Plug-in machine learning for partially linear mixed-effects models with repeated measurements

Journal Article

Author(s):
Emmenegger, Corinne; Bühlmann, Peter

Publication date:
2023-12

Permanent link:
https://doi.org/10.3929/ethz-b-000614568

Originally published in:
Scandinavian Journal of Statistics 50(4), https://doi.org/10.1111/sjos.12639

Funding acknowledgement:
786461 - Statistics, Prediction and Causality for Large-Scale Data (EC)

Received: 22 February 2022 | Revised: 20 November 2022 | Accepted: 3 March 2023

DOI: 10.1111/sjos.12639

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

Plug-in machine learning for partially linear mixed-effects models with repeated measurements

Corinne Emmenegger | **Peter Bühlmann**

Seminar for Statistics, ETH Zurich, Zurich, Switzerland

Correspondence
Corinne Emmenegger, Seminar for Statistics, ETH Zurich, Zurich 8092 Zurich, Switzerland
Email: emmenegger@stat.math.ethz.ch

Funding information
H2020 European Research Council, Grant/Award Number: 786461

Abstract

Traditionally, spline or kernel approaches in combination with parametric estimation are used to infer the linear coefficient (fixed effects) in a partially linear mixed-effects model for repeated measurements. Using machine learning algorithms allows us to incorporate complex interaction structures, nonsmooth terms, and high-dimensional variables. The linear variables and the response are adjusted nonparametrically for the nonlinear variables, and these adjusted variables satisfy a linear mixed-effects model in which the linear coefficient can be estimated with standard linear mixed-effects methods. We prove that the estimated fixed effects coefficient converges at the parametric rate, is asymptotically Gaussian distributed, and semiparametrically efficient. Two simulation studies demonstrate that our method outperforms a penalized regression spline approach in terms of coverage. We also illustrate our proposed approach on a longitudinal dataset with HIV-infected individuals. Software code for our method is available in the R-package `dmlalg`.

KEYWORDS

between-group heterogeneity, CD4 dataset (HIV), dependent data, double machine learning, fixed effects estimation, longitudinal data, semiparametric estimation

1 INTRODUCTION

Repeated measurements data consists of observations from several experimental units, subjects, or groups under different conditions. This grouping or clustering of the individual responses into experimental units typically introduces dependencies: the different units are assumed to be independent, but there may be heterogeneity across units and correlation within units.

Mixed-effects models provide a powerful and flexible tool to analyze grouped data by incorporating fixed and random effects. Fixed effects are associated with the entire population, and random effects are associated with individual groups and model the heterogeneity across them and the dependence structure within them (Pinheiro & Bates, 2000). Linear mixed-effects models (Demidenko, 2004; Laird & Ware, 1982; Pinheiro & Bates, 2000; Verbeke & Molenberghs, 2000) impose a linear relationship between all covariates and the response. Partially linear mixed-effects models (Zeger & Diggle, 1994) extend the linear ones.

We consider the partially linear mixed-effects model

$$\mathbf{Y}_i = \mathbf{X}_i \beta_0 + g(\mathbf{W}_i) + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \tag{1}$$

for groups $i \in \{1,\ldots,N\}$. There are n_i observations per group i. The unobserved random variable $\mathbf{b}_i$, called random effect, introduces correlation within its group i because all n_i observations within this group share $\mathbf{b}_i$. We make the assumption generally made that both the random effect $\mathbf{b}_i$ and the error term $\boldsymbol{\varepsilon}_i$ follow a Gaussian distribution (Pinheiro & Bates, 2000). The matrices $\mathbf{Z}_i$ assigning the random effects to group-level observations are fixed. The linear covariables $\mathbf{X}_i$ and the nonparametric and potentially high-dimensional covariables $\mathbf{W}_i$ are observed and random, and they may have dependent columns. Furthermore, the nonparametric covariables may contain nonlinear transformations and interaction terms of the linear ones. Please see Assumption 1 in Section 2 for further details.

Our aim is to estimate and make inference for the so-called fixed effect β_0 in (1) in the presence of a highly complex g using general machine learning algorithms. The parametric component β_0 provides a simple summary of the covariate effects that are of main scientific interest. The nonparametric component g enhances model flexibility because time trends and further covariates with possibly nonlinear and interaction effects can be modeled nonparametrically.

Repeated measurements, or longitudinal, data is omnipresent in empirical research. For example, assume we want to study the effect of a treatment over time. Observing the same subjects repeatedly presents three main advantages over having cross-sectional data. First, subjects can serve as their own controls. Second, the between-subject variability is explicitly modeled and can be excluded from the experimental error. This yields more efficient estimators of the relevant model parameters. Third, data can be collected more reliably (Davis, 2002; Fitzmaurice et al., 2011).

Various approaches have been considered in the literature to estimate the nonparametric component g in (1): kernel methods (Chen & Cao, 2017; Hart & Wehrly, 1986; Taavoni & Arashi, 2021b; Zeger & Diggle, 1994), backfitting (Taavoni & Arashi, 2021b; Zeger & Diggle, 1994), spline methods (Aniley et al., 2019; Kim et al., 2017; Li & Zhu, 2010; Qin & Zhu, 2007, 2009; Rice & Silverman, 1991; Zhang, 2004), and local linear regression (Liang, 2009; Taavoni & Arashi, 2021b).

Our aim is to make inference for β_0 in the presence of potentially highly complex effects of $\mathbf{W}_i$ on $\mathbf{X}_i$ and $\mathbf{Y}_i$. First, we adjust $\mathbf{X}_i$ and $\mathbf{Y}_i$ for $\mathbf{W}_i$ by regressing $\mathbf{W}_i$ out of them using machine learning algorithms. These machine learning algorithms may yield biased results, especially if regularization methods are used, like for instance with the lasso (Tibshirani, 1996). Second, we fit a linear mixed-effects model to these regression residuals to estimate β_0. Our estimator of β_0 converges at the optimal $1/\sqrt{N}$ rate, follows a Gaussian distribution asymptotically, and is semiparametrically efficient.

We adapt double machine learning techniques of Chernozhukov et al. (2018) to estimate β_0 using general machine learning algorithms. To the best of our knowledge, this is the first work to allow the nonparametric nuisance components of a partially linear mixed-effects model to be estimated with arbitrary machine learners like random forests (Breiman, 2001) or the lasso (Bühlmann & van de Geer, 2011; Tibshirani, 1996). In contrast to the setting and proofs of Chernozhukov et al. (2018), we have dependent data and need to incorporate this accordingly. Chernozhukov et al. (2018) introduce double machine learning and develop estimation of the low-dimensional linear regression parameter vector in a partially linear model. Their estimator converges at the parametric rate and is asymptotically Gaussian due to Neyman orthogonality and sample splitting with cross-fitting. We would like to remark that nonparametric nuisance components can be estimated without sample splitting and cross-fitting if the underlying function class satisfies some entropy conditions; see for instance Mammen and van de Geer (1997). However, these conditions limit the complexity of the function class, and machine learning algorithms usually do not satisfy them. Particularly, these conditions fail to hold if the dimension of the nonparametric variables increases with the sample size (Chernozhukov et al., 2018). We show that the desirable properties of double machine learning also hold in the context of partially linear mixed-effects models: such a further development of plug-in machine learning methods is nontrivial and practically highly relevant.

