this section focuses on the parametric approach to obtaining the posterior mean predictor. Section 6.1 assesses prediction accuracy, while Section 6.2 evaluates the performance of prediction intervals.

6.1 Forecast accuracy

The large time dimension of the data allows to use the rolling window approach. Two window lengths, $T = 5$ and $T = 15$, are considered. In both cases, all models are evaluated on a number of subsamples. The k^{th} subsample encompasses observations $Y_{i,k}$ through $Y_{i,k+T-1}$, with $k \in [1, 60]$ and $k \in [1, 50]$ for $T = 5$ and $T = 15$, respectively. Each time the models are used to compute predictions up to 5 periods ahead and forecasts errors are averaged over subsamples. Additionally, as the time series are strictly non-stationary, the demeaning transformation is applied to each series. The performance of PMPP models is compared with the standard benchmarks from Section 5: Difference-GMM, System-GMM and Pooled OLS. The results, averaged over forecast horizons, are presented in Table 6.

Table 6: Average RMSE over 5 periods ahead forecasts from a rolling window sample with window length $T = 5$ and $T = 15$, for original and demeaned data series. Values lower than the lowest benchmark in each column are bolded. For detailed results see Tables 10 and 11 in Appendix B.

	$T = 5$		$T = 15$	
	original	demeaned	original	demeaned
PMPP:				
(QMLE, par)	828 152	684 326	**532 557**	**483 268**
(DIFF, par)	1 063 220	2 237 530	**584 178**	550 164
(SYS, par)	734 205	613 904	**561 690**	**496 145**
(LEV, par)	823 953	689 143	592 404	522 327
(SSYS3, par)	798 567	679 295	**580 407**	**514 153**
(SSYS1, par)	816 500	683 510	**580 901**	**514 851**
DIFF	3 884 057	8 412 915	2 241 478	1 810 361
SYS	714 424	643 966	586 380	529 703
Pooled OLS	672 042	588 395	586 589	521 389

On average, for the short window of $T = 5$ none of the PMPP models yielded more accurate predictions than Pooled OLS, which turned out to be the best benchmark. System-GMM was the most accurate among all the underperforming PMPP implementations. In case of a longer window of $T = 15$, on the contrary, four out of six PMPP implementations beat the benchmark on average, with QMLE being the most accurate model. This empirical evidence confirms some of the insights from the simulation study from Section 5 and contradicts the argumentation of Liu et al. (2016): gains from employing PMPP tend to be larger for a longer sample size. The simulations have shown it to be the case when the random effects are normally distributed. This assumption seems to be plausible to hold for the paint data: major factors expected to co-move with paint demand, such as new constructions rate, can be expected to affect all types of paint similarly, resulting in demands for individual paint types exhibiting little heterogeneity.

Moreover, the gains from using PMPP increase with the forecast horizon (see Tables 10 and 11 in Appendix B), which is also in accordance with the simulation results. It is worth noting that for a longer window the performances of 3- and 1-step SSYS estimators are roughly equal. The poor performance of Difference-GMM, both on its own as well as within PMPP framework, might be explained by the strong persistence of the process. Estimates of the auto-regressive parameter from various models (not reported) vary between 0.95 and 1.00. In this case Arellano-Bond estimator might collapse, as explained in Section 3.

For both window lengths the predictions are more accurate for demeaned data, both among PMPP and benchmark models. This was expected, as after demeaning the process is more linear and thus easier to describe by linear models. It is also worth noting that the gains from demeaning are larger for the shorter window compared to the longer one as well as for PMPP models compared to the benchamarks.

6.2 Forecast uncertainty

This section examines the performance of prediction intervals. For the QMLE model estimated in the previous section prediction bounds have been computed according to the algorithm described in Section 4.2. We focus only on the one most accurate PMPP implementation in order to preserve computational feasibility. Moreover, we consider the model evaluated on demeaned data, as bootstrap-based inference requires stationarity assumption. Table 7 shows the percentage of point forecasts that fall into the prediction interval at the 0.95 confidence level.

