Australian & New Zealand Journal of Statistics

Aust. N. Z. J. Stat. **62**(4), 2020, 536–557 doi: 10.1111/anzs.12316

Bayesian estimation and model comparison for linear dynamic panel models with missing values

Christian Aßmann[1,2*] and Marcel Preising[3]

Otto-Friedrich-Universität Bamberg, Leibniz Institute for Educational Trajectories Bamberg, Federal Statistical Office Germany

Summary

Panel data are collected over several time periods for the same units and hence allow for modelling both latent heterogeneity and dynamics. Since in a dynamic setup, the dependent variable also appears as an explanatory variable in later periods, missing values lead to substantial loss of information and the possibility of inefficient estimation. For linear dynamic panel models with fixed or random effects, we suggest a Bayesian approach to deal with missing values. The Gibbs sampling scheme providing a sample from the posterior distribution is thereby augmented by draws from the full conditional distribution of the missing values. While the full conditional distribution for missing values in the dependent variable is implied by the model setup, we incorporate a flexible non-parametric approximation to the full conditional posterior distribution of missing values in the explaining variables. Also, we provide accurate non-nested model comparison in terms of the marginal likelihood from the resulting hybrid Gibbs sampling output. The properties and possible efficiency gains of the suggested approach are illustrated by means of a simulation study and an empirical application using a macroeconomic panel data set.

Key words: data augmentation; dynamic linear panel regression; marginal likelihood; missing values.

1. Introduction

Linear dynamic panel regression models are workhorse tools in economics and the social sciences. Depending on the specification of latent heterogeneity, established estimation routines in form of generalised methods of moments, instrumental variable or maximum likelihood estimators are available (see, e.g. Arellano & Bond 1991; Arellano & Bover 1995;

*Author to whom correspondence should be addressed.
[1]Chair of Survey Statistics and Data Analysis, Otto-Friedrich-Universität Bamberg, 96047 Bamberg, Bayern Germany, e-mail: christian.assmann@uni-bamberg.de
[2]Leibniz Institute for Educational Trajectories Bamberg, 96047 Bamberg, Bayern Germany
[3]Federal Statistical Office Germany, 65189 Wiesbaden, Germany

Acknowledgements. We thank for the inspiring ideas received at the Joint Statistical Meetings 2017 of the American Statistical Association and participants of the statistics seminars of the University of Cologne and the Leibniz Institute for Educational Trajectories, where this work was presented. The first author acknowledges financial support by the Deutsche Forschungsgesellschaft (DFG) within priority programme SPP 1646 under grants AS 368/3-1 and 368/3-2. We thank the Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities for the provision of the system resources the simulation studies were performed with.

Blundell & Bond 1998; Everaert 2013; Hsiao 2014). Note that these estimation approaches are all based on variable transformations, for example within transformations or difference operations, which apply to linear model setups only. Popular software implementations are also available. For non-linear model frameworks, random effects models requiring simulation-based estimation approaches (see, e.g. Richard & Zhang 2007; Liesenfeld & Richard 2010; Mesters & Koopman 2014; Mesters, van der Geest & Bijleveld 2016) are typically used, as fixed effects specifications cause the occurrence of identification problems. Also Bayesian estimation of linear dynamic panel regression models incorporating latent heterogeneity at different degrees is discussed within the literature (Hsiao & Pesaran 2008; Liu *et al.* 2016).

However, these estimation approaches are typically discussed for completely observed, yet possibly unbalanced, data sets. That is, the set of variables per individual has to be observed in consecutive time periods, where the number of time periods may differ across the individuals. This requirement is often not met in empirical application. Instead, missing values may occur at any time period for any variable due to nonresponse in surveyed data. Also, even in macroeconomic data sets obtained from administrative data sources, typically not all considered variables are available for the same set of time periods. This might be, for example due to changes in national accounting standards. Hence, as not all macroeconomic aggregates are subject to subsequent revisions, missing values occur for some time periods in some variables. As a missing value in the dependent variable affects at least two time periods in the dynamic context, the typical strategy suggested by Arellano & Bond (1991) to restrict the data set to those individuals and periods where information is completely available, may cause a higher information loss compared to the static context. Indeed, strategies for the efficient handling of the information loss resulting from missing values (see for an introduction Little & Rubin 2002) are almost absent from the literature on dynamic panel models. An exception is given by Millimet & McDonough (2016), who correct for incomplete data in a linear dynamic framework including individual-specific effects. However, their approach deals with missing values within the explaining variables only.

This article suggests a Bayesian estimation approach to handle missing values in the context of linear lagged dependent variable models of order one with latent individual- and time-specific heterogeneity captured by fixed or random effects. The approach can handle any unbalanced panel data pattern. As we develop our estimation approach in a Bayesian framework, conjugate prior distributions are chosen for the model parameters and Markov Chain Monte Carlo (MCMC) methods (see e.g. Geweke 1989), namely Gibbs Sampling (e.g. Geman & Geman 1984; Gelfand & Smith 1990) are at hand to provide the approximations of the corresponding posterior densities. To correct for the non-response within the dynamic response variable, we derive the full conditional distribution of the missing values as implied by the model setup and augment the sampler accordingly (Tanner & Wong 1987; Li 1988). Our data augmentation approach secures the identifiability of the necessary full conditional distributions and profits from the properties of the normal distribution. To handle incomplete covariate data, we consider non-parametric approximations of the full conditional distributions of the missing values provided by classification and regression trees (CART) (Burgette & Reiter 2010). This is in line with non-parametric prior distributions for the missing values in the covariates.

Comparison of the different non-nested model specifications incorporating latent heterogeneity at different degrees is straightforward in terms of Bayes factors (Jeffreys 1961; Kass & Raftery 1995; Frühwirth-Schnatter 2004; Frühwirth-Schnatter & Kaufmann 2008) and hence a further conceptual advantage of the Bayesian estimation approach. Typical model selection devices such as information-based criteria are not readily available for missing data problems in the panel context, but require substantial numerical calculation efforts (see e.g. Ibrahim, Zhu & Tang 2008). Calculation of information criteria is based on the log-likelihood, which is often obtainable for special missing data problems by use of excessive numerical or analytical integration. Due to that, we extend the Bayesian approach of Chib (1995) towards dynamic linear regression models with missing values to compute marginal likelihoods.

The accuracy of our approach in terms of parameter estimation and model comparison is evaluated by means of a simulation study and an empirical application for a macroeconomic panel data set. The results show an increase of efficiency regarding the parameter estimation for our data augmentation scheme in comparison to complete case estimations both in the simulations and the empirical application. Additionally, the marginal likelihood computations allow for precise model decisions under various data generating processes within the simulation study. Thus, our non-nested evaluation approach provides reliable model selections, even for incomplete data situations. Alternatives to our suggested data augmentation approach are available within the literature (for multiple imputation by chained equations, see Van Buuren *et al.* 2015). In contrast, alternatives for non-nested model comparison based on posterior predictive assessment (see Gelman, Meng & Stern 1996) are typically based on ad hoc assumptions regarding the partitioning of the data utilised for the estimation and prediction. An extensive comparison is, however, beyond the scope of the paper.

The remainder of this paper is organised as follows. Section 2 provides the suggested linear dynamic panel model, the choice of conjugate prior distributions, and the resulting full conditional posterior densities in detail. In Section 3, we develop our data augmentation method and the calculation of the marginal likelihood. Section 4 evaluates the performance of the suggested approach by means of a simulation study and an empirical illustration. Section 5 concludes.

2. Model formulation and estimation

For a set of individuals enumerated by $i = 1, \ldots, N$ and a set of periods enumerated by $t = 1, \ldots, T$, $y_{i,t}$ denotes the continuous dependent variable. To start, consider a completely observed balanced panel structure. Thus, a dynamic linear model for the response variable $y_{i,t}$ is then given by

$$y_{i,t} = c_i + a_t + \xi + \boldsymbol{X}_{i,t}\boldsymbol{\beta} + \rho y_{i,t-1}\mathbf{1}(t>1) + \boldsymbol{X}_{i,1}\boldsymbol{\delta}\mathbf{1}(t=1) + \varepsilon_{i,t}. \tag{1}$$

Here $\mathbf{1}(\cdot)$ denotes the indicator function and $\varepsilon_{i,t}$ is a normally distributed error term with expectation zero and period-specific variance σ_t^2, whereas c_i and a_t represent either fixed or random individual- and time-specific effects. For the random effects specification, we assume all individual-specific effects to be normally and independently identically distributed with mean zero and variance σ_c^2, while all time-specific random effects are by assumption also normally and independently distributed with mean zero and variance σ_a^2. ξ denotes the global intercept, $\boldsymbol{X}_{i,t}$ is a $1 \times K$ vector of observable explanatory variables, and $\boldsymbol{\beta}$ is the corresponding $K \times 1$ parameter vector. ρ denotes a scalar to specify the dynamic relationship between $y_{i,t}$ and $y_{i,t-1}$. Parameters $\boldsymbol{\beta}$ and $\boldsymbol{\rho}$ are assumed to be homogeneous across all individuals and periods, whereas $\boldsymbol{\delta}$ captures deviations in the regression relation in the initial period. The unconditional model allows for handling missing values also in the initial period (see Section 3), where the suggested approach remains feasible in case of a conditional model setup.