1.1 | Additional literature

Expositions and overviews of mixed-effects modeling techniques can be found in Pinheiro (1994), Davidian and Giltinan (1995), Vonesh and Chinchilli (1997), Pinheiro and Bates (2000), and Davidian and Giltinan (2003).

Zhang et al. (1998) consider partially linear mixed-effects models and estimate the nonparametric component with natural cubic splines. They treat the smoothing parameter as an extra variance component that is jointly estimated with the other variance components of the model. Masci et al. (2019) consider partially linear mixed-effects models for unsupervised classification with discrete random effects. Schelldorfer et al. (2011) consider high-dimensional linear mixed-effects models where the number of fixed effects coefficients may be much larger than the overall sample size. To estimate and make inference for the first, say, d components of the linear coefficient in such a high-dimensional mixed-effects model, our approach may consider the remaining components as an additive contribution $\mathbf{W}_i\beta_{0,-(1:d)}$ in the model and may adjust for them using the lasso (Tibshirani, 1996). Debiased fixed effects estimators in high-dimensional linear mixed effects models are studied by Li et al. (2021) and Bradic et al. (2020). Taavoni and Arashi (2021a) employ a regularization approach in generalized partially linear mixed-effects models using regression splines to approximate the nonparametric component. Wood and Scheipl (2020) use penalized regression splines where the penalized components are treated as random effects.

The unobserved random variables in the partially linear mixed-effects model (1) are assumed to follow a Gaussian distribution. Taavoni et al. (2021) introduce multivariate t partially linear mixed-effects models for longitudinal data. They consider t-distributed random effects to account for outliers in the data. Fahrmeir and Kneib (2011, chapter 4) relax the assumption of Gaussian random effects in generalized linear mixed models. They consider nonparametric Dirichlet processes and Dirichlet process mixture priors for the random effects. Ohinata (2012, chapter 3) consider partially linear mixed-effects models and make no distributional assumptions for the random terms, and the nonparametric component is estimated with kernel methods. Lu (2016) consider a partially linear mixed-effects model that is nonparametric in time and that features asymmetrically distributed errors and missing data.

Furthermore, methods have been developed to analyze repeated measurements data that are robust to outliers. Guoyou and Zhongyi (2008) consider robust estimating equations and estimate the nonparametric component with a regression spline. Tang et al. (2015) consider median-based regression methods in a partially linear model with longitudinal data to account for highly skewed responses. Lin et al. (2018) present an estimation technique in partially linear models for longitudinal data that is doubly robust in the sense that it simultaneously accounts for missing responses and mismeasured covariates.

It is prespecified in the partially linear mixed-effects model (1) which covariates are modeled with random effects. Simultaneous variable selection for fixed effects variables and random effects has been developed by Bondell et al. (2010); Ibrahim et al. (2011). They use penalized likelihood approaches. Li and Zhu (2010) use a nonparametric test to test the existence of random effects in partially linear mixed-effects models. Zhang and Xue (2020) propose a variable selection procedure for the linear covariates of a generalized partially linear model with longitudinal data.

1.2 | Outline of the paper

Section 2 presents our plug-in machine learning estimator of the linear coefficient in a partially linear mixed-effects model. Section 3 presents our numerical results. Proofs and technical assumptions are presented in the Data S1.

1.3 | Notation

We denote by $[N]$ the set $\{1, 2, \ldots, N\}$. We add the probability law P as a subscript to the probability operator $\mathbb{P}$ and the expectation operator $\mathbb{E}$ whenever we want to emphasize the corresponding dependence. We denote the $L^p(P)$ norm for $p \geq 1$ by $\|\cdot\|_{P,p}$ and the Euclidean or operator norm by $\|\cdot\|$, depending on the context. We implicitly assume that given expectations and conditional expectations exist. We denote by $\xrightarrow{\mathcal{L}}$ convergence in distribution. The symbol $\perp\!\!\!\perp$ denotes independence of random variables. We denote by $\mathbb{1}_n$ the $n \times n$ identity matrix and omit the subscript n if we do not want to emphasize the dimension. We denote the d-variate Gaussian distribution by $\mathcal{N}_d$.

2 | MODEL FORMULATION AND THE PLUG-IN MACHINE LEARNING ESTIMATOR

We consider repeated measurements data that is grouped according to experimental units or subjects. This grouping structure introduces dependency in the data. The individual experimental units or groups are assumed to be independent, but there may be some between-group heterogeneity and within-group correlation. We consider the partially linear mixed-effects model

$$\mathbf{Y}_i = \mathbf{X}_i\beta_0 + g(\mathbf{W}_i) + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad i \in [N], \tag{2}$$

for groups i as in (1) to model the between-group heterogeneity and within-group correlation with random effects. We have n_i observations per group that are concatenated row-wise into $\mathbf{Y}_i \in \mathbb{R}^{n_i}$, $\mathbf{X}_i \in \mathbb{R}^{n_i\times d}$, and $\mathbf{W}_i \in \mathbb{R}^{n_i\times v}$. The nonparametric variables may be high-dimensional, but d is fixed. Both $\mathbf{X}_i$ and $\mathbf{W}_i$ are random. The $\mathbf{X}_i$ and $\mathbf{W}_i$ belonging to the same group i may be dependent. For groups $i \neq j$, we assume $\mathbf{X}_i \perp\!\!\!\perp \mathbf{X}_j$, $\mathbf{W}_i \perp\!\!\!\perp \mathbf{W}_j$, and $\mathbf{X}_i \perp\!\!\!\perp \mathbf{W}_j$. Moreover, we assume that all within-unit observations of the linear and nonlinear covariates, namely $((\mathbf{X}_i)_{t,\cdot}, (\mathbf{W}_i)_{t,\cdot})$ for all $i \in [N]$ and all $t \in [n_i]$, are independent and identically distributed. We assume that $\mathbf{Z}_i \in \mathbb{R}^{n_i\times q}$ is fixed. The random variable $\mathbf{b}_i \in \mathbb{R}^q$ denotes a group-specific vector of random regression coefficients that is assumed to follow a Gaussian distribution. The dimension q of the random effects model is fixed. Also the error terms are assumed to follow a Gaussian distribution as is commonly done in a mixed-effects models framework (Pinheiro & Bates, 2000). All groups i share the common linear coefficient β_0 and the potentially complex function $g : \mathbb{R}^v \to \mathbb{R}$. The function g is applied row-wise to $\mathbf{W}_i$, denoted by $g(\mathbf{W}_i)$.