Table 7: Average percentage of point forecasts over the rolling window that fall into the estimated 0.95 prediction interval.

Horizon	$T = 5$	$T = 15$
+1	0.901	0.990
+2	0.938	0.981
+3	0.885	0.973
+4	0.909	0.972
+5	0.875	0.967
Average	0.902	0.977

The model using short time series, $T = 5$, produces too narrow prediction intervals, as too little observations fall into them. A positive feature is that the percentage of observation within the interval, although too low, stays constant over the forecast horizon. Because the bootstrap method relies on re-fitting the model multiple times, it could be expected that the setting using $T = 15$ observations will be more accurate, as this model delivered more accurate forecasts in the analysis in Section 6.1. In this setting, however, the bootstrapped interval is too wide, which, as shown in the simulation study in Section 5.3, is the case if the process is highly persistent. Yet, the average percentages of 0.977 and 0.902 observations within estimated 0.95 bounds are not unacceptable. Increasing the number of bootstrap replications or simulating data multiple times might yield results of higher accuracy. It is left for further research to investigate the performance of the algorithm for prediction intervals in more comprehensive studies.

7 Conclusion

The main purpose of this paper has been to construct a forecasting dynamic panel data model that would deliver as accurate predictions as possible. In order to achieve this goal, the empirical-Bayes approach by Liu et al. (2016) has been extended to propose a set of Posterior Mean Panel Predictor (PMPP) models. For the homogeneous parameters and the heterogeneous intercept, this paper proposed, respectively, five and two estimation methods, which yields ten combinations of the PMPP specification.

A series of simulation studies found that under the assumption of normality of the heterogeneous intercept, PMPP models outperform benchmark models in all settings, with the gains from employing PMPP being largest for long-run forecasting, especially if the data generating process is highly persistent. Contradictory to previous literature's findings, PMPP's advanage has been shown to be be stronger for the data with a large time dimension. However, when the random effects are non-normal and correlated with the data, PMPP framework is able to outperform benchmark models only for series with a short time dimension. The magnitude of this advantage also increases with the forecast horizon.

An application to empirical data has confirmed simulation results: the gains from employing PMPP have been proved to be largest for long time series and to increase with the forecast horizon. However, for the analysed data set, forecast accuracy improvement was only noted for the data with large time dimension.

In addition to the forecasting framework, this paper has also proposed a Random-Window Block Bootstrap algorithm to compute prediction intervals for PMPP models. Its two-step procedure allows to account for the sampling variability of both parameter estimates and residuals. A simulation study has found that the algorithm's accuracy is satisfying as long as the process is not very persistent, while it is less so for highly auto-correlated series.

An application to empirical data showed that the algorithm yields reliable, albeit slightly too narrow prediction intervals for the data with a short time dimension, while it is less reliable if the time dimension is larger.

Recapitulating, Posterior Mean Panel Predictor models have been proved to be a valuable addition to the existing set of forecasting tools in the panel setting. Under some conditions, they are able to deliver predictions of higher accuracy than the state-of-the-art models applied in this field nowadays. It is left for further research to investigate the performance of the bootstrap method for prediction intervals in a more comprehensive way. Follow-up studies might also find it interesting to test the PMPP framework on empirical data expected to exhibit more cross-sectional heterogeneity or to benchmark the PMPP's accuracy against fully Bayesian approaches.

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.

Andrews, W. (2002). Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators. *Econometrica*, 70(1):119–162.

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

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

Baltagi, B. H. (2008). Forecasting with panel data. *Journal of Forecasting*, 27(2):153–173.

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

Blundell, R., Bond, S., and Windmeijer, F. (2000). Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator. In Baltagi, B. H., editor, *Nonstationary Panels, Cointegrating Panels and Dynamic Panels*, page 53–92. Elsevier.