To simplify calculation, denote $\boldsymbol{y}=[y_{1,1},y_{1,2},\ldots,y_{N,T}]^\top$ as the $NT\times 1$ vector of the response variable. Let the vectors $\boldsymbol{c}=[c_1,c_2,\ldots,c_N]^\top$, which is $N\times 1$, and the $T\times 1$ vector $\boldsymbol{a}=[a_1,a_2,\ldots,a_T]^\top$, and thereby $\boldsymbol{h}=[\boldsymbol{c}^\top,\boldsymbol{a}^\top]^\top$ summarise all individual- and time-specific effects in a $(N+T)\times 1$ column vector. Let $\boldsymbol{D_c}=\boldsymbol{I}_N\otimes\boldsymbol{\iota}_{T\times 1}$, where $\boldsymbol{I}_N$ is the identity matrix of dimension N, $\boldsymbol{\iota}_{T\times 1}$ is a column vector of ones of length T, and $\otimes$ denotes the Kronecker product. Let $\boldsymbol{D_a}=\mathbf{1}_{N\times 1}\otimes\boldsymbol{I}_T$, where $\mathbf{1}_{N\times 1}$ is a column vector of ones of length N and $\boldsymbol{I}_T$ is the identity matrix of dimension T. Summarise $\boldsymbol{D}=[\boldsymbol{D_c},\boldsymbol{D_a}]$ as the $NT\times(N+T)$ indicator matrix to control for the fixed or random effects intercepts. The global constant, as well as the terms $\boldsymbol{X}_{i,t}\boldsymbol{\beta}$, $\rho y_{i,t-1}$, and $\boldsymbol{X}_{i,1}\boldsymbol{\delta}$ are summarised as $\boldsymbol{Z}_{i,t}\boldsymbol{\gamma}$, where $\boldsymbol{Z}_{i,t}=[1,\boldsymbol{X}_{i,t},y_{i,t-1},0]$ for $t\neq 1$ and $\boldsymbol{Z}_{i,1}=[1,\boldsymbol{X}_{i,1},0,\boldsymbol{X}_{i,1}]$ is $1\times(2K+1)$, $\boldsymbol{\gamma}=[\xi,\boldsymbol{\beta}^\top,\rho,\delta^\top]^\top$ is $(2K+2)\times 1$, whereas $\boldsymbol{Z}=[\boldsymbol{Z}_{1,1}^\top,\boldsymbol{Z}_{1,2}^\top,\ldots,\boldsymbol{Z}_{N,T}^\top]^\top$ and $\boldsymbol{X}=[\boldsymbol{X}_{1,1}^\top,\boldsymbol{X}_{1,2}^\top,\ldots,\boldsymbol{X}_{N,T}^\top]^\top$ denote $NT\times(2K+2)$ or $NT\times K$ matrices of covariate information for all individuals and periods. With $\boldsymbol{\varepsilon}=[\varepsilon_{1,1},\varepsilon_{1,2},\ldots,\varepsilon_{N,T}]^\top$ and $\boldsymbol{\Sigma}=\mathrm{diag}[\sigma^2]\otimes\boldsymbol{I}_N$ with $\boldsymbol{\sigma}^2=[\sigma_1^2,\sigma_2^2,\ldots,\sigma_T^2]^\top$ denoting the $NT\times NT$ diagonal covariance matrix, the model stated in (1) in matrix notation is now given by

$$\boldsymbol{y}=\boldsymbol{Dh}+\boldsymbol{Z\gamma}+\boldsymbol{\varepsilon}\quad\text{and}\quad\boldsymbol{\varepsilon}\sim\mathrm{N}(\mathbf{0}_{NT\times 1},\boldsymbol{\Sigma}),\tag{2}$$

where for the random effects specification, we have $\boldsymbol{h}\sim N(\mathbf{0}_{(N+T)\times 1},\boldsymbol{\Sigma_h})$, with $\mathbf{0}_{(N+T)\times 1}$ expectation null vector of length $(N+T)$ and $\boldsymbol{\Sigma_h}=\mathrm{diag}[\sigma_c^2\boldsymbol{\iota}_{N\times 1},\sigma_a^2\boldsymbol{\iota}_{T\times 1}]$ as $(N+T)\times(N+T)$ covariance matrix of the random effects.

Collecting all structural model parameters into the vectors $\boldsymbol{\theta}_{\mathrm{FE}}$ and $\boldsymbol{\theta}_{\mathrm{RE}}$ for the fixed and random effects model, respectively, the corresponding joint density of all observations for the fixed effects model with therefore $2K+T+N+T$ structural parameters summarised in $\boldsymbol{\theta}_{\mathrm{FE}}=[\boldsymbol{\gamma},\sigma^2,\boldsymbol{h}^\top]^\top$ and the Nth individual- and the Tth time-specific effect each set to zero due to model identification, is given as

$$f(\boldsymbol{y}|\boldsymbol{\theta}_{\mathrm{FE}},\boldsymbol{X})=(2\pi)^{-\frac{NT}{2}}|\boldsymbol{\Sigma}|^{-0.5}\exp\left\{-\frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{Dh}-\boldsymbol{Z\gamma}\right)^\top\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}-\boldsymbol{Dh}-\boldsymbol{Z\gamma}\right)\right\}.\tag{3}$$

The joint density of all observations for the random effects specification is likewise given by

$$f(\boldsymbol{y}|\boldsymbol{\theta}_{\mathrm{RE}},\boldsymbol{X})=\int f(\boldsymbol{y},\boldsymbol{h}|\boldsymbol{\theta}_{\mathrm{RE}},\boldsymbol{X})\,\mathrm{d}\boldsymbol{h}=\int f(\boldsymbol{y}|\boldsymbol{\theta}_{\mathrm{RE}},\boldsymbol{X},\boldsymbol{h})f(\boldsymbol{h}|\boldsymbol{\theta}_{\mathrm{RE}})\,\mathrm{d}\boldsymbol{h},\tag{4}$$

where $f(\boldsymbol{y}|\boldsymbol{\theta}_{\mathrm{RE}},\boldsymbol{X},\boldsymbol{h})$ has the same functional form as the density provided in (3), the $2K+4+T$ structural parameters $\boldsymbol{\theta}_{\mathrm{RE}}=[\boldsymbol{\gamma}^\top,\sigma^2,\sigma_c^2,\sigma_a^2]^\top$, and $f(\boldsymbol{h}|\boldsymbol{\theta}_{\mathrm{RE}})$ is given as a multivariate normal distribution with mean zero and covariance $\boldsymbol{\Sigma}_h$.

From a Bayesian perspective, the main interest lies in the posterior distributions of the structural model parameters. We assume mutually independent conjugate prior distributions, that is multivariate normal and inverse gamma. A Gibbs sampler (e.g. Geman & Geman 1984; Gelfand & Smith 1990; Casella & George 1992) generates a random sample of the joint posterior distribution. Thus, we draw M values iteratively, for $m=1,\ldots,M$, from the respective full conditional distributions of the considered parameter blocks. The corresponding full conditional posterior distributions are presented in the supplementary material.

3. Missing values and model comparison

Next, the suggested data augmentation approach for estimation and handling of missing values is discussed. We assume the missing data mechanism to be ignorable (see, e.g. Little & Rubin 2002). The device of data augmentation is especially suited to deal with missing data problems, as augmenting the missing values into the parameter vector allows for handling all other model quantities as in the case of completely observed data. In addition, data augmentation within the Bayesian framework constitutes an estimation routine where no combining rules are required to provide valid estimators and corresponding variances. Furthermore, we describe the approximation of the marginal likelihood allowing for model comparison.

3.1. Missing values in the dynamic panel context

To deal with missing values, the Gibbs sampler is augmented to include draws from the corresponding full conditional distributions, whose functional forms are in principle the same for the fixed as for the random effects specification. Thus, the discussion will be based on the fixed effects specification. To handle missing values in each iteration $m = 1, \ldots, M$, let $\boldsymbol{y}^{(m)} = [\boldsymbol{y}^{\text{obs}}, \boldsymbol{y}^{\text{mis,(m)}}]$ denote the dependent variable with missing values completed by draws from their corresponding full conditional distributions, with superscripts $^{\text{mis}}$ and $^{\text{obs}}$ indicating missing and observed values respectively. $\boldsymbol{X}^{(m)} = [\boldsymbol{X}^{\text{obs}}, \boldsymbol{X}^{\text{mis,(m)}}]$ denotes the corresponding set of completed regressors. Given an initialisation $\boldsymbol{\theta}_{\text{FE}}^{(0)}, \boldsymbol{y}^{(0)}$, and $\boldsymbol{X}^{(0)}$, we obtain the following iterative procedure.

Step *I* Sample $\boldsymbol{\theta}_{\text{FE}}^{(m)}$ from $f(\boldsymbol{\theta}_{\text{FE}} | \boldsymbol{y}^{(m-1)}, \boldsymbol{X}^{(m-1)})$.
Step *II* Sample $\boldsymbol{y}^{\text{mis,(m)}}$ from $f(\boldsymbol{y}^{\text{mis}} | \boldsymbol{y}^{\text{obs}}, \boldsymbol{\theta}_{\text{FE}}^{(m)}, \boldsymbol{X}^{(m-1)})$.
Step *III* Sample $\boldsymbol{X}^{\text{mis,(m)}}$ from $f(\boldsymbol{X}^{\text{mis}} | \boldsymbol{y}^{(m)}, \boldsymbol{\theta}_{\text{FE}}^{(m)}, \boldsymbol{X}^{\text{obs}})$.