We denote the total number of observations by $N_T := \sum_{i=1}^N n_i$. We assume that the numbers n_i of within-group observations are uniformly upper bounded by $n_{\max} < \infty$. Asymptotically, the number of groups, N, goes to infinity.

Our distributional and independency assumptions are summarized as follows:

Assumption 1. Consider the partially linear mixed-effects model (2). We assume that there is some $\sigma_0 > 0$ and some symmetric positive definite matrix $\Gamma_0 \in \mathbb{R}^{q\times q}$ such that the following conditions hold.

1.1 The random effects $\mathbf{b}_1,\ldots,\mathbf{b}_N$ are independent and identically distributed $\mathcal{N}_q(\mathbf{0}, \Gamma_0)$.
1.2 The error terms $\boldsymbol{\varepsilon}_1,\ldots,\boldsymbol{\varepsilon}_N$ are independent and follow a Gaussian distribution, $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma_0^2\mathbb{1}_{n_i})$ for $i \in [N]$, with the common variance component σ_0^2.
1.3 The variables $\mathbf{b}_1,\ldots,\mathbf{b}_N, \boldsymbol{\varepsilon}_1,\ldots,\boldsymbol{\varepsilon}_N$ are independent.
1.4 For all $i,j \in [N]$, $i \neq j$, we have $(\mathbf{b}_i, \boldsymbol{\varepsilon}_i) \perp\!\!\!\perp (\mathbf{W}_i, \mathbf{X}_i)$ and $(\mathbf{b}_i, \boldsymbol{\varepsilon}_i) \perp\!\!\!\perp (\mathbf{W}_j, \mathbf{X}_j)$.
1.5 For all $i \in [N]$ and all $t \in [n_i]$, we have that $((\mathbf{X}_i)_{t,\cdot}, (\mathbf{W}_i)_{t,\cdot})$ are independent and identically distributed.

We would like to remark that the distribution of the error terms $\boldsymbol{\varepsilon}_i$ in Assumption 1.2 can be generalized to $\boldsymbol{\varepsilon}_i \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma_0^2\Lambda_i(\lambda))$, where $\Lambda_i(\lambda) \in \mathbb{R}^{n_i\times n_i}$ is a symmetric positive definite matrix parametrized by some finite-dimensional parameter vector λ that all groups have in common. For the sake of notational simplicity, we restrict ourselves to Assumption 1.2.

Moreover, we may consider stochastic random effects matrices $\mathbf{Z}_i$. Alternatively, the nonparametric variables $\mathbf{W}_i$ may be part of the random effects matrix. In this case, we consider the random effects matrix $\widetilde{\mathbf{Z}}_i = \zeta(\mathbf{Z}_i, \mathbf{W}_i)$ for some known function ζ in (2) instead of $\mathbf{Z}_i$. Please see Section D in the Data S1 for further details. For simplicity, we restrict ourselves to fixed random effects matrices $\mathbf{Z}_i$ that are disjoint from $\mathbf{W}_i$.

The unknown parameters in our model are β_0, Γ_0, and σ_0. Our aim is to estimate β_0 and make inference for it. Although the variance parameters Γ_0 and σ_0 need to be estimated consistently to construct an estimator of β_0, it is not our goal to perform inference for them.

2.1 The plug-in machine learning estimator

Subsequently, we describe our plug-in machine learning estimator of β_0 in (2). To motivate our procedure, we first consider the population version with the residual terms

$$\mathbf{R}_{\mathbf{X}_i} := \mathbf{X}_i - \mathbb{E}[\mathbf{X}_i|\mathbf{W}_i] \quad \text{and} \quad \mathbf{R}_{\mathbf{Y}_i} := \mathbf{Y}_i - \mathbb{E}[\mathbf{Y}_i|\mathbf{W}_i] \quad \text{for} \quad i \in [N],$$

that adjust $\mathbf{X}_i$ and $\mathbf{Y}_i$ for $\mathbf{W}_i$. On this adjusted level, we have the linear mixed-effects model

$$\mathbf{R}_{\mathbf{Y}_i} = \mathbf{R}_{\mathbf{X}_i}\beta_0 + \mathbf{Z}_i\mathbf{b}_i + \varepsilon_i, \quad i \in [N], \tag{3}$$

due to (2) and Assumption 1.4. In particular, the adjusted and grouped responses in this model are independent in the sense that we have $\mathbf{R}_{\mathbf{Y}_i} \perp\!\!\!\perp \mathbf{R}_{\mathbf{Y}_j}$ for $i \neq j$. The strategy now is to first estimate the residuals with machine learning algorithms and then use linear mixed model techniques to infer β_0. This is done with sample splitting and cross-fitting, and the details are described next.

Let us define $\Sigma_0 := \sigma_0^{-2}\Gamma_0$ and $\mathbf{V}_{0,i} := (\mathbf{Z}_i\Sigma_0\mathbf{Z}_i^T + \mathbb{1}_{n_i})$ so that we have

$$(\mathbf{R}_{\mathbf{Y}_i}|\mathbf{W}_i, \mathbf{X}_i) \sim \mathcal{N}_{n_i}(\mathbf{R}_{\mathbf{X}_i}\beta_0, \sigma_0^2\mathbf{V}_{0,i}). \tag{4}$$

We assume that there exist functions $m_X^0 : \mathbb{R}^v \to \mathbb{R}^d$ and $m_Y^0 : \mathbb{R}^v \to \mathbb{R}$ that we can apply row-wise to $\mathbf{W}_i$ to have $\mathbb{E}[\mathbf{X}_i|\mathbf{W}_i] = m_X^0(\mathbf{W}_i)$ and $\mathbb{E}[\mathbf{Y}_i|\mathbf{W}_i] = m_Y^0(\mathbf{W}_i)$, which is conceivable due to Assumption 1.5. In particular, m_X^0 and m_Y^0 do not depend on the grouping index i. Let $\eta^0 := (m_X^0, m_Y^0)$ denote the true unknown nuisance parameter. Let us denote by $\theta_0 := (\beta_0, \sigma_0^2, \Sigma_0)$ the complete true unknown parameter vector and by $\theta := (\beta, \sigma^2, \Sigma)$ and $\mathbf{V}_i := \mathbf{Z}_i\Sigma\mathbf{Z}_i^T + \mathbb{1}_{n_i}$ respective general parameters. The log-likelihood of group i is given by

$$\begin{aligned}\ell_i(\theta, \eta^0) = &-\frac{n_i}{2}\log(2\pi) - \frac{n_i}{2}\log(\sigma^2) - \frac{1}{2}\log\left(\det(\mathbf{V}_i)\right) \\ &- \frac{1}{2\sigma^2}(\mathbf{R}_{\mathbf{Y}_i} - \mathbf{R}_{\mathbf{X}_i}\beta)^T\mathbf{V}_i^{-1}(\mathbf{R}_{\mathbf{Y}_i} - \mathbf{R}_{\mathbf{X}_i}\beta) - \log\left(p(\mathbf{W}_i, \mathbf{X}_i)\right),\end{aligned} \tag{5}$$

