A new class of Semi-Mixed Effects Models
and its Application to Tourist Expenditures in Small Areas

María José Lombardía
Universidade da Coruña
Stefan Sperlich
Georg-August Universität Göttingen

April 17, 2009

Abstract

A new class of semi-mixed effects models is introduced. It includes random or mixed and fixed effects models as extreme cases. In multi-level regression, such as small area studies, and in panel data studies, using a fixed effect for each region leads to models that are flexible but that have poor estimation accuracy; they are over-parameterized. Regarding region as a random effect reduces the number of parameters, and hence the flexibility, but needs crucial assumptions, such as that of independence between covariates and the random effects. The proposed class of models constitutes a continuum of models, indexed by a "slider", that determines the position of the model between these two extremes. So one can choose a model that is close to the parsimonious random effects case, but far enough away from it to filter out unwanted dependencies. The methodology is used for a small area analysis of tourist expenditures in Galicia.[1]

Keywords and Phrases: Semi-mixed effects models, semiparametric regression, multilevel models, small area statistics, panel data analysis.

[1]We gratefully thank Walter Zucchini, Jean Opsomer, Thomas Kneib, and Bernd Fitzenberger for helpful discussion and comments.

1 Introduction

For a response $Y_{dj} \in I\!\!R$ and covariates $\boldsymbol{X}_{dj} \in I\!\!R^p$, including the intercept, consider a generalized linear Mixed Effects Model (MEM) with known link g

$$E\left[Y_{dj}|\boldsymbol{u}_d, \boldsymbol{X}_{dj}\right] \;=\; g\left\{\boldsymbol{X}_{dj}^t\boldsymbol{\beta} + \boldsymbol{Z}_{dj}^t\boldsymbol{u}_d\right\}, \quad d = 1, \ldots, D;\; j = 1, \ldots, n_d, \qquad (1)$$

with $\boldsymbol{Z}_{dj} \subseteq \boldsymbol{X}_{dj}$ of dimension ρ, $\boldsymbol{\beta} \in I\!\!R^p$ the fixed effect, and $\boldsymbol{u}_d \in I\!\!R^\rho$ the i.i.d. unobservable random effect with mean zero and unknown variances-covariance matrix Σ_u. The latter has to be estimated. Suppose to have sample size $n = \sum_{d=1}^D n_d$, where D is the number of areas (domains or groups) with the typical assumption that $D \to \infty$ at rate $O(n)$. In panel data analysis i may be time and d the individual. A crucial assumption for the existing methodology is that $\boldsymbol{X}_{dj}$ and $\boldsymbol{u}_d$ are independent and that $g(\cdot)$ is known. Note that, if g is the identity, model (1) includes the nested-error model ($\boldsymbol{Z}_{dj} = 1$ and $u_d \in I\!\!R$), the random regression coefficient model ($\boldsymbol{Z}_{dj} = \boldsymbol{X}_{dj}$), and the Fay-Herriot model (only area specific information, Fay and Herriot, 1979); see Prasad and Rao (1990) for a summary.

Today, mixed effects models are popular in many areas of statistics, especially in small area statistics, see Jiang and Lahiri (2006) or Rao (2003) for reviews; for panel data analysis Diggle, Heagerty, Liang and Zeger (2002), and Ghosh, Nangia and Kim (1996) for a typical example. They are widely applied in biomedical, forestry, agricultural, economic and social science studies. Although the different research areas favor different terminology, like small area statistics, multi-level or simply mixed effects models, the statistical problems of modeling, estimation and testing are basically the same; the differences arise mainly in the subsequent inferences. For example, in biometrics they serve to analyze data with repeated measurements; in panel data analysis they account for possible heterogeneity over the cross sectional samples; in small area statistics they serve to improve the prediction of area-level parameters, while in econometrics they improve the calculation of macro indices from micro-data. Apart from the increasing interest in multi-level modeling (see Goldstein, 2003), they have also become popular in economics for data matching, i.e. to impute a certain factor for the individuals in the sample of interest with the aid of a different sample (see Elbers, Lanjouw and Lanjouw, 2003, for a recent example in poverty mapping). At the end, they all have in common that they try to account for a certain clustering, may it be due to space, time or individuals over time in panels, climate