In each of the three steps, sampling from the indicated distribution is performed in terms of sampling from a corresponding set of full conditional distributions. The sampling performed within step *I* is facilitated by use of the set of full conditional distributions outlined in the supplementary material (section A). The full conditional distributions and corresponding draws involved in step *II* are obtained as follows. Formulate the joint density given in (3) as the conditional density of the second until the last period given the initial observations for the fixed effects model, that is,

$$f(\boldsymbol{y} | \boldsymbol{\theta}_{\text{FE}}, \boldsymbol{X}) = \prod_{i=1}^{N} f(\boldsymbol{y}_{i,2:T} | \boldsymbol{\theta}_{\text{FE}}, \boldsymbol{X}_{i,2:T}, y_{i,1}) f(y_{i,1} | \boldsymbol{\theta}_{\text{FE}}, \boldsymbol{X}_{i,1}). \tag{5}$$

The formulation allows to derive the full conditional distribution of the observations for an individual i in the initial period as a normal distribution with expectation $\mu_{y_{i,1}}$ and variance $\sigma^2_{y_{i,1}}$, where

$$\sigma^2_{y_{i,1}} = \frac{1}{\frac{1}{\sigma_1^2} + \frac{\rho^2}{\sigma_2^2}} \text{ and } \mu_{y_{i,1}} = \sigma^2_{y_{i,1}} \left(\frac{c_i + a_1 + \boldsymbol{Z}_{i,1}\boldsymbol{\gamma}}{\sigma_1^2} + \frac{\rho(y_{i,2} - c_i - a_2 - \xi - \boldsymbol{X}_{i,2}\boldsymbol{\beta})}{\sigma_2^2} \right).$$

To sample missing values in periods 2 to T, we obtain from (5) that the $(T-1)$ observations of a particular individual are, conditioned on the initial period, normally distributed with expectation vector $\boldsymbol{\mu}_{\boldsymbol{y}_{i,2:T}} = [\mu_{y_{i,2}}, \ldots, \mu_{y_{i,T}}]^\top$ of length $T-1 \times 1$, where

$$\mu_{y_{i,t}} = \rho^{t-1} y_{i,1} + \sum_{j=0}^{t-2} \rho^j (c_i + a_{t-j} + \xi + \boldsymbol{X}_{i,t-j}\boldsymbol{\beta}), \quad \text{for } t = 2, \ldots, T,$$

and corresponding covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{y}_{i,2:T}}$ of dimension $(T-1) \times (T-1)$ denoted as

$$\boldsymbol{\Sigma}_{\boldsymbol{y}_{i,2:T}} = \begin{pmatrix} \sigma^2_{y_{i,2}} & \sigma_{y_{i,2},y_{i,3}} & \cdots & \sigma_{y_{i,2},y_{i,T}} \\ & \sigma^2_{y_{i,3}} & & \sigma_{y_{i,3},y_{i,T}} \\ & & \ddots & \vdots \\ & & & \sigma^2_{y_{i,T}} \end{pmatrix},$$

with variances and covariances given as

$$\sigma^2_{y_{i,t}} = \sum_{j=0}^{t-2} \rho^{2j} \sigma^2_{t-j}, \quad \text{and } \sigma_{y_{i,t},y_{i,t+s}} = \rho^s \sigma^2_{y_{i,t}}, \text{ for } s = 0, \ldots, T-t \text{ and } 2 \leqslant t \leqslant T.$$

Given that the individuals are, conditioning on the fixed or random effects, mutually independent, draws of the missing values for each individual can be obtained by utilising the multivariate normal distribution properties. The conditional distribution of the missing values in $\boldsymbol{y}_{i,2:T}$, that is $\boldsymbol{y}^{\text{mis}}_{i,2:T}$ conditional on $\boldsymbol{y}^{\text{obs}}_{i,2:T}$, is hence given as a multivariate normal distribution with expectation and covariance $\mu^{\text{mis}}_{\boldsymbol{y}_{i,2:T}} + \boldsymbol{\Sigma}_{\boldsymbol{y}^{\text{mis}}_{i,2:T},\boldsymbol{y}^{\text{obs}}_{i,2:T}} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{y}^{\text{obs}}_{i,2:T}} (\boldsymbol{y}^{\text{obs}}_{i,2:T} - \mu^{\text{obs}}_{\boldsymbol{y}_{i,2:T}})$ and $\boldsymbol{\Sigma}_{\boldsymbol{y}^{\text{mis}}_{i,2:T}} - \boldsymbol{\Sigma}_{\boldsymbol{y}^{\text{mis}}_{i,2:T},\boldsymbol{y}^{\text{obs}}_{i,2:T}} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{y}^{\text{obs}}_{i,2:T}} \boldsymbol{\Sigma}_{\boldsymbol{y}^{\text{obs}}_{i,2:T},\boldsymbol{y}^{\text{mis}}_{i,2:T}}$. The superscripts mis and obs indicate those cells of $\mu_{\boldsymbol{y}_{i,2:T}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{y}_{i,2:T}}$ corresponding to the missing values and observed values respectively. Furthermore, $\boldsymbol{\Sigma}_{\boldsymbol{y}^{\text{mis}}_{i,2:T},\boldsymbol{y}^{\text{obs}}_{i,2:T}}$ denotes the covariance rows and columns of the missing and observed values of $\boldsymbol{\Sigma}_{\boldsymbol{y}_{i,2:T}}$. Note that this set of full conditional distributions is implied by the unconditional model setup given in (1), where an explicit model for the initial period is formulated. An unconditional model approach requires additional parameters to be specified for the initial period. However, at the same time, the formulation enables an increased estimation accuracy. This trade-off depends on the overall missing pattern of all variables. Thus, for some missing patterns, estimation conditional on the initial period may be favourable. However, the handling of missing values in $\boldsymbol{y}$ in period 2 to T in the context of a conditional modelling approach is the same as discussed above and may still enhance the estimation accuracy in combination with handling of missing values in $\boldsymbol{X}$.

With regard to step *III*, referring to data situations with missing values also occurring in the remaining covariates, that is $\boldsymbol{X} = [\boldsymbol{X}^{\text{mis}}, \boldsymbol{X}^{\text{obs}}]$, we suggest to implement nonparametric approximations for the full conditional distributions of missing values that is $f(\boldsymbol{X}^{\text{mis}} | \boldsymbol{y}, \boldsymbol{\theta}_{\text{FE}}, \boldsymbol{X}^{\text{obs}})$. These can be provided in form of classification and sequential regression trees (CART), see Burgette & Reiter (2010). Thus, consider the set of full conditional distributions $f(\boldsymbol{X}^{\text{mis}}_k | \boldsymbol{y} - \boldsymbol{D}\boldsymbol{h} - \rho \boldsymbol{y}_{-1}, \boldsymbol{X}_{\setminus k})$ for $k = 1, \ldots, K$, where $\boldsymbol{X}_{\setminus k}$ denotes the set of iteratively completed covariate variables except variable k, and $\boldsymbol{y}_{-1}$ denotes the vector of the lagged dependent variable. It is a major advantage of the data augmentation approach that the complete latent structure enters the approximation of the full conditional distribution of missing values most efficiently. Note that an adaption towards time or individual specific full conditional distributions of the missing variables is possible by adjusting the conditioning quantities by time- or individual-specific subsets. Sampling from the implied full conditional distributions is then performed using the Bayesian bootstrap (Rubin 1981). We favour this approach as it has shown to be highly capable of dealing with discrete and possible non-linear relationships among the variables (Doove, van Buuren & Dusseldorp 2014). In terms of assumed prior distributions, the considered non-parametric approximations are compatible with non-parametric prior distributions implied by the observed values. Further details are provided in the supplementary material.

3.2. Model evaluation via the marginal likelihood

As the comparison of non-nested model specifications is often of analytical interest, Bayes factors (see, e.g. for a general introduction Kass & Raftery 1995) are the choice at hand. The implied marginal likelihoods incorporate two competitive properties of models, parsimony and fit to the data. Note that the well-known and popular Bayesian information criterion (BIC) according to Schwarz (1978) is consistent only under fulfilment of certain regularity conditions (Kass & Raftery 1995). For an approximation of the marginal likelihood by the BIC, see Ando (2010), we require that the prior information of the model parameters can be ignored due to a sufficiently large number of observations compared to the amount of model parameters (Raftery 1995). However, this condition is probably unfulfilled in case of the fixed effects specification, as a growing number of observations directly yield a growing number of prior distributions due to the amount of parameters. Throughout the paper, we will assume two alternative model specifications a priori to be equally likely. The calculation of the Bayes factor hence corresponds to the computation of the marginal likelihoods, that is the normalising constant of the posterior density given as

$$f(\boldsymbol{y}|\boldsymbol{X}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{X}) f(\boldsymbol{\theta}|\boldsymbol{X})}{f(\boldsymbol{\theta}|\boldsymbol{y},\boldsymbol{X})}. \tag{6}$$

Chib (1995) and Chib & Jeliazkov (2001) suggest to approximate the natural logarithm of the marginal likelihood via the decomposition $\ln\widehat{f}(\boldsymbol{y}|\boldsymbol{X}) = \ln f(\boldsymbol{y}|\tilde{\boldsymbol{\theta}},\boldsymbol{X}) + \ln f(\tilde{\boldsymbol{\theta}}|\boldsymbol{X}) - \ln\widehat{f}(\tilde{\boldsymbol{\theta}}|\boldsymbol{y},\boldsymbol{X})$, where $\tilde{\boldsymbol{\theta}}$ is chosen as a point within the highest density region, typically the posterior estimate, to enhance the numerical accuracy of the approximation. Furthermore, $\ln f(\boldsymbol{y}|\tilde{\boldsymbol{\theta}},\boldsymbol{X})$ denotes the logarithm of the likelihood, and $\ln\widehat{f}(\tilde{\boldsymbol{\theta}}|\boldsymbol{y},\boldsymbol{X})$ is the posterior distribution including its normalising constant which can be approximated via recursively shortened Gibbs sequences.