where $p(\mathbf{W}_i, \mathbf{X}_i)$ denotes the joint density of $\mathbf{W}_i$ and $\mathbf{X}_i$. We assume that $p(\mathbf{W}_i, \mathbf{X}_i)$ does not depend on θ. The true nuisance parameter η^0 in the log-likelihood (5) is unknown and estimated with machine learning algorithms (see below). Denote by $\eta := (m_X, m_Y)$ some general nuisance parameter. The terms that adjust $\mathbf{X}_i$ and $\mathbf{Y}_i$ for $\mathbf{W}_i$ with this general nuisance parameter are given by $\mathbf{X}_i - m_X(\mathbf{W}_i)$ and $\mathbf{Y}_i - m_Y(\mathbf{W}_i)$. Up to additive constants that do not depend on θ and η, we thus consider maximum likelihood estimation with the likelihood

$$\begin{aligned}\ell_i(\theta, \eta) = &-\frac{n_i}{2}\log(\sigma^2) - \frac{1}{2}\log\left(\det(\mathbf{V}_i)\right) \\ &- \frac{1}{2\sigma^2}(\mathbf{Y}_i - m_Y(\mathbf{W}_i) - (\mathbf{X}_i - m_X(\mathbf{W}_i))\,\beta)^T\mathbf{V}_i^{-1}\left(\mathbf{Y}_i - m_Y(\mathbf{W}_i) - (\mathbf{X}_i - m_X(\mathbf{W}_i))\,\beta\right),\end{aligned}$$

which is a function of both the finite-dimensional parameter θ and the infinite-dimensional nuisance parameter η.

Our estimator of β_0 is constructed as follows adapting double machine learning techniques. We estimate η^0 with machine learning algorithms and plug these estimators into the estimating equations for θ_0, Equation (6) below, to obtain an estimator for β_0. This procedure is done with sample splitting and cross-fitting as explained next.

Consider repeated measurements from N experimental units, subjects, or groups as in (2). Denote by $\mathbf{S}_i := (\mathbf{W}_i, \mathbf{X}_i, \mathbf{Z}_i, \mathbf{Y}_i)$ the observations of group i. First, we split the group indices $[N]$ into $K \geq 2$ disjoint sets $I_1, \ldots, I_K$ of approximately equal size in the sense that the number of unit-level observations belonging to each set are asymptotically of the same order. The number of observations per unit may differ, but is assumed to be uniformly bounded. That is, we avoid too unbalanced settings. Please see Section B in the Data S1 for further details.

For each $k \in [K]$, we estimate the conditional expectations $m_X^0(W)$ and $m_Y^0(W)$ with data from I_k^c. We call the resulting estimators $\hat{m}_X^{I_k^c}$ and $\hat{m}_Y^{I_k^c}$, respectively. Then, the adjustments $\widehat{\mathbf{R}}_{\mathbf{X}_i}^{I_k} := \mathbf{X}_i - \hat{m}_X^{I_k^c}(\mathbf{W}_i)$, and $\widehat{\mathbf{R}}_{\mathbf{Y}_i}^{I_k} := \mathbf{Y}_i - \hat{m}_Y^{I_k^c}(\mathbf{W}_i)$ for $i \in I_k$ are evaluated on I_k, the complement of I_k^c. Let $\hat{\eta}^{I_k^c} := (\hat{m}_X^{I_k^c}, \hat{m}_Y^{I_k^c})$ denote the estimated nuisance parameter. Consider the score function $\psi(\mathbf{S}_i; \theta, \eta) := \nabla_\theta \ell_i(\theta, \eta)$, where ∇_θ denotes the gradient with respect to θ interpreted as a vector. On each set I_k, we consider an estimator $\hat{\theta}_k = (\hat{\beta}_k, \hat{\sigma}_k^2, \hat{\Sigma}_k)$ of θ_0 that, approximately, in the sense of Assumption 4.3 in the Data S1, solves

$$\frac{1}{n_{T,k}} \sum_{i \in I_k} \psi(\mathbf{S}_i; \hat{\theta}_k, \hat{\eta}^{I_k^c}) = \frac{1}{n_{T,k}} \sum_{i \in I_k} \nabla_\theta \ell_i(\theta, \eta) \stackrel{!}{=} \mathbf{0}, \tag{6}$$

where $n_{T,k} := \sum_{i \in I_k} n_i$ denotes the total number of observations from experimental units that belong to the set I_k. These K estimators $\hat{\theta}_k$ for $k \in [K]$ are assembled to form the final cross-fitting estimator

$$\hat{\beta} := \frac{1}{K} \sum_{k=1}^K \hat{\beta}_k, \tag{7}$$

of β_0. We remark that one can simply use linear mixed model computation and software to compute $\hat{\beta}_k$ based on the estimated residuals $\widehat{\mathbf{R}}^{I_k}$. The estimator $\hat{\beta}$ fundamentally depends on the particular sample split. To alleviate this effect, the overall procedure may be repeated S times (Chernozhukov et al., 2018). The S point estimators are aggregated by the median, and an additional term accounting for the random splits is added to the variance estimator of $\hat{\beta}$; please see Algorithm 1 that presents the complete procedure.

2.2 | Theoretical properties of the plug-in machine learning estimator

The estimator $\hat{\beta}$ as in (7) converges at the parametric rate, is asymptotically Gaussian distributed, and semiparametrically efficient.

Theorem 1. *Consider grouped observations $\{\mathbf{S}_i = (\mathbf{W}_i, \mathbf{X}_i, \mathbf{Y}_i)\}_{i \in [N]}$ from the partially linear mixed-effects model (2) that satisfy Assumption 1 such that $p(\mathbf{W}_i, \mathbf{X}_i)$ does not depend on θ. Let $N_T := \sum_{i=1}^N n_i$ denote the total number of unit-level observations. Furthermore, suppose the assumptions in Section B in the Data S1 hold, and consider the symmetric positive-definite matrix T_0 given in Assumption 3.8 in the Data S1. Then, $\hat{\beta}$ as in (7) concentrates in a $1/\sqrt{N_T}$ neighborhood of β_0, is centered Gaussian, namely*

$$\sqrt{N_T} T_0^{\frac{1}{2}} (\hat{\beta} - \beta_0) \xrightarrow{\mathcal{L}} \mathcal{N}_d(\mathbf{0}, \mathbb{1}_d) \quad (N \to \infty), \tag{8}$$

and semiparametrically efficient. The convergence in (8) is in fact uniformly over the law P of $\{\mathbf{S}_i = (\mathbf{W}_i, \mathbf{X}_i, \mathbf{Y}_i)\}_{i \in [N]}$.

Algorithm 1: Plug-in machine learning for partially linear mixed-effects models with repeated measurements

Input : N grouped observations $\{\mathbf{S}_i = (\mathbf{W}_i, \mathbf{X}_i, \mathbf{Z}_i, \mathbf{Y}_i)\}_{i\in[N]}$ from model (2) satisfying Assumption 1, a natural number K, a natural number $\mathcal{S}$.