Botev, Z., Grotowski, J., and Kroese, D. (2010). Kernel density estimation via diffusion. *The Annals of Statistics*, 38(5):2916 – 2957.

Brown, R. G. (1963). *Smoothing, Forecasting and Prediction of Descrete Time Series*. Prentice Hall.

Davison, A. C. and Hinkley, C. V. (1997). *Bootstrap method and their application*. Cambridge University Press.

Divakar, S., Ratchford, B. T., and Shankar, V. (2005). CHAN4CAST: A multichannel, multiregion sales forecasting model and decision support system for consumer packaged goods. *Marketing Science*, 24(3):334–350.

Effron, B. (2011). Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602 – 1614.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7(1):1–26.

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association*, 57(298):369–375.

Hall, P. and Horowitz, J. (1996). Bootstrap critcal values for tests based on generalized-method-of-moments estimators. *Econometrica*, 64(4):891–916.

Inoue, A. and Shintani, M. (2001). Bootstrapping GMM estimators for time series. *Vanderbilt University, Department of Economics, Working Paper No. 01-W29R*.

Jung, H. and Kwon, H. (2007). An alternative system GMM estimation in dynamic panel models. *Hi-Stat Discussion Paper no. 217. Hitotsubashi University, Tokyo, Japan.*

Kapoor, S. G., Madhok, P., and Wu, S. M. (1981). Modeling and forecasting sales data by time series analysis. *Journal of Marketing Research*, XVIII:94–100.

Lahiri, S. (1999). Theoretical comparisons of block bootstrap methods. *The Annals of Statistics*, 27(1):386–404.

Li, J. (2013). Block bootstrap prediction intervals for vector autoregression. *Miami University, Department of Economics, Working Paper - 2013-04.*

Liu, L., Moon, H. R., and Schorfheide, F. (2016). Forecasting with dynamic panel data models. *PIER Working Paper No. 16022.*

Lu, C. and Shao, Y. E. (2012). Forecasting computer products sales by integrating ensemble empirical mode decomposition and extreme learning machine. *Mathematical Problems in Engineering*, 2012.

Penrose, R. (1955). A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, 51:406 – 413.

Pindyck, R. S. and Rubinfeld, D. L. (1976). *Econometric Models And Economic Forecasts.* McGraw-Hill.

Robbins, H. (1951). Asymptocially subminimax solutions of compound decision problems. *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, I.

Robinson, G. K. (1991). That BLUP is a good thing: the estimation of random effects. *Statistical Science*, 6(1):15–32.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, 27:832 – 837.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis.* London, Chapman and Hall.

Thiesing, F. M. and Vornberger, O. (1997). Sales forecasting using neural networks. *Proceedings ICNN'97, Houston, Texas, USA*, 4:2125 – 2128.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, 6:324–342.

Wittink, D. R., Addona, M., Hawkes, W., and Porter, J. (1988). SCAN*PRO: The estimation, validation and use of promotional effects based on scanner data. *Working paper, AC Nielsen, Schaumburg, Illinois, USA.*

Youssef, A. and Abonazel, M. (2015). Alternative GMM estimators for first-order autoregressive panel model: An improving efficiency approach. *MPRA Paper No. 68674; Forthcoming in: Communications in Statistics - Simulation and Computation.*

Youssef, A., Abonazel, M., and El-sheikh, A. (2014). New GMM estimators for dynamic panel data models. *International Journal of Innovative Research in Science, Engineering and Technology*, 3(10):16414 – 16425.

Appendix A Detailed simulation results

Table 8: RMSE in the Simulation Study Design 1: Gaussian Random Effects. Study design is described in Table 1. Values lower than the lowest benchmark in each column are bolded. SSYS3 and SSYS1 stand for three- and one-step Sub-optimal System GMM estimator from Section 3.4.5, while DIFF, LEV and SYS stand for Difference-, Level- and System-GMM estimators. PMPP shows different variants of Posterior Mean Panel Predictor, where the first term in parenthesis denotes estimation method for common parameters, while *par* and *kde* stand for parametric and kernel-based estimation of the intercept parameters respectively.