To calculate the marginal likelihood in case of missing values in the covariates, we treat them in the sense of latent variables as discussed by Chib (1995). By denoting $\tilde{\boldsymbol{X}} = [\boldsymbol{X}^{\text{obs}}, \tilde{\boldsymbol{X}}^{\text{mis}}]$, this alters the approximation in case of the fixed effects model into

$$\begin{aligned} \ln\widehat{f}(\boldsymbol{y}^{\text{obs}}|\boldsymbol{X}^{\text{obs}}) = \ln f(\boldsymbol{y}^{\text{obs}}|\tilde{\boldsymbol{\theta}}_{\text{FE}},\tilde{\boldsymbol{X}}) &+ \ln f(\tilde{\boldsymbol{\theta}}_{\text{FE}},\tilde{\boldsymbol{X}}^{\text{mis}}|\boldsymbol{X}^{\text{obs}}) \\ &- \ln\widehat{f}(\tilde{\boldsymbol{\theta}}_{\text{FE}},\tilde{\boldsymbol{X}}^{\text{mis}}|\boldsymbol{y}^{\text{obs}},\boldsymbol{X}^{\text{obs}}), \end{aligned} \tag{7}$$

and for the random effects model into

$$\begin{aligned} \ln\widehat{f}(\boldsymbol{y}^{\text{obs}}|\boldsymbol{X}^{\text{obs}}) = \ln f(\boldsymbol{y}^{\text{obs}}|\tilde{\boldsymbol{\theta}}_{\text{RE}},\tilde{\boldsymbol{X}},\tilde{\boldsymbol{h}}) &+ \ln f(\tilde{\boldsymbol{\theta}}_{\text{RE}},\tilde{\boldsymbol{X}}^{\text{mis}},\tilde{\boldsymbol{h}}|\boldsymbol{X}^{\text{obs}}) \\ &- \ln\widehat{f}(\tilde{\boldsymbol{\theta}}_{\text{RE}},\tilde{\boldsymbol{X}}^{\text{mis}},\tilde{\boldsymbol{h}}|\boldsymbol{y}^{\text{obs}},\boldsymbol{X}^{\text{obs}}), \end{aligned} \tag{8}$$

hence indicating that the random effects can also be conceptualised as latent variables. Note that in case of continuous covariates, $\tilde{\boldsymbol{X}}^{\text{mis}}$ denotes the posterior mean of the M drawn missing values. For ordered variables, we suggest to compute medians. For categorical variables, calculations of modes from the MCMC output are appropriate. As the functional form of the fixed effects log-likelihood from (7) is the same as that of the augmented log-likelihood function of the random effects model given in (8), further illustrations concerning the log-likelihood evaluations will be based on the fixed effects specification. When missing values occur in the dependent variable $\boldsymbol{y}$, we in fact have to calculate

$$f(\boldsymbol{y}^{\text{obs}}|\tilde{\boldsymbol{\theta}}_{\text{FE}},\tilde{\boldsymbol{X}}) = \int f(\boldsymbol{y}^{\text{obs}},\boldsymbol{y}^{\text{mis}}|\tilde{\boldsymbol{\theta}}_{\text{FE}},\tilde{\boldsymbol{X}})d\boldsymbol{y}^{\text{mis}} = \int f(\boldsymbol{y}^{\text{obs}},\boldsymbol{y}^{\text{mis}}_{2:T},\boldsymbol{y}^{\text{mis}}_{1}|\tilde{\boldsymbol{\theta}}_{\text{FE}},\tilde{\boldsymbol{X}})\,\text{d}\boldsymbol{y}^{\text{mis}}_{1}\text{d}\boldsymbol{y}^{\text{mis}}_{2:T}.$$

For this purpose, we first integrate the missing values of the initial period $\boldsymbol{y}^{\text{mis}}_{1}$. The resulting integrand takes the form of a normal distribution. The corresponding expectations do not depend on missing values in the lagged dependent variable, and thus, allow for an analytical solution due to the properties of the multivariate normal distribution. This is in contrast to a formulation of the joint density as implied by (3), where the dependent variables of all periods, and thus, potentially missing values are components of the expectations of subsequent periods. In order to check the numerical accuracy of the approach of Chib (1995) in this context, we randomly pick out one of the before deletion data sets generated within the simulation study below and calculate the log marginal likelihood both by an analytical and a 10 times repetition of the numerical integration for the random effects model, with the latter procedure requiring the evaluation of the prior and posterior as proposed by Chib (1995). The results show a mean difference not until the third decimal place.

For the involved prior evaluations, we calculate the prior ordinates at $\tilde{\boldsymbol{\theta}}_{\text{FE}}$, $\tilde{\boldsymbol{\theta}}_{\text{RE}}$, and $\tilde{\boldsymbol{X}}^{\text{mis}}$ respectively, given the corresponding prior distributions. As the prior of the random effects formulation decomposes into

$$\ln f(\tilde{\boldsymbol{\theta}}_{\text{RE}},\tilde{\boldsymbol{X}}^{\text{mis}},\tilde{\boldsymbol{h}}|\boldsymbol{X}^{\text{obs}}) = \ln f(\tilde{\boldsymbol{\theta}}_{\text{RE}}) + \ln f(\tilde{\boldsymbol{X}}^{\text{mis}}|\boldsymbol{X}^{\text{obs}}) + \ln f(\tilde{\boldsymbol{h}}|\tilde{\boldsymbol{\theta}}_{\text{RE}}),$$

the distribution of the latent individual- and time-specific intercepts. which is given as a multivariate normal distribution with mean zero and variance $\tilde{\boldsymbol{\Sigma}}_h$ as implied by $\tilde{\sigma}^2_c$ and $\tilde{\sigma}^2_a$, is evaluated at $\tilde{h}$. To evaluate the posterior parts of (7) and (8), we perform shortened Gibbs samplers for each structural parameter block as described in Chib (1995). In case of random effects, we apply additional shortened Gibbs sequences for both latent quantity blocks and evaluate the full conditional densities given the generated moments from the two reduced samplers at $\tilde{c}$ and $\tilde{a}$, respectively. Finally, we perform runs of shortened MCMC sequences for the missing covariates given all other structural and, in case of random effects, latent quantities. To assess the corresponding posterior density ordinates, we calculate the ordinates of the Gaussian kernel densities in case of continuous and the empirical frequency distributions in case of discrete variables within each iteration of CART. The average over the iterated ordinates implies the estimated missing values posterior density ordinates.

4. Evaluation

The evaluation focuses on highlighting the estimation accuracy gains of our data augmentation approach. The estimation routines have been implemented in R (R Core Team 2017). All computer code necessary for the simulation study and empirical application are available from the authors upon request. Using simulated data generated as outlined below, we first assess in Section 4.1 the relative performance of the suggested approach when missing values occur at different degrees. Within a benchmark scenario, completely observed data is considered (Scenario I). Within a second scenario, we erase entries of the dependent variable (Scenario II), and within a third scenario, observations of one of the covariates are additionally deleted (Scenario *III*). All scenarios are subject to complete cases analyses and estimations via data augmentation. The performance is assessed via coverage rates, root mean squared errors (RMSE), and biases. In a next step, we compare the suggested approach to a classical estimation routine available in the literature (Section 4.2). Finally, Section 4.3 provides an illustrative empirical comparison between a classical estimation routine and our new data augmentation approach.

We use the following data generating process employed within the three scenarios. We generate 500 data sets for both the fixed and random effects specifications given the model formulation of (1). We set $N=50$ but differentiate between a sample with $T=5$ and a sample with $T=20$. The covariates $\boldsymbol{X}_{i,t}=[X_{1,i,t},X_{2,i,t},X_{3,i,t}]$ are drawn from a trivariate normal distribution with expectation vector $\mu_{\boldsymbol{X}_{i,t}}=(0,0,0)$ and correlations $r_{X_{1,i,t},X_{2,i,t}}=0.35$, $r_{X_{1,i,t},X_{3,i,t}}=r_{X_{2,i,t},X_{3,i,t}}=0.5$, where, next to the first two variables, the categorical variable implied by $\mathbf{1}(X_{3,i,t}<-1)+2\mathbf{1}(-1\leqslant X_{3,i,t}\leqslant 1)+3\mathbf{1}(X_{3,i,t}>1)$ is used as a regressor. The corresponding regression coefficients are set to $[\xi,\boldsymbol{\beta}^{\top},\delta^{\top}]=[-1.630,4.900,4.528,1.437,0,0,0]$. We consider two settings concerning the dynamic dependence via assuming $\rho=0.2$ and $\rho=0.8$. Error terms are normally distributed with mean zero and period-specific variances drawn from a continuous uniform distribution with bounds $[0.5;4]$. Fixed effects are generated from a probability density function given as

$$f(c,a)=\prod_{i=1}^{N-1}\frac{|c_i|}{100}\mathbf{1}_{[-10,10]}(c_i)\prod_{t=1}^{T-1}\frac{|a_t|}{100}\mathbf{1}_{[-10,10]}(a_t).$$

Both the N-th individual- and T-th time-specific fixed effect are set to zero to solve the identification issue. Random effects are drawn from a normal distribution with expectation zero and variances $\sigma_c^2=\sigma_a^2=0.75$. For Scenario *II*, we consider the following missing at random mechanism related to $X_{2,i,t}$,

$$\Pr(y_{it}\text{ is missing})=\frac{1}{1+\exp\{2-2.5X_{2,i,t}\}},\qquad i=1,\ldots,N,\quad t=1,\ldots,T.$$

This mechanism implies that around 25% of the entries of the dependent and thus lagged dependent variable are missing. For Scenario *III*, we additionally consider a missing mechanism for $X_{3,i,t}$ conditional on $X_{1,i,t}$. Thus, $X_{3,i,t}$ is missing if $F_U(U_{i,t})>0.9$, where $F_U(U_{i,t})$ denotes the empirical distribution function of the random variable

$$U_{i,t}=\frac{1}{1+\exp\{0.2\phi_{i,t}X_{1,i,t}\}+\tau_{i,t}},\qquad i=1,\ldots,N,\quad t=1,\ldots,T,$$

with $\phi_{i,t}$ and $\tau_{i,t}$ being both standard normally distributed. Thus, exactly 10% of the variable is set to be missing.