Output: An estimator of β_0 in (2) together with its estimated asymptotic variance.

for $s \in [\mathcal{S}]$ **do**
- Split the grouped observation index set $[N]$ into K sets $I_1, \dots, I_K$ of approximately equal size.
- **for** $k \in K$ **do**
  - Compute the conditional expectation estimators $\hat{m}_X^{I_k^c}$ and $\hat{m}_Y^{I_k^c}$ with some machine learning algorithm and data from I_k^c.
  - Evaluate the adjustments $\widehat{\mathbf{R}}_{\mathbf{X}_i}^{I_k} = \mathbf{X}_i - \hat{m}_X^{I_k^c}(\mathbf{W}_i)$ and $\widehat{\mathbf{R}}_{\mathbf{Y}_i}^{I_k} = \mathbf{Y}_i - \hat{m}_Y^{I_k^c}(\mathbf{W}_i)$ for $i \in I_k$.
  - Compute $\hat{\theta}_{k,s} = (\hat{\beta}_{k,s}, \hat{\sigma}^2_{k,s}, \hat{\Sigma}_{k,s})$ using, for instance, linear mixed model techniques.
- **end**
- Compute $\hat{\beta}_s = \frac{1}{K}\sum_{k=1}^K \hat{\beta}_{k,s}$ as an approximate solution to (6).
- Compute an estimate $\hat{T}_{0,s}$ of the asymptotic variance-covariance matrix T_0 in Theorem 1.

end

Compute $\hat{\beta} = \text{median}_{s\in[\mathcal{S}]}(\hat{\beta}_s)$.

Estimate T_0 by $\hat{T}_0 = \text{median}_{s\in[\mathcal{S}]}(\hat{T}_{0,s} + (\hat{\beta} - \hat{\beta}_s)(\hat{\beta} - \hat{\beta}_s)^T)$.

Please see Section C.4 in the Data S1 for a proof of Theorem 1. Our proof builds on Chernozhukov et al. (2018), but we have to take into account the correlation within units that is introduced by the random effects.

The inverse asymptotic variance-covariance matrix T_0 can be consistently estimated; see Lemma 23 in the Data S1. Semiparametric efficiency follows from (Lin & Carroll, 2001, section 5).

The assumptions in Section B of the Data S1 specify regularity conditions and required convergence rates of the machine learning estimators. The machine learning errors need to satisfy the product relationship

$$\|m_X^0(W) - \hat{m}_X^{I_k^c}(W)\|_{P,2}(\|m_Y^0(W) - \hat{m}_Y^{I_k^c}(W)\|_{P,2} + \|m_X^0(W) - \hat{m}_X^{I_k^c}(W)\|_{P,2}) \ll N^{-\frac{1}{2}}.$$

This bound requires that only the products of the machine learning estimation errors $\|m_X^0(W) - \hat{m}_X^{I_k^c}(W)\|_{P,2}$ and $\|m_Y^0(W) - \hat{m}_Y^{I_k^c}(W)\|_{P,2}$ but not the individual ones need to vanish at a rate smaller than $N^{-1/2}$. In particular, the individual estimation errors may vanish at the rate smaller than $N^{-1/4}$. This is achieved by many machine learning methods (cf. Chernozhukov et al., 2018): ℓ_1-penalized and related methods in a variety of sparse models (Belloni et al., 2011; Belloni et al., 2012; Belloni & Chernozhukov, 2011, 2013; Bickel et al., 2009; Bühlmann & van de Geer, 2011), forward selection in sparse models (Kozbur, 2020), L_2-boosting in sparse linear models (Luo & Spindler, 2016), a class of regression trees and random forests (Wager & Walther, 2016), and neural networks (Chen & White, 1999).

We note that so-called Neyman orthogonality makes score functions insensitive to inserting potentially biased machine learning estimators of the nuisance parameters. A score function is Neyman orthogonal if its Gateaux derivative vanishes at the true θ_0 and the true η^0. In particular, Neyman orthogonality is a first-order property. The product relationship of the machine learning estimating errors described above is used to bound second-order terms. We refer to Section C.4 in the Data S1 for more details.

3 | NUMERICAL EXPERIMENTS

Subsequently, we apply our plug-in machine learning method to an empirical and a pseudo-random dataset and in a simulation study. Our implementation is available in the R-package `dmlalg` (Emmenegger, 2021).

3.1 | Empirical analysis: CD4 cell count data

First, we apply our method to longitudinal CD4 cell counts data collected from human immunodeficiency virus (HIV) seroconverters. This data has previously been analyzed by Zeger and Diggle (1994) and is available in the R-package `jmcm` (Pan & Pan, 2017) as `aids`. It contains 2376 observations of CD4 cell counts measured on 369 subjects. The data was collected during a period ranging from 3 years before to 6 years after seroconversion. The number of observations per subject ranges from 1 to 12, but for most subjects, 4–10 observations are available. Please see Zeger and Diggle (1994) for more details on this dataset.

Apart from time, five other covariates are measured: the age at seroconversion in years (age), the smoking status measured by the number of cigarette packs consumed per day (smoking), a binary variable indicating drug use (drugs), the number of sex partners (sex), and the depression status measured on the Center for Epidemiologic Studies Depression (CESD) scale (cesd), where higher CESD values indicate the presence of more depression symptoms.

We incorporate a random intercept per person. Furthermore, we consider a square-root transformation of the CD4 cell counts to reduce the skewness of this variable as proposed by Zeger and Diggle (1994). The CD4 counts are our response. The covariates that are of scientific interest are considered as X's, and the remaining covariates are considered as W's in the partially linear mixed-effects model (2). The effect of time is modeled nonparametrically, but there are several options to model the other covariates. Other models than partially linear mixed-effects model have also been considered in the literature to analyze this dataset. For instance, Fan and Zhang (2000) consider a functional linear model where the linear coefficients are a function of the time.

We consider two partially linear mixed-effects models for this dataset. First, we incorporate all covariates except time linearly. Most approaches in the literature employing a partially linear mixed-effects model for this data that model time nonparametrically report that sex and cesd are significant and that either smoking or drugs is significant as well; see for instance Zeger and Diggle (1994), Taavoni and Arashi (2021b), and Wang et al. (2005). Guoyou and Zhongyi (2008) develop a robust estimation method for longitudinal data and estimate nonlinear effects from time with regression splines. With the CD4 dataset, they find that smoking and cesd are significant.

We apply our method with $K = 2$ sample splits, $\mathcal{S} = 100$ repetitions of splitting the data, and learn the conditional expectations with random forests that consist of 500 trees whose minimal node size is 5. Like Guoyou and Zhongyi (2008), we conclude that smoking and cesd are significant; please see the first row of Table 1 for a more precise account of our findings. Apart from sex, our point estimators are larger or of about the same size in absolute value as what Guoyou and

TABLE 1 Estimates of the linear coefficient and its SD in parentheses with our method for nonparametrically adjusting for time (first row) and for time, age, and sex (second row).