	$\rho = 0.50$					$\rho = 0.95$				
Horizon:	+1	+2	+3	+4	+5	+1	+2	+3	+4	+5
	$T = 5$									
PMPP:										
(QMLE, par)	**1.09**	**1.29**	**1.38**	**1.42**	**1.44**	**1.09**	**1.62**	**2.06**	**2.46**	**2.84**
(QMLE, kde)	**1.09**	**1.30**	**1.38**	**1.42**	**1.44**	**1.10**	**1.63**	**2.09**	**2.50**	**2.89**
(DIFF, par)	1.18	1.44	1.55	1.60	1.62	1.17	**1.81**	**2.39**	**2.91**	**3.40**
(DIFF, kde)	**1.12**	**1.34**	**1.42**	**1.47**	**1.49**	**1.12**	**1.68**	**2.17**	**2.61**	**3.02**
(SYS, par)	**1.11**	**1.33**	**1.43**	**1.50**	**1.55**	1.15	1.89	2.77	3.93	5.45
(SYS, kde)	1.13	1.37	1.49	1.56	1.62	1.20	2.05	3.10	4.48	6.26
(LEV, par)	1.16	1.43	1.55	1.62	1.67	**1.11**	**1.70**	**2.27**	**2.91**	**3.66**
(LEV, kde)	1.18	1.60	2.02	2.48	2.99	**1.14**	**1.80**	**2.49**	**3.29**	**4.26**
(SSYS3, par)	1.24	1.49	1.66	1.80	1.96	1.41	2.24	3.12	4.12	5.39
(SSYS3, kde)	1.18	1.43	1.62	1.81	2.05	1.31	2.10	2.96	3.98	5.31
(SSYS1, par)	1.19	1.46	1.62	1.76	1.91	1.27	2.15	3.18	4.51	6.27
(SSYS1, kde)	1.15	1.42	1.61	1.80	2.03	1.24	2.10	3.13	4.47	6.26
DIFF	1.45	1.89	2.09	2.19	2.24	1.42	2.40	3.31	4.17	4.97
SYS	1.13	1.39	1.54	1.66	1.75	1.15	1.89	2.77	3.93	5.45
Pooled OLS	1.12	1.36	1.46	1.52	1.57	1.14	1.84	2.66	3.72	5.10
	$T = 15$									
PMPP:										
(QMLE, par)	**1.03**	**1.18**	**1.23**	**1.24**	**1.25**	**1.03**	**1.47**	**1.80**	**2.09**	**2.33**
(QMLE, kde)	**1.03**	**1.18**	**1.23**	**1.25**	**1.25**	**1.03**	**1.47**	**1.80**	**2.09**	**2.34**
(DIFF, par)	**1.06**	**1.23**	**1.29**	**1.31**	**1.32**	3.85	6.07	7.44	8.35	9.00
(DIFF, kde)	**1.04**	**1.19**	**1.23**	**1.25**	**1.26**	2.44	3.87	4.83	5.52	6.06
(SYS, par)	**1.06**	**1.23**	**1.29**	**1.32**	**1.33**	**1.09**	**1.61**	**2.05**	**2.45**	**2.81**
(SYS, kde)	**1.06**	**1.23**	**1.27**	**1.29**	**1.29**	**1.05**	**1.50**	**1.87**	**2.19**	**2.47**
(LEV, par)	1.16	1.42	1.54	1.60	**1.64**	**1.09**	**1.65**	**2.19**	**2.75**	**3.36**
(LEV, kde)	1.32	2.06	2.89	3.79	4.74	**1.11**	**1.69**	**2.23**	**2.77**	**3.35**
(SSYS3, par)	**1.12**	**1.33**	**1.42**	**1.45**	1.47	**1.07**	**1.57**	**2.02**	**2.44**	**2.88**
(SSYS3, kde)	**1.12**	**1.35**	**1.48**	1.58	1.68	**1.08**	**1.61**	**2.09**	**2.56**	**3.03**
(SSYS1, par)	1.14	1.37	**1.47**	**1.51**	**1.53**	**1.14**	**1.80**	**2.47**	**3.20**	**4.02**
(SSYS1, kde)	1.13	1.40	1.59	1.79	2.01	**1.15**	**1.83**	**2.52**	**3.28**	**4.13**
DIFF	1.47	1.93	2.14	2.23	2.28	5.56	8.74	10.64	11.84	12.66
SYS	1.18	1.51	1.73	1.88	2.00	1.61	2.80	3.91	4.94	5.90
Pooled OLS	1.12	1.36	1.48	1.57	1.66	1.53	2.64	3.67	4.64	5.55