4.1. Simulation results

This subsection assesses the relative performance of the suggested approach, when missing values are considered at different degrees. Details on the employed prior and MCMC setup are provided in the supplementary material (section C). We examine each data set by fixed and random effects analysis models both with our data augmentation scheme and complete case analyses for Scenarios *II* and *III*. Note that estimation is performed under the restriction $\delta=0$ as the relative performance of the suggested approach is of interest.

Discussion of simulation results focus on the setting with $\rho = 0.2$. The corresponding results are provided in Tables 1–3. Results for $\rho = 0.8$ are provided in the supplementary material (section D in Tables 6–8) and are in line with results for $\rho = 0.2$. Differences are highlighted after discussing the results for $\rho = 0.2$.

Table 1 provides the coverage rates of the mean parameters that is the rate of cases where the true value of a parameter is located within the 95% highest posterior density region interval, and the ratios of how often the fixed effects and random effects specification is chosen as the favoured analysis model respectively. To provide a benchmark, we calculate the BIC according to and the log marginal likelihood to compare the specifications for Scenario *I* and check both for congruence in terms of model decision. Therefore, we integrate the random effects of the joint density given in (4) analytically. As the approximation of the log marginal likelihood described in Section 3.2 allows to evaluate models in the event of missing values, we use our approach to compare the fixed and random effects specifications for Scenarios *II* and *III*. While the BIC could be obtained for the fixed effects model with missing values only in the dependent variable, the log-likelihood of the random effects specification required for model comparison involves excessive integration. To check for efficiency gains of our data augmentation approach compared to the complete cases estimators, Table 2 provides the respective RMSEs of the parameters. Additionally, Table 3 provides the respective biases of the parameter estimators.

If we consider the fixed effects model as the true data generating process (DGP), the results for the fixed effects analyses model including data augmentation show coverages at the nominal 95%-level for each parameter for both the small and large sample sizes (left half of Table 1). Note that the coverage rate at nominal 95%-level based on 500 replications approximately ranges in the interval $0.95 \pm 1.96\sqrt{0.95 \times 0.05/500}$. The parameter estimate corresponding to the covariate with missing values, that is $\tilde{\boldsymbol{\beta}}_3$, shows in Scenario *III* slightly lower but still satisfying rates (0.924 for the small sample; 0.912 for the large sample). Table 2 demonstrates the advantages of our data augmentation approach in terms of RMSEs. For example, the RMSEs of the small sample analyses of Scenario *III* decrease between around 20% (for $\tilde{\beta}_1$ with data augmentation 0.583; complete case 0.721) and more than 40% (for $\tilde{\rho}$ with data augmentation 0.017; complete case 0.029). The large sample setup supports these findings as we observe increased efficiencies of up to more than 30% (for $\tilde{\rho}$ with data augmentation: 0.006; complete case 0.009) for Scenario *III*. Considering a random effects specification as the DGP, the fixed effects analysis model results in lower coverage rates for the intercept and the dynamic component, while for the remaining parameters we again observe rates of around the nominal 95% (left half of Table 1). The corresponding complete case estimates reveal similar coverage rates but again tend to higher RMSEs except for the intercept (Table 2). Applying random effects analysis models with data augmentation (right half of Tables 1 and 2) yields, apart from the intercepts if a small sample size and a fixed effects DGP setup is chosen, coverage rates near the nominal 95%-level. Table 8 provides corresponding biases, where results indicate overall modest biases. The data augmentation procedure tends to achieve remarkably higher efficiencies compared to the complete case estimates, independent of the data generating process and the sample size.

The model decisions for Scenario *I* that is before deletion, in terms of both the BIC and the log marginal likelihood point in the same direction. Assuming a fixed effects model as the DGP, both key figures favour the fixed effects analysis model (Table 1) in 100%, as the random effects analysis specification is unable to cope with the high extent of not normally distributed latent heterogeneity. To check both model selection criteria for sensitivity in terms

Table 1. Simulated data with fixed and random effects as true data generating process analysed with data augmentation and complete case approach. 95% coverage rates and preference rates for Scenarios *I*, *II* and *III* with $\rho = 0.2$.

		Fixed effects analysis						Random effects analysis					
		Small sample			Large sample			Small sample			Large sample		
		I	*II*	*III*	*I*	*II*	*III*	*I*	*II*	*III*	*I*	*II*	*III*
Data generating process according to fixed effects specification													
Data augmentation	$\tilde{\xi}$	0.970	0.966	0.968	0.942	0.948	0.948	0.858	0.848	0.848	0.986	0.974	0.986
	$\tilde{\beta}_1$	0.958	0.946	0.962	0.940	0.934	0.948	0.958	0.946	0.962	0.938	0.932	0.946
	$\tilde{\beta}_2$	0.938	0.940	0.946	0.930	0.954	0.956	0.932	0.934	0.950	0.934	0.954	0.960
	$\tilde{\beta}_3$	0.940	0.940	0.924	0.938	0.932	0.912	0.948	0.940	0.922	0.940	0.934	0.906
	$\tilde{\rho}$	0.944	0.938	0.944	0.958	0.956	0.964	0.944	0.932	0.940	0.958	0.956	0.958
Complete case	$\tilde{\xi}$	–	0.958	0.958	–	0.954	0.950	–	0.932	0.916	–	1.000	0.998
	$\tilde{\beta}_1$	–	0.952	0.950	–	0.948	0.952	–	0.954	0.952	–	0.952	0.952
	$\tilde{\beta}_2$	–	0.940	0.932	–	0.952	0.954	–	0.936	0.946	–	0.952	0.952
	$\tilde{\beta}_3$	–	0.936	0.942	–	0.934	0.948	–	0.936	0.944	–	0.934	0.946
	$\tilde{\rho}$	–	0.924	0.904	–	0.954	0.940	–	0.916	0.896	–	0.952	0.932
Preference rate													
Marginal likelihood		1.000	0.998	0.988	1.000	1.000	0.996	0.000	0.002	0.012	0.000	0.000	0.004
BIC		1.000	–	–	1.000	–	–	0.000	–	–	0.000	–	–
Data generating process according to random effects specification													
Data augmentation	$\tilde{\xi}$	0.548	0.616	0.628	0.386	0.452	0.438	0.974	0.978	0.972	0.956	0.946	0.940
	$\tilde{\beta}_1$	0.944	0.954	0.968	0.964	0.960	0.958	0.950	0.952	0.954	0.960	0.958	0.956
	$\tilde{\beta}_2$	0.956	0.952	0.956	0.950	0.956	0.942	0.954	0.954	0.956	0.952	0.950	0.934
	$\tilde{\beta}_3$	0.940	0.948	0.904	0.970	0.960	0.902	0.950	0.956	0.910	0.962	0.960	0.908
	$\tilde{\rho}$	0.896	0.874	0.870	0.948	0.926	0.932	0.934	0.940	0.942	0.952	0.940	0.940

(Continued)

Table 1. *(Continued)*

		Fixed effects analysis						Random effects analysis					
		Small sample			Large sample			Small sample			Large sample		
		I	*II*	*III*	*I*	*II*	*III*	*I*	*II*	*III*	*I*	*II*	*III*
Complete case	$\tilde{\xi}$	–	0.726	0.762	–	0.548	0.564	–	0.974	0.976	–	0.950	0.950
	$\tilde{\beta}_1$	–	0.946	0.940	–	0.920	0.934	–	0.948	0.960	–	0.936	0.946
	$\tilde{\beta}_2$	–	0.956	0.960	–	0.950	0.954	–	0.966	0.966	–	0.942	0.952
	$\tilde{\beta}_3$	–	0.950	0.954	–	0.954	0.942	–	0.950	0.952	–	0.956	0.940
	$\tilde{\rho}$	–	0.892	0.916	–	0.914	0.922	–	0.954	0.966	–	0.936	0.932
Preference rate													
Marginal likelihood		0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000
BIC		0.000	–	–	0.074	–	–	1.000	–	–	0.926	–	–

Notes: The preference rates indicate how often the specifications are chosen as the favourable analysis model in terms of the Bayesian Information Criterion (BIC) and the logarithm of the marginal likelihood, based on analyses including data augmentation for Scenarios *II* and *III*.

Table 2. Simulated data with fixed and random effects as true data generating process analysed with data augmentation and complete case approach. Root mean squared errors of mean estimates for Scenarios *I*, *II* and *III* with $\rho = 0.2$.