	Age	Smoking	Drugs	Sex	cesd
W = (time)	0.004 (0.027)	0.752 (0.123)	0.704 (0.360)	0.001 (0.043)	−0.042 (0.015)
W = (time, age, sex)	—	0.620 (0.126)	0.602 (0.335)	—	−0.047 (0.015)
Zeger and Diggle (1994)	0.037 (0.18)	0.27 (0.15)	0.37 (0.31)	0.10 (0.038)	−0.058 (0.015)
Taavoni and Arashi (2021b)	$1.5 \cdot 10^{-17}$ ($3.5 \cdot 10^{-17}$)	0.152 (0.208)	0.130 (0.071)	0.0184 (0.0039)	−0.0141 (0.0061)
Wang et al. (2005)	0.010 (0.033)	0.549 (0.144)	0.584 (0.331)	0.080 (0.038)	−0.045 (0.013)
Guoyou and Zhongyi (2008)	0.006 (0.038)	0.538 (0.136)	0.637 (0.350)	0.066 (0.040)	−0.042 (0.015)

Notes: The remaining rows display the results from (Zeger & Diggle, 1994, section 5), Taavoni and Arashi (2021b, table 1, "Kernel"), (Wang et al., 2005, table 2, "Semiparametric efficient scenario I"), and (Guoyou & Zhongyi, 2008, table 5, "Robust"), respectively.

Zhongyi (2008) obtain. However, apart from age, the SDs are slightly larger with our method. This can be expected because random forests are more complex than the regression splines Guoyou and Zhongyi (2008) employ.

We consider a second estimation approach where we model the variables time, age, and sex nonparametrically and allow them to interact. It is conceivable that these variables are not (causally) influenced by smoking, drugs, and cesd and that they are therefore exogenous. The variables smoking, drugs, and cesd are modeled linearly, and they are considered as treatment variables. Some direct causal effect interpretations are possible if one is willing to assume, for instance, that the nonparametric adjustment variables are causal parents of the linear variables or the response. However, we do not pursue this line of thought further. We estimate the conditional expectations given the three nonparametric variables time, age, and sex again with random forests that consist of 500 trees whose minimal node size is 5 and use $K = 2$ and $\mathcal{S} = 100$ in Algorithm 1. We again find that smoking and cesd are significant; please see the second row of Table 1. This cannot be expected a priori because this second model incorporates more complex adjustments, which can lead to less significant variables.

3.2 | Pseudorandom simulation study: CD4 cell count data

Subsequently, we consider the CD4 cell count data from the previous subsection and perform a pseudorandom simulation study. The variables smoking, drugs, and cesd are modeled linearly and the variables time, age, and sex nonparametrically. We condition on these six variables in our simulation. That is, they are the same in all repetitions. The function g in (2) is chosen as a regression tree that we built beforehand. We let $\beta_0 = (0.62, 0.6, -0.05)^T$, where the first component corresponds to smoking, the second one to drugs, and the last one to cesd, consider a standard deviation of the random intercept per subject of 4.36, and a SD of the error term of 4.35. These are the point estimates of the respective quantities obtained in the previous subsection.

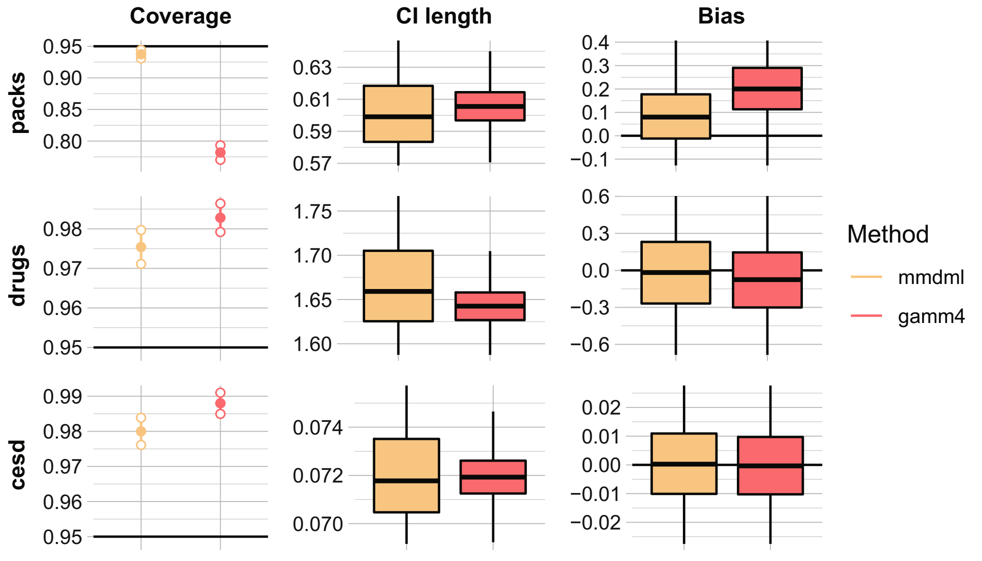

FIGURE 1 Coverage and length of two-sided confidence intervals at significance level 5% and bias for our method, `mmdml`, and `gamm4`. In the coverage plot, solid dots represent point estimators, and circles represent 95% confidence bands with respect to the 5000 simulation runs. The confidence interval length and bias are displayed with box plots without outliers.

Our fitting procedure uses random forests consisting of 500 trees whose minimal node size is 5 to estimate the conditional expectations, and we use $K = 2$ and $S = 10$ in Algorithm 1. We perform 5000 simulation runs. We compare the performance of our method with that of the spline-based function `gamm4` from the package `gamm4` (Wood & Scheipl, 2020) for the statistical software R (R Core Team, 2021). This method represents the nonlinear part of the model by smooth additive functions and estimates them by penalized regression splines. The penalized components are treated as random effects and the unpenalized components as fixed effects.

The results are displayed in Figure 1. With our method, `mmdml`, the two-sided confidence intervals for β_0 are of about the same length but achieve a coverage that is closer to the nominal 95% confidence level than with `gamm4`. The `gamm4` method largely undercovers the packs component of β_0, which can be explained by the incorporated bias.

3.3 | Simulation study

Finally, we carry out a simulation study with a partially linear mixed-effects model with $q = 3$ random effects and where β_0 is one-dimensional. Every subject has their own random intercept term and a nested random effect with two levels. Thus, the random effects structure is more complex than in the previous two subsections because these models only used a random intercept. We compare three data generating mechanisms: one where the function g is nonsmooth and the number of observations per group is balanced, one where the function g is smooth and the number of observations per group is balanced, and one where the function g is nonsmooth and the number of observations per group is unbalanced; please see Section A in the Data S1 for more details.