Table 9: RMSE in the Simulation Study Design 2: Non-Gaussian Correlated Random Effects. Study design is described in Table 3. Values lower than the lowest benchmark in each column are bolded. SSYS3 and SSYS1 stand for three- and one-step Sub-optimal System GMM estimator from Section 3.4.5, while DIFF, LEV and SYS stand for Difference-, Level- and System-GMM estimators. PMPP shows different variants of Posterior Mean Panel Predictor, where the first term in parenthesis denotes estimation method for common parameters, while *par* and *kde* stand for parametric and kernel-based estimation of the intercept parameters respectively.

	$T = 5$					$T = 15$				
Horizon:	+1	+2	+3	+4	+5	+1	+2	+3	+4	+5
PMPP:										
(QMLE, par)	1.91	3.07	4.11	**5.12**	**6.18**	1.68	2.45	3.01	3.49	3.92
(QMLE, kde)	8.71	16.84	24.92	33.29	42.22	4.14	7.79	11.37	14.93	18.50
(DIFF, par)	2.48	3.58	4.13	**4.41**	**4.55**	3.88	4.53	4.64	4.66	4.66
(DIFF, kde)	4.76	7.07	8.22	8.79	9.08	2.69	3.11	3.19	3.20	3.20
(SYS, par)	1.63	2.83	4.20	5.75	7.46	1.20	1.55	1.80	2.02	2.25
(SYS, kde)	1.88	3.50	5.35	7.42	9.70	1.24	1.71	2.15	2.61	3.09
(LEV, par)	**1.52**	**2.45**	**3.40**	**4.37**	**5.35**	1.20	1.55	1.80	2.03	2.26
(LEV, kde)	1.67	2.83	4.02	**5.24**	**6.46**	1.25	1.76	2.25	2.76	3.31
(SSYS3, par)	**1.50**	**2.43**	**3.40**	**4.42**	**5.49**	1.20	1.55	1.80	2.03	2.26
(SSYS3, kde)	1.82	3.22	4.70	6.25	7.88	1.24	1.71	2.16	2.62	3.11
(SSYS1, par)	**1.50**	**2.42**	**3.39**	**4.41**	**5.47**	1.20	1.55	1.81	2.04	2.27
(SSYS1, kde)	1.79	3.17	4.63	6.15	7.73	1.24	1.72	2.18	2.65	3.15
DIFF	6.82	10.18	11.85	12.69	13.11	11.19	13.12	13.46	13.52	13.53
SYS	1.55	2.64	3.88	5.27	6.83	1.15	1.41	1.53	1.58	1.60
Pooled OLS	1.56	2.66	3.91	5.33	6.91	1.17	1.45	1.59	1.68	1.75

Appendix B Detailed empirical application results

Table 10: RMSE of forecasts from a rolling window sample with window length T, for the original data series. Values lower than the lowest benchmark in each column are bolded. SSYS3 and SSYS1 stand for three- and one-step Sub-optimal System GMM estimator from Section 3.4.5, while DIFF, LEV and SYS stand for Difference-, Level- and System-GMM estimators. PMPP shows different variants of Posterior Mean Panel Predictor, where the first term in parenthesis denotes estimation method for common parameters, while *par* stands for parametric estimation of the intercept parameters.