		Fixed effects analysis						Random effects analysis					
		Small sample			Large sample			Small sample			Large sample		
		I	*II*	*III*	*I*	*II*	*III*	*I*	*II*	*III*	*I*	*II*	*III*
Data generating process according to fixed effects specification													
Data augmentation	$\tilde{\xi}$	0.523	0.589	0.583	0.344	0.382	0.394	2.837	2.773	2.847	0.930	0.920	0.903
	$\tilde{\beta}_1$	0.126	0.154	0.153	0.054	0.064	0.064	0.126	0.154	0.154	0.054	0.064	0.064
	$\tilde{\beta}_2$	0.132	0.184	0.187	0.054	0.072	0.073	0.132	0.185	0.187	0.054	0.072	0.073
	$\tilde{\beta}_3$	0.240	0.296	0.314	0.098	0.116	0.135	0.240	0.298	0.315	0.098	0.116	0.135
	$\tilde{\rho}$	0.013	0.017	0.017	0.005	0.006	0.006	0.014	0.017	0.017	0.005	0.006	0.006
Complete case	$\tilde{\xi}$	–	0.694	0.721	–	0.452	0.485	–	2.799	2.894	–	0.756	0.704
	$\tilde{\beta}_1$	–	0.188	0.204	–	0.073	0.079	–	0.190	0.205	–	0.073	0.079
	$\tilde{\beta}_2$	–	0.241	0.253	–	0.087	0.092	–	0.242	0.255	–	0.087	0.092
	$\tilde{\beta}_3$	–	0.364	0.389	–	0.136	0.145	–	0.368	0.394	–	0.136	0.145
	$\tilde{\rho}$	–	0.025	0.029	–	0.008	0.009	–	0.026	0.031	–	0.008	0.009
Data generating process according to random effects specification													
Data augmentation	$\tilde{\xi}$	1.223	1.207	1.196	1.217	1.196	1.201	0.426	0.428	0.424	0.237	0.243	0.242
	$\tilde{\beta}_1$	0.105	0.133	0.133	0.045	0.056	0.058	0.098	0.119	0.121	0.045	0.056	0.058
	$\tilde{\beta}_2$	0.100	0.144	0.148	0.048	0.065	0.069	0.095	0.132	0.134	0.049	0.066	0.069
	$\tilde{\beta}_3$	0.197	0.245	0.285	0.086	0.104	0.125	0.183	0.218	0.251	0.086	0.104	0.124
	$\tilde{\rho}$	0.015	0.021	0.022	0.005	0.007	0.007	0.012	0.014	0.015	0.005	0.006	0.006
Complete case	$\tilde{\xi}$	–	1.167	1.160	–	1.181	1.184	–	0.442	0.444	–	0.247	0.248
	$\tilde{\beta}_1$	–	0.166	0.181	–	0.069	0.071	–	0.139	0.149	–	0.068	0.070
	$\tilde{\beta}_2$	–	0.182	0.200	–	0.080	0.087	–	0.154	0.166	–	0.080	0.086
	$\tilde{\beta}_3$	–	0.298	0.324	–	0.124	0.134	–	0.252	0.267	–	0.123	0.132
	$\tilde{\rho}$	–	0.028	0.029	–	0.009	0.009	–	0.018	0.019	–	0.008	0.009

Table 3. Simulated data with fixed and random effects as true data generating process analysed with data augmentation and complete case approach. Biases of mean estimates for Scenarios I, II and III with $\rho = 0.2$.

		Fixed effects analysis						Random effects analysis					
		Small sample			Large sample			Small sample			Large sample		
		I	II	III	I	II	III	I	II	III	I	II	III
Data generating process according to fixed effects specification													
Data augmentation	$\tilde{\xi}$	0.179	0.221	0.221	0.052	0.065	0.070	2.647	2.590	2.689	0.001	−0.075	−0.083
	$\tilde{\beta}_1$	−0.003	−0.010	0.010	−0.003	−0.001	0.010	−0.003	−0.011	0.010	−0.002	−0.001	0.010
	$\tilde{\beta}_2$	−0.007	−0.003	0.017	0.004	0.000	0.010	−0.008	−0.004	0.016	0.004	0.000	0.010
	$\tilde{\beta}_3$	0.002	−0.002	−0.126	−0.003	−0.003	−0.065	0.002	−0.001	−0.126	−0.003	−0.003	−0.065
	$\tilde{\rho}$	0.001	0.001	0.001	−0.000	−0.000	−0.000	0.001	0.001	0.001	0.000	0.000	0.000
Complete case	$\tilde{\xi}$	–	0.264	0.286	–	0.087	0.103	–	2.660	2.767	–	−0.056	−0.059
	$\tilde{\beta}_1$	–	−0.001	−0.000	–	−0.002	−0.003	–	0.000	0.002	–	−0.002	−0.003
	$\tilde{\beta}_2$	–	0.006	0.003	–	−0.001	−0.002	–	0.006	0.002	–	−0.001	−0.002
	$\tilde{\beta}_3$	–	−0.006	−0.005	–	0.004	0.006	–	−0.006	−0.008	–	0.004	0.006
	$\tilde{\rho}$	–	0.009	0.012	–	0.001	0.001	–	0.009	0.012	–	0.001	0.002
Data generating process according to random effects specification													
Data augmentation	$\tilde{\xi}$	−0.079	−0.086	−0.087	0.015	0.011	0.009	0.005	0.002	−0.004	0.015	0.014	0.015
	$\tilde{\beta}_1$	−0.006	−0.012	0.010	−0.001	−0.002	0.009	0.001	0.000	0.018	−0.000	−0.001	0.009
	$\tilde{\beta}_2$	−0.004	−0.006	0.018	0.003	0.002	0.013	−0.000	−0.005	0.015	0.003	0.001	0.012
	$\tilde{\beta}_3$	−0.012	−0.017	−0.158	−0.004	−0.002	−0.068	−0.009	−0.016	−0.134	−0.003	−0.002	−0.066
	$\tilde{\rho}$	−0.008	−0.013	−0.013	−0.002	−0.002	−0.002	−0.001	−0.001	−0.001	−0.000	−0.000	−0.000
Complete case	$\tilde{\xi}$	–	−0.109	−0.112	–	−0.008	−0.006	–	0.007	0.007	–	0.020	0.021
	$\tilde{\beta}_1$	–	−0.024	−0.025	–	−0.003	−0.001	–	−0.002	−0.002	–	−0.003	−0.000
	$\tilde{\beta}_2$	–	−0.013	−0.005	–	0.001	−0.002	–	0.000	0.006	–	0.001	−0.001
	$\tilde{\beta}_3$	–	−0.019	−0.024	–	0.002	0.003	–	−0.012	−0.017	–	0.002	0.004
	$\tilde{\rho}$	–	−0.015	−0.016	–	−0.002	−0.002	–	0.000	0.001	–	0.001	0.001

of prior uncertainty, we perform an additional analysis of a randomly chosen small sample size data set with fixed effects DGP of Scenario I, now with all multivariate normal prior variances set to 100. The resulting log marginal likelihoods mark the random effects model as the superior specification. As the resulting log-likelihoods just changed negligibly, the fixed effects model remains the favoured specification in terms of the BIC. This contradiction in model selections clarifies the higher sensitivity of the marginal likelihood to prior uncertainty compared to the BIC. For the small sample sizes, in Scenarios II and III the rates remain at 99.8% and 98.8%, respectively, and are congruent with the decisions for Scenario I. For the large samples, the rates amount to 100% and 99.6% and hence only differ slightly from those of the small samples. In contrast, if we assume a random effects model as the DGP, the random effects specification is almost always in favour (Table 1), both deciding according to the BIC and the log marginal likelihood for Scenario I. This is due to the penalisation of the non-parsimonious number of parameters of the fixed effects model, as the latent heterogeneity is sufficiently captured by the random effects estimators. An exception for this is the large samples, as the random effects specification is the model of choice in 92.6%. In case of missing values, the random effects model is chosen as the favourable specification in 100% of the cases both for Scenarios II and III. Thus, according to our model evaluation criteria, the choice of analysis model predominantly coincides with the underlying data generating process, thus yielding reliable model selections.

The estimations for the DGPs with $\rho = 0.8$ indicate equivalent results in terms of coverage rates, RMSEs, biases and model decisions. However, biases and RMSEs are slightly higher. Given small sample sizes, the coverage rates of the fixed effects model estimations of around 0.91 for all three scenarios (see Table 6) point at a Nickell bias (Nickell 1981; Hsiao, Hashem Pesaran & Kamil Tahmiscioglu 2002), diminishing for a growing T, as the large sample results reveal. The efficiency of the data augmentation approach remains higher than for the complete case estimators. Again, the coverage rates of the intercept and the dynamic component decrease if the DGP is determined by random effects. The random effects analyses yield high coverage rates for both sample sizes, as long as the data are also generated due to a random effects specification. If the degree of latent heterogeneity is due to the fixed effects DGP and thus not in line with the random effects specification, we obtain lower coverage rates for the intercepts both for the small and large sample size. Also the model evaluation criteria altogether indicate similar model selection rates as for $\rho = 0.2$.

4.2. Comparison to classical estimation

To further investigate the efficiency of our suggested methods, Table 4 provides a comparison of our procedure to the classical estimation approach developed by Arellano & Bond (1991), For this comparison, we generated additional 500 data sets due to the fixed effects model specification with only taking individual- but no time-specific latent heterogeneity into account. Also, we consider the situation with $\rho = 0.8$ and, thus, stronger serial dependence and a homoscedastic variance scenario. The homoscedastic variance is implied via a single draw from the uniform distribution used in the heteroscedastic case. The parameter setting is given as before, whereas the MCMC and prior settings are as given in the supplementary material (Section C). Accordingly, the Arellano–Bond estimator is based on differences as well as orthogonal deviations, where we also refer to robust standard errors. Furthermore, we consider $T = 5$, as this mimics a typical situation in application and yields slightly reduced computation times. Furthermore, we specified a continuous regressor

Table 4. Simulated data with individual-specific fixed effects as true data generating process with homoscedastic and heteroscedastic variance structure. Analysis with fixed effects model with data augmentation and Arellano–Bond type estimator. Results for the Arellano–Bond type estimator are reported for first difference instruments and orthogonal deviations. Results obtained using robust standard errors are given in parenthesis. Coverage rates, root mean squared errors of mean estimates, and biases of mean estimates for Scenarios I, II and III with $\rho = 0.8$, $N = 50$, and $T = 5$.