We estimate the nonparametric nuisance components, that is, the conditional expectations, with random forests consisting of 500 trees whose minimal node size is 5. Furthermore, we use

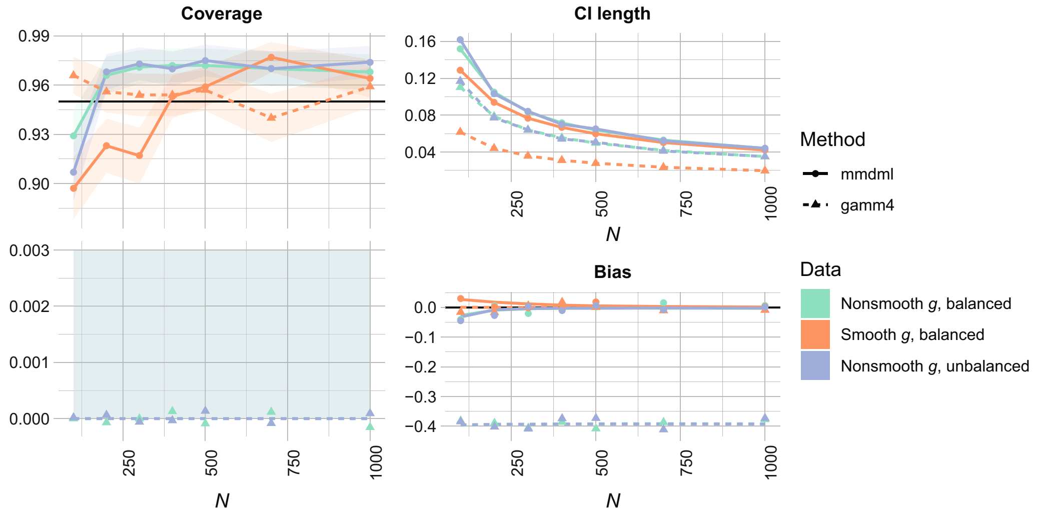

FIGURE 2 Coverage and median length of two-sided confidence intervals for β_0 at significance level 5% (true $\beta_0 = 0.5$) and median bias for three data generating scenarios for our method, mmdml, and gamm4. The shaded regions in the coverage plot represent 95% confidence bands with respect to the 1000 simulation runs. The dots in the coverage and bias plot are jittered, but neither are their interconnecting lines nor their confidence bands.

$K = 2$ and $S = 10$ in Algorithm 1. We perform 1000 simulation runs and consider different numbers of groups N. As in the previous subsection, we compare the performance of our method with gamm4.

The results are displayed in Figure 2. Our method, mmdml, highly outperforms gamm4 in terms of coverage for nonsmooth g because the coverage of gamm4 equals 0 due to its substantial bias. Our method overcovers slightly due to the correction factor that results from the S repetitions. However, this correction factor is highly recommended in practice. With smooth g, gamm4 is closer to the nominal coverage and has shorter confidence intervals than our method. Because the underlying model is smooth and additive, a spline-based estimator is better suited. In all scenarios, our method outputs longer confidence intervals than gamm4 because we use random forests; consistent with theory, the difference in absolute value decreases though when N increases.

4 | CONCLUSION

Our aim was to develop inference for the linear coefficient β_0 of a partially linear mixed-effects model that includes a linear term and potentially complex nonparametric terms. Such models can be used to describe heterogeneous and correlated data that feature some grouping structure, which may result from taking repeated measurements. Traditionally, spline or kernel approaches are used to cope with the nonparametric part of such a model. We presented a plug-in machine learning scheme that adapts double machine learning techniques of Chernozhukov et al. (2018) to estimate any nonparametric components with arbitrary machine learning algorithms. This allowed us to consider complex nonparametric components with interaction structures and high-dimensional variables.

Our proposed method is as follows. First, the nonparametric variables are regressed out from the response and the linear variables. This step adjusts the response and the linear variables for the nonparametric variables and may be performed with any machine learning algorithm. The adjusted variables satisfy a linear mixed-effects model, where the linear coefficient β_0 can be estimated with standard linear mixed-effects techniques. We showed that the estimator of β_0 asymptotically follows a Gaussian distribution, converges at the parametric rate, and is semiparametrically efficient. This asymptotic result allows us to perform inference for β_0.

Empirical experiments demonstrated the performance of our proposed method. We conducted an empirical and pseudorandom data analysis and a simulation study. The simulation study and the pseudorandom experiment confirmed the effectiveness of our method in terms of coverage, length of confidence intervals, and estimation bias compared to a penalized regression spline approach relying on additive models. In the empirical experiment, we analyzed longitudinal CD4 cell counts data collected from HIV-infected individuals. In the literature, most methods only incorporate the time component nonparametrically to analyze this dataset. Because we estimate nonparametric components with machine learning algorithms, we can allow several variables to enter the model nonlinearly, and we can allow these variables to interact.

Implementations of our method are available in the R-package `dmlalg` (Emmenegger, 2021).

ACKNOWLEDGMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 786461). We would like to thank Michael Law for constructive comments. Open access funding provided by Eidgenössische Technische Hochschule Zrich.

ORCID

Corinne Emmenegger https://orcid.org/0000-0003-0353-8888

REFERENCES

Aniley, T. T., Debusho, L. K., Nigusie, Z. M., Yimer, W. K., & Yimer, B. B. (2019). A semi-parametric mixed models for longitudinally measured fasting blood sugar level of adult diabetic patients. *BMC Medical Research Methodology*, *19*, 13.

Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, *80*, 2369–2429.

Belloni, A., & Chernozhukov, V. (2011). l1-penalized quantile regression in high-dimensional sparse models. *The Annals of Statistics*, *39*, 82–130.

Belloni, A., & Chernozhukov, V. (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli*, *19*, 521–547.

Belloni, A., Chernozhukov, V., & Wang, L. (2011). Square-root lasso: Pivotal recovery of sparse signals via conic programming. *Biometrika*, *98*, 791–806.

Bickel, P. J., Ritov, Y., & Tsybakov, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *The Annals of Statistics*, *37*, 1705–1732.

Bondell, H. D., Krishna, A., & Ghosh, S. K. (2010). Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics*, *66*, 1069–1077.

Bradic, J., Claeskens, G., & Gueuning, T. (2020). Fixed effects testing in high-dimensional linear mixed models. *Journal of the American Statistical Association*, *115*, 1835–1850.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Bühlmann, P., & van de Geer, S. (2011). *Statistics for high-dimensional data: Methods, theory and applications Springer Series in Statistics*. Springer.

Chen, L., & Cao, H. (2017). Analysis of asynchronous longitudinal data with partially linear models. *Electronic Journal of Statistics*, *11*, 1549–1569.

Chen, X., & White, H. (1999). Improved rates and asymptotic normality for nonparametric neural network estimators. *IEEE Transactions on Information Theory*, *45*, 682–691.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, *21*, C1–C68.

Davidian, M., & Giltinan, D. M. (1995). *Nonlinear models for repeated measurement data, vol. 62 of monographs on statistics and applied probability*. Chapman & Hall/CRC.