	Original data					
Horizon:	+1	+2	+3	+4	+5	Average
	$T = 5$					
PMPP:						
(QMLE, par)	370 345	632 122	871 359	1 050 005	1 216 931	828 152
(DIFF, par)	636 686	635 089	953 041	1 179 065	1 912 221	1 063 220
(SYS, par)	353 375	592 199	798 683	918 889	1 007 879	734 205
(LEV, par)	365 069	633 093	877 415	1 044 787	1 199 400	823 953
(SSYS3, par)	361 564	621 078	858 656	1 010 964	1 140 572	798 567
(SSYS1, par)	362 131	627 783	872 404	1 037 478	1 182 704	816 500
DIFF	2 516 841	2 280 046	3 431 826	4 155 703	7 035 871	3 884 057
SYS	377 261	607 043	779 963	869 731	938 121	714 424
Pooled OLS	342 128	558 336	739 423	830 367	889 956	672 042
	$T = 15$					
PMPP:						
(QMLE, par)	318 456	**477 251**	**590 743**	**632 354**	**643 983**	**532 557**
(DIFF, par)	449 754	570 947	**621 189**	**637 953**	**641 048**	**584 178**
(SYS, par)	**305 927**	**484 556**	**629 572**	**683 068**	**705 329**	**561 690**
(LEV, par)	**310 754**	501 465	662 314	729 160	758 325	592 404
(SSYS3, par)	**308 484**	**494 742**	**649 853**	**710 762**	**738 194**	**580 407**
(SSYS1, par)	**309 475**	**496 392**	**650 938**	**710 834**	**736 866**	**580 901**
DIFF	1 756 784	2 216 666	2 365 915	2 421 947	2 446 078	2 241 478
SYS	311 617	498 577	654 565	718 655	748 485	586 380
Pooled OLS	312 352	498 735	652 939	717 772	751 146	586 589

Table 11: RMSE of forecasts from a rolling window sample with window length T, for the demeaned data series. Values lower than the lowest benchmark in each column are bolded. SSYS3 and SSYS1 stand for three- and one-step Sub-optimal System GMM estimator from Section 3.4.5, while DIFF, LEV and SYS stand for Difference-, Level- and System-GMM estimators. PMPP shows different variants of Posterior Mean Panel Predictor, where the first term in parenthesis denotes estimation method for common parameters, while *par* stands for parametric estimation of the intercept parameters.

	Demeaned data					
Horizon:	+1	+2	+3	+4	+5	Average
	$T = 5$					
PMPP:						
(QMLE, par)	326 998	536 442	730 351	858 303	969 535	684 326
(DIFF, par)	595 450	640 377	1 127 297	2 240 846	6 583 679	2 237 530
(SYS, par)	311 326	500 668	666 486	761 305	829 734	613 904
(LEV, par)	320 773	535 926	738 521	866 696	983 797	689 143
(SSYS3, par)	319 718	532 701	727 353	851 864	964 840	679 295
(SSYS1, par)	317 767	530 550	732 836	861 346	975 053	683 510
DIFF	2 007 172	2 062 685	3 876 520	8 312 258	25 805 939	8 412 915
SYS	342 053	543 556	696 047	784 334	853 840	643 966
Pooled OLS	308 427	492 285	649 560	722 365	769 339	588 395
	$T = 15$					
PMPP:						
(QMLE, par)	290 947	**436 462**	**537 576**	**570 115**	**581 240**	**483 268**
(DIFF, par)	427 347	537 284	583 169	**600 140**	**602 882**	550 164
(SYS, par)	**279 139**	**433 120**	**557 715**	**596 966**	**613 785**	**496 145**
(LEV, par)	**281 139**	**445 137**	583 817	637 201	664 339	522 327
(SSYS3, par)	**281 561**	**442 652**	**577 679**	**624 572**	**644 301**	**514 153**
(SSYS1, par)	**280 991**	**442 955**	**577 234**	**625 267**	**647 810**	**514 851**
DIFF	1 433 529	1 797 480	1 908 994	1 947 612	1 964 191	1 810 361
SYS	285 594	451 462	590 211	645 986	675 264	529 703
Pooled OLS	283 417	445 979	581 681	634 555	661 315	521 389