		Homoscedasticity			Heteroscedasticity		
		I	II	III	I	II	III
		Coverage rates			Coverage rates		
Fixed effects - data augmentation	$\tilde{\beta}_1$	0.934	0.954	0.950	0.928	0.918	0.928
	$\tilde{\beta}_2$	0.950	0.946	0.948	0.946	0.952	0.972
	$\tilde{\beta}_3$	0.928	0.942	0.942	0.938	0.942	0.926
	$\tilde{\rho}$	0.946	0.954	0.966	0.934	0.944	0.948
Arellano-Bond - differencing	$\tilde{\beta}_1$	0.946	0.944	0.946	0.958 (0.948)	0.942 (0.928)	0.938 (0.924)
	$\tilde{\beta}_2$	0.956	0.942	0.936	0.956 (0.944)	0.936 (0.922)	0.928 (0.904)
	$\tilde{\beta}_3$	0.936	0.944	0.922	0.964 (0.950)	0.952 (0.910)	0.964 (0.926)
	$\tilde{\rho}$	0.952	0.946	0.934	0.962 (0.974)	0.950 (0.958)	0.968 (0.958)
Arellano-Bond - orthogonal	$\tilde{\beta}_1$	0.950	0.962	0.952	0.970 (0.954)	0.942 (0.918)	0.952 (0.926)
	$\tilde{\beta}_2$	0.960	0.950	0.938	0.962 (0.944)	0.944 (0.930)	0.946 (0.922)
	$\tilde{\beta}_3$	0.938	0.956	0.922	0.956 (0.950)	0.952 (0.918)	0.956 (0.922)
	$\tilde{\rho}$	0.976	0.978	0.964	0.976 (0.970)	0.968 (0.956)	0.982 (0.964)
		Homoscedasticity			Heteroscedasticity		
		I	II	III	I	II	III
		Root mean squared errors			Root mean squared errors		
Fixed effects - data augmentation	$\tilde{\beta}_1$	0.172	0.203	0.214	0.173	0.221	0.227
	$\tilde{\beta}_2$	0.166	0.224	0.227	0.166	0.217	0.220
	$\tilde{\beta}_3$	0.188	0.227	0.239	0.186	0.230	0.234
	$\tilde{\rho}$	0.012	0.013	0.013	0.011	0.013	0.013
Arellano-Bond - differencing	$\tilde{\beta}_1$	0.263	0.345	0.407	0.272	0.405	0.456
	$\tilde{\beta}_2$	0.246	0.399	0.451	0.264	0.444	0.499
	$\tilde{\beta}_3$	0.217	0.347	0.396	0.224	0.354	0.372
	$\tilde{\rho}$	0.063	0.077	0.090	0.076	0.103	0.107
Arellano-Bond - orthogonal	$\tilde{\beta}_1$	0.235	0.302	0.376	0.233	0.364	0.425
	$\tilde{\beta}_2$	0.219	0.344	0.410	0.234	0.387	0.463
	$\tilde{\beta}_3$	0.203	0.306	0.377	0.212	0.329	0.370
	$\tilde{\rho}$	0.064	0.071	0.085	0.074	0.092	0.102
		Homoscedasticity			Heteroscedasticity		
		I	II	III	I	II	III
		Biases of mean estimates			Biases of mean estimates		
Fixed effects - data data augmentation	$\tilde{\beta}_1$	0.013	0.014	0.037	0.000	−0.004	0.022
	$\tilde{\beta}_2$	−0.006	−0.008	0.018	−0.001	−0.001	0.024

(Continued)

Table 4. *(Continued)*

	$\tilde{\beta}_3$	−0.007	−0.003	−0.071	0.004	0.007	−0.065
	$\tilde{\rho}$	−0.001	−0.001	−0.001	0.001	0.001	0.001
Arellano-Bond - differencing	$\tilde{\beta}_1$	−0.001	−0.004	−0.016	−0.027	−0.071	−0.086
	$\tilde{\beta}_2$	−0.013	0.006	−0.010	−0.028	−0.039	−0.047
	$\tilde{\beta}_3$	−0.016	−0.024	−0.028	−0.006	−0.009	−0.024
	$\tilde{\rho}$	−0.008	−0.010	−0.017	−0.011	−0.019	−0.030
Arellano-Bond - orthogonal	$\tilde{\beta}_1$	−0.003	−3.4e − 5	−0.016	−0.018	−0.037	−0.061
	$\tilde{\beta}_2$	−0.013	0.006	−0.008	−0.021	−0.028	−0.034
	$\tilde{\beta}_3$	−0.011	−0.015	−0.011	−0.007	−0.008	−0.029
	$\tilde{\rho}$	−0.007	−0.009	−0.016	−0.010	−0.016	−0.028

Notes: Arellano-Bond type one-step estimators calculated in STATA 15 (StataCorp 2017) using command `xtabond2` and options `robust` or `orthogonal`, see Roodman (2006).

$X_{3,i,t}$ instead of the ordered categorical variable from above in order to simplify the parameterisation and interpretation without the necessity of recoding. The data sets are generated with $\delta = 0$. However, the data augmentation estimation is not performed under this restriction. For comparison, we report results on β and ρ only.

In the case of homoscedastic error variances (left half of Table 4), the Arellano–Bond estimates show satisfying coverage rates, efficiencies and biases compared to our suggested approach. However, the insertion of missing values (Scenarios *II* and *III*) highlights the efficiency of our approach, as RMSEs for the data augmentation approach just worsen sparsely in contrast to the Arellano–Bond estimates. The results reveal that the use of orthogonal deviations is beneficial in the considered case with substantial serial dependence, as the bias is reduced further. However, the bias reduction does not outperform the positive effects on efficiency in terms of RMSEs arising from handling the missing values via data augmentation. This impression is confirmed if period-specific error variances are additionally taken into account. When comparing the RMSEs, it should be taken into account that the suggested unconditional model formulation considers not only additional N observations, but also includes $K+2$, that is $K+1$, more parameters in the heteroscedastic and homoscedastic case respectively. The differences in the RMSEs are not solely driven by this additional information, but generally by the extra information available for estimation via data augmentation. Data augmentation allows to consider information from entities $[y_{it}, \boldsymbol{Z}_{it}]$ having missing values in single variables.

4.3. Empirical illustration

In our empirical illustration, we analyse the current account balance dynamics of the 35 OECD member states and China from 1991 until 2015, comparing our data augmentation approach to a complete case estimation. We use the World Development Indicators provided by the World Bank (The World Bank 2017) and regress the current account balance (percent of the GDP) on its previous year's outcome, adding the covariates foreign direct investment in net flows, the annual GDP growth in percent, GDP per capita (constant 2010 in 1000 U.S. dollars) and trade (percent of GDP), each lagged of order one to account for possible inconsistencies due to endogeneity problems (see e.g. Chinn & Prasad 2003; Aßmann 2012). We decide for 1992 as the initial period of the dependent variable, following the general model structure of (1), to mitigate the effects of structural change occurring after 1990. The prior specifications and starting values are chosen the same as in our simulation studies. For the empirical illustration 40,000 draws after a burn-in phase of 10,000 iterations have been found sufficient to ensure convergence. For completeness, the estimation results with regard to the variance and latent heterogeneity parameters are provided in Tables 9 and 10.

As Table 5 also contains the rates of missing values of each variable, the complete case estimations yield a loss of 86 and thus about 10% of the 864 country-period observations at hand, resulting in 778 observations for which all variables are available. As the Arellano–Bond approach conditions on the initial period, 828 observations are available for estimation. The number of observations is further reduced as 111 missings occur over all variables, where the used STATA 15 implementation does not check for overlap in the periods and units with missing values. The results illustrate the increase of efficiency for both model types due to the data augmentation approach, as we obtain smaller standard deviations for all slope and dynamic component parameters compared to the complete case analyses. Hence, the 90% highest posterior density region estimated by the data augmentation approach for GDP per capita does not contain zero. In contrast, the corresponding 90% highest posterior density region estimated by the complete case analysis does include zero. An increase in the level of influence is also obtained for the life expectancy parameters, as for the data augmentation the 90% highest posterior density region of the fixed effects estimator does not incorporate zero, while for the complete case estimator the effect does not differ from zero at any conventional level. Furthermore, the data augmentation estimator for the random effects model exceeds the 95% highest posterior density region, while the complete case estimator reaches the 90% level. The Arellano–Bond estimates are in line with the obtained estimates of the Bayesian approach. The estimated confidence intervals of the two estimation approaches are overlapping and no substantial differences are revealed. However, the efficiency gains documented in the simulation studies also show off in the estimated standard deviations, which are smaller for the Bayesian approach. We are aware that this comparison may be criticised given the different conceptual frameworks of the two estimation approaches.

To complete the discussion of estimation results, the random effects specification is according to the log marginal likelihood the preferred model specification with -1947.330 compared to -2047.714 for the random and fixed effect specification respectively. The robustness of this model selection with regard to prior variance has been checked and confirmed via changing variance parameter of the considered prior distribution of $\boldsymbol{\gamma}$, c, and a from 50 to 10 and 100. This is in line with the moderate heterogeneity in variance as well as individual and time specific effects documented in Tables 8 and 9.

5. Conclusion and outlook

This paper provides a Bayesian estimation approach to deal with missing values in linear dynamic panel models incorporating individual- and time-specific heterogeneity. The set of full conditional distributions used to sample from the corresponding posterior distribution is thereby augmented with the full conditional posterior distribution of the missing values in the dependent variable. Likewise, to handle also incomplete data in the covariates via data augmentation, classification and sequential regression trees are used to provide non-parametric approximations for the corresponding full conditional distributions. This way to model the full conditional distributions of missing values offers a flexible yet parsimonious

Table 5. Estimates of empirical illustration – current account balance (% of GDP, 7.75% missing values).