Davidian, M., & Giltinan, D. M. (2003). Nonlinear models for repeated measurement data: An overview and update. *Journal of Agricultural, Biological, and Environmental Statistics*, *8*, 387–419.

Davis, C. S. (2002). *Statistical methods for the analysis of repeated measurements. Springer texts in statistics*. Springer.

Demidenko, E. (2004). *Mixed models: Theory and applications Wiley Series in Probability and Statistics*. John Wiley & Sons, Ltd.

Emmenegger, C. (2021). *dmlalg: Double machine learning algorithms*. R-package available on CRAN. Retrieved from. https://cran.r-project.org/web/packages/dmlalg/index.html

Fahrmeir, L., & Kneib, T. (2011). *Bayesian smoothing and regression for longitudinal, spatial and event history data Oxford statistical science series* (Vol. *36*). Oxford University Press.

Fan, J., & Zhang, J.-T. (2000). Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *62*, 303–322.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied longitudinal analysis Wiley Series in Probability and Statistics* (2nd ed.). Wiley.

Guoyou, Q., & Zhongyi, Z. (2008). Robust estimation in partial linear mixed model for longitudinal data. *Acta Mathematica Scientia*, *28*, 333–347.

Hart, J. D., & Wehrly, T. E. (1986). Kernel regression estimation using repeated measurements data. *Journal of the American Statistical Association*, *81*, 1080–1088.

Ibrahim, J. G., Zhu, H., Garcia, R. I., & Guo, R. (2011). Fixed and random effects selection in mixed effects models. *Biometrics*, *67*, 495–503.

Kim, S., Zeng, D., & Taylor, J. M. G. (2017). Joint partially linear model for longitudinal data with informative drop-outs. *Biometrics*, *73*, 72–82.

Kozbur, D. (2020). Analysis of testing-based forward model selection. *Econometrica*, *88*, 2147–2173.

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*, 963–974.

Li, S., Cai, T. T., & Li, H. (2021). Inference for high-dimensional linear mixed-effects models: A quasi-likelihood approach. *Journal of the American Statistical Association*. *117*, 1835–1846.

Li, Z., & Zhu, L. (2010). On variance components in semiparametric mixed models for longitudinal data. *Scandinavian Journal of Statistics*, *37*, 442–457.

Liang, H. (2009). Generalized partially linear mixed-effects models incorporating mismeasured covariates. *Annals of the Institute of Statistical Mathematics*, *61*, 27–46.

Lin, H., Qin, G., Zhang, J., & Fung, W. K. (2018). Doubly robust estimation of partially linear models for longitudinal data with dropouts and measurement error in covariates. *Statistics*, *52*, 84–98.

Lin, X., & Carroll, R. (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association*, *96*, 1045–1056.

Lu, T. (2016). Skew-t partially linear mixed-effects models for AIDS clinical studies. *Journal of Biopharmaceutical Statistics*, *26*, 899–911.

Luo, Y., & Spindler, M. (2016). High-dimensional $l_2$boosting: Rate of convergence. Preprint *arXiv:1602.08927*.

Mammen, E., & van de Geer, S. (1997). Penalized quasi-likelihood estimation in partial linear models. *The Annals of Statistics*, *25*, 1014–1035.

Masci, C., Paganoni, A. M., & Ieva, F. (2019). Semiparametric mixed effects models for unsupervised classification of Italian schools. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *182*, 1313–1342.

Ohinata, R. (2012). *Three essays on application of semiparametric regression: Partially linear mixed effects model and index model* (PhD thesis). Wirtschaftswissenschaftlichen Fakultät der Universität Göttingen.

Pan, J., & Pan, Y. (2017). Jmcm: An R package for joint mean-covariance modeling of longitudinal data. *Journal of Statistical Software*, *82*, 1–29.

Pinheiro, J. C. (1994). *Topics in mixed effects models* (PhD thesis), University of Wisconsin.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS. Statistics and computing*. Springer.
Qin, G., & Zhu, Z. (2007). Robust estimation in generalized semiparametric mixed models for longitudinal data. *Journal of Multivariate Analysis*, *98*, 1658–1683.
Qin, G. Y. Q., & Zhu, Z. Y. (2009). Robustified maximum likelihood estimation in generalized partial linear mixed model for longitudinal data. *Biometrics*, *65*, 52–59.
R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/
Rice, J. A., & Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society Series B (Methodological)*, *53*, 233–243.
Schelldorfer, J., Bühlmann, P., & van de Geer, S. (2011). Estimation for high-dimensional linear mixed-effects models using ℓ_1-penalization. *Scandinavian Journal of Statistics*, *38*, 197–214.
Taavoni, M., & Arashi, M. (2021a). High-dimensional generalized semiparametric model for longitudinal data. *Statistics*, *55*, 831–850.
Taavoni, M., & Arashi, M. (2021b). Kernel estimation in semiparametric mixed effect longitudinal modeling. *Statistical Papers*, *62*, 1095–1116.
Taavoni, M., Arashi, M., Wang, W.-L., & Lin, T.-I. (2021). Multivariate *t* semiparametric mixed-effects model for longitudinal data with multiple characteristics. *Journal of Statistical Computation and Simulation*, *91*, 260–281.
Tang, Y., Sinha, D., & Pati, D. (2015). Bayesian partial linear model for skewed longitudinal data. *Biostatistics*, *16*, 441–453.
Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B: Methodological*, *58*, 267–288.
Verbeke, G., & Molenberghs, G. (2000). *Linear mixed models for longitudinal data. Springer series in statistics*. Springer.
Vonesh, E. F., & Chinchilli, V. M. (1997). *Linear and nonlinear models for the analysis of repeated measurements Statistics: Textbooks and monographs* (Vol. *154*). Chapman & Hall/CRC.
Wager, S., & Walther, G. (2016). Adaptive concentration of regression trees, with application to random forests. Preprint *arXiv:1503.06388*.
Wang, N., Carroll, R. J., & Lin, X. (2005). Efficient semiparametric marginal estimation for longitudinal/clustered data. *Journal of the American Statistical Association*, *100*, 147–157.
Wood, S., & Scheipl, F. (2020). *gamm4: Generalized Additive Mixed Models using "mgcv" and "lme4"*. R package version 0.2–6. Retrieved from. https://CRAN.R-project.org/package=gamm4
Zeger, S. L., & Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics*, *50*, 689–699.
Zhang, D. (2004). Generalized linear mixed models with varying coefficients for longitudinal data. *Biometrics*, *60*, 8–15.
Zhang, D., Lin, X., Raz, J., & Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *Journal of the American Statistical Association*, *93*, 710–719.
Zhang, J., & Xue, L. (2020). Variable selection for generalized partially linear models with longitudinal data. *Evolutionary Intelligence*, *15*, 2473–2483.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Emmenegger, C., & Bühlmann, P. (2023). Plug-in machine learning for partially linear mixed-effects models with repeated measurements. *Scandinavian Journal of Statistics*, *50*(4), 1553–1567. https://doi.org/10.1111/sjos.12639