Appendix C R-package pmpp

The methods analysed in this research can be easily applied in the statistical software R with use of the `pmpp` package accompanying this paper. The package provides a set of functions allowing for Posterior Mean Panel Predictor modelling and forecasting, as well as for bootstrapping prediction intervals. This appendix gives a brief description of the main functions included in the package.

```
# Functions ------------------------------------------------------------------

pmpp(depVar, data, panelIndex  = colnames(data[,1:2]), expVar = NULL,
    csiVar = NULL, postMean = 'gaussian', commonPar = 'QMLE', optimMethod =
    'quadratic', densGrid = 2^10, GMMmodel = 'twosteps', GMMinst = 99,
    pureData = TRUE)
# depVar        character string indicating name of dependet variable
# panelIndex    vector of length 2 indicating names of variables indexing
#               units and time periods respectively
# expVar        vector of character strings indicating names of
#               exogeneuos explanatory variables
# csiVar        vector of character strings indicating names of
#               cross-sectionally invariant explanatory variables
# data          data.frame or matrix with input data
# postMean      method for estimating the heterogeneous intercept
#               parameters, one of "gaussian", "kernel"
# commonPar     method for estimating the common parameters, one of
#               "QMLE", "GMM.ABond", "GMM.BBond", GMM.ABover", "GMM.SSYS"
# optimMethod   which optimisation routine to use, one of "gradient",
#               "quadratic", "annealing"
# densGrid      size of the grid over which data is interpolated for
#               kernel density estimation; larger value may yield higher
#               accuracy, but increases computation time
# GMMmodel      number of steps for computing optimal GMM matrix,
#               one of "onestep", "twosteps", "threesteps";
#               "threesteps" can be used for "GMM.SSYS" only
# GMMinst       number of lagged values of the dependent variable to be
#               used as GMM instruments in Arellano-Bond/Blundell-Bond
#               setting
# pureData      if TRUE, removes indexing/subsetting from model's call
#               on data, faciliating use in a loop

pmpp.predict(model, n.ahead = 1, newdata = NULL, iter = NULL)
# model     PMPP model as returned by pmpp()
# newdata   data frame with future observations of explanatory variables
# n.ahead   numeric, forecast horizon
# iter      iterating constant, to be used in a loop

pmpp.predInterval(model, n.ahead = 1, bootReps = 1000, blockSize = NULL,
                  confidence = 0.95, iter = NULL)
# model        PMPP model, as returned by pmpp()
# n.ahead      integer; forecast horizon
# bootReps     integer; number of bootstrap replications
# blockSize    integer; width of the re-sampled block of residuals
# confidence   numeric in (0,1); confidence level of the interval
# iter         iterating constant, to be used in a loop
```

```
# Examples ------------------------------------------------------------------

# Prepare the data
data(EmplUK, package = 'plm')
EmplUK <- filter(EmplUK, year %in% c(1978,1979,1980,1981,1982))

# Estimate the parameters of PMPP model
model <- pmpp(depVar = 'emp', expVar = c('wage', 'capital'), data = EmplUK)

# Compute forecasts
prediction <- pmpp.predict(model, n.ahead = 5)

# Bootstrap predcition interval
intervals <- pmpp.predInterval(model, bootReps = 1000, n.ahead = 5)
```