Variable (Rate of missing values per variable in %)	Data augmentation		Complete case		Arellano-Bond
	Fixed effects	Random effects	Fixed effects	Random effects	AB
Regression coefficients ξ, β, ρ					
Constant	−6.656	−4.924*	−5.921	−5.056*	–
	(5.584)	(3.467)	(5.722)	(3.529)	
Foreign direct investment, net inflows$_{t-1}$	−0.004	−0.003	−0.002	−0.002	−0.009
(4.51)	(0.008)	(0.008)	(0.008)	(0.008)	(0.005)
GDP growth (annual %)$_{t-1}$	−0.144**	−0.099**	−0.151**	−0.093**	−0.246**
(2.08)	(0.036)	(0.033)	(0.036)	(0.034)	(0.074)
GDP per capita (constant 2010 in 1000 US$)$_{t-1}$	−0.029	0.013*	−0.013	0.011	0.043
(1.50)	(0.026)	(0.009)	(0.029)	(0.009)	(0.075)
Life expectancy at birth, total (years)$_{t-1}$	0.098*	0.057	0.074	0.059	0.331
(0,00)	(0.070)	(0.046)	(0.073)	(0.047)	(0.267)
Trade (% of GDP)$_{t-1}$	0.016**	0.004*	0.025**	0.005*	0.032**
(1.39)	(0.006)	(0.003)	(0.007)	(0.003)	(0.016)
Current account balance (% of GDP)$_{t-1}$	0.731**	0.796**	0.726**	0.802**	0.645**
(8.09)	(0.024)	(0.025)	(0.025)	(0.026)	(0.063)
Initial period regression coefficients δ					
Foreign direct investment, net inflows$_1$	0.510*	0.302	0.237	0.027	–
(19.44)	(0.372)	(0.343)	(0.551)	(0.548)	
GDP growth (annual %)$_1$	0.138*	0.170*	0.259**	0.268**	–
(13.89)	(0.109)	(0.106)	(0.116)	(0.116)	
GDP per capita (constant 2010 in 1000 US$)$_1$	0.036	0.049*	0.062**	0.070**	–
(11.11)	(0.037)	(0.032)	(0.036)	(0.034)	
Life expectancy at birth, total (years)$_1$	−0.135*	−0.065**	−0.125*	−0.066**	–
(0.00)	(0.095)	(0.020)	(0.093)	(0.021)	
Trade (% of GDP)$_1$	0.033*	0.037*	0.018	0.027	–
(8.33)	(0.025)	(0.024)	(0.025)	(0.026)	
NT available for analysis due to missing values	864		778		717
Marg. Likelihood	−2047.714	−1947.33	–	–	–

Notes: Bayesian standard deviations in parentheses; ** indicates 95% highest posterior density region excludes 0; * 90% denotes highest posterior density region excludes 0; AB refers to Arellano–Bond type estimates based on orthogonal transformation and robust standard errors; values in parentheses below estimates denote corresponding estimated standard deviations.

handling of relationships among the variables and latent structures. However, as pointed out by a reviewer, parametric models for the missing covariate values could also be considered.

Next to estimation, the paper extends the approach of Chib (1995) to calculate the marginal likelihood in order to provide non-nested model comparisons in case of incomplete data in the linear dynamic panel context. As simulations reveal, the suggested procedure offers both, efficiency gains in parameter estimation and reliable model selection. Comparison with the Arellano–Bond estimator yields that in some missing data situations neglect of missing values seems worse than neglect of unbiased instrument variable estimation. The empirical analysis exemplifies the wide field of applicability of the suggested approach and confirms the improved estimation efficiency. The efficiency gains stem from the additional information that can be exploited when augmenting the missing values. Augmenting the missing values allows for consideration of entities $[y_{it}, \boldsymbol{Z}_{it}]$ that show missing values in single variables and would otherwise be dismissed. This extends also to the suggested unconditional modelling approach allowing to consider initial period observations of the dependent variable. Augmentation seems to improve the efficiency in situations where the extra information is not set off by the extra parameters involved in the set of full conditional distributions of missing values. Moreover, the calculation of the log marginal likelihood allows for model selection.

Furthermore, the suggested approach is not limited to linear dynamic panel regressions and the set of full conditional distributions for missing values discussed in this paper, as the sampling scheme could handle other models as well and any valid alternative to model the full conditional distributions of missing values. Another extension could focus on polychotomous response variables via an additional augmentation step (Albert & Chib 1993; Geweke & Keane 2001), thus opening a wider range of application fields. Extending the approach towards specifications where the latent individual-specific heterogeneity is also captured via the slope parameters and lagged-dependent variable models of higher order could be considered. Also possibilities to combine the instrument estimation and data augmentation approach may be addressed in future research.

Supporting information

Additional Supporting Information may be found in the online version of this article at http://wileyonlinelibrary.com/journal/anzs

Data S1. Supplementary materials

References

ALBERT, J.H. & CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669.

ANDO, T. (2010). *Bayesian Model Selection and Statistical Modeling*. Statistics: textbooks and monographs, Boca Raton: CRC Press/Chapman & Hall Book.

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

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

AßMANN, C. (2012). Determinants and costs of current account reversals under heterogeneity and serial correlation. *Applied Economics* **44**, 1685–1700.

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

BURGETTE, L.F. & REITER, J.P. (2010). Multiple imputation for missing data via sequential regression trees. *American Journal of Epidemiology* **172**, 1070–1076.

CASELLA, G. & GEORGE, E.I. (1992). Explaining the Gibbs sampler. *The American Statistician* **46**, 167.

CHIB, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* **90**, 1313–1321.

CHIB, S. & JELIAZKOV, I. (2001). Marginal likelihood from the Metropolis–Hastings output. *Journal of the American Statistical Association* **96**, 270–281.

CHINN, M.D. & PRASAD, E.S. (2003). Medium-term determinants of current accounts in industrial and developing countries: an empirical exploration. *Journal of International Economics* **59**, 47–76.

DOOVE, L.L., VAN BUUREN, S. & DUSSELDORP, E. (2014). Recursive partitioning for missing data imputation in the presence of interaction effects. *Computational Statistics & Data Analysis* **72**, 92–104.

EVERAERT, G. (2013). Orthogonal to backward mean transformation for dynamic panel data models. *The Econometrics Journal* **16**, 179–221.

FRÜHWIRTH-SCHNATTER, S. (2004). Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *The Econometrics Journal* **7**, 143–167.

FRÜHWIRTH-SCHNATTER, S. & KAUFMANN, S. (2008). Model-based clustering of multiple time series. *Journal of Business & Economic Statistics* **26**, 78–89.

GELFAND, A.E. & SMITH, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398.

GELMAN, A., MENG, X.L. & STERN, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica* **6**, 733–760.

GEMAN, S. & GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.

GEWEKE, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* **57**, 1317.

GEWEKE, J. & KEANE, M. (2001). Computationally intensive methods for integration in econometrics. In *Handbook of Econometrics*, eds. J.J. Heckman & E.E. Leamer, *Handbook of Econometrics*, vol. 5. pp. 3463–3568. Amsterdam: Elsevier.

HSIAO, C. (2014). *Analysis of Panel Data*. Econometric Society Monographs, 3rd edn. New York, NY: Cambridge University Press.

HSIAO, C., HASHEM PESARAN, M. & KAMIL TAHMISCIOGLU, A. (2002). Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. *Journal of Econometrics* **109**, 107–150.

HSIAO, C. & PESARAN, M.H. (2008). Random coefficients models. In *The Econometrics of Panel Data*, ed. L. Mátyás, *Advanced Studies in Theoretical and Applied Econometrics*, vol. 46. pp. 185–213. Berlin and Heidelberg: Springer.

IBRAHIM, J.G., ZHU, H. & TANG, N. (2008). Model selection criteria for missing-data problems using the EM algorithm. *Journal of the American Statistical Association* **103**, 1648–1658.

JEFFREYS, H. (1961). *The Theory of Probability*, 3rd edn. Oxford: Clarendon Press and Oxford University Press.

KASS, R.E. & RAFTERY, A.E. (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773.

LI, K.H. (1988). Imputation using Markov chains. *Journal of Statistical Computation and Simulation* **30**, 57–79.

LIESENFELD, R. & RICHARD, J.F. (2010). Efficient estimation of probit models with correlated errors. *Journal of Econometrics* **156**, 367–376.

LITTLE, R.J.A. & RUBIN, D.B. (2002). *Statistical Analysis with Missing Data*. Wiley Series in Probability and Statistics, 2nd edn. Hoboken, NJ: Wiley.

LIU, F., ZHANG, P., ERKAN, I. & SMALL, D.S. (2016). Bayesian inference for random coefficient dynamic panel data models. *Journal of Applied Statistics* **120**, 1–17.

MESTERS, G. & KOOPMAN, S.J. (2014). Generalized dynamic panel data models with random effects for cross-section and time. *Journal of Econometrics* **180**, 127–140.

MESTERS, G., VAN DER GEEST, V. & BIJLEVELD, C. (2016). Crime, employment and social welfare: an individual-level study on disadvantaged males. *Journal of Quantitative Criminology* **32**, 159–190.

MILLIMET, D.L. & MCDONOUGH, I.K. (2016). Dynamic panel data models with irregular spacing: with an application to early childhood development. *Journal of Applied Econometrics* **77**, 598.

NICKELL, S. (1981). Biases in dynamic models with fixed effects. *Econometrica* **49**, 1417.
R CORE TEAM (2017). R: A language and environment for statistical computing.
RAFTERY, A.E. (1995). Bayesian model selection in social research. *Sociological Methodology* **25**, 111.
RICHARD, J.F. & ZHANG, W. (2007). Efficient high-dimensional importance sampling. *Journal of Econometrics* **141**, 1385–1411.
ROODMAN, D. (2006). How to Do xtabond2: an introduction to “Difference” and “System” GMM in Stata, Center for Global Development, Washington, Working Paper 103.
RUBIN, D.B. (1981). The Bayesian bootstrap. *The Annals of Statistics* **9**, 130–134.
SCHWARZ, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464.
STATACORP (2017). Stata statistical software: Release 15.
TANNER, M.A. & WONG, W.H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82**, 528–540.
THE WORLD BANK (2017). World development indicators (1960–2016).
VAN BUUREN, S., GROOTHUIS-OUDSHOORN, K., ROBITZSCH, A., VINK, G., DOOVE, L. & JOLANI, S. (2015). MICE: Multivariate imputation by chained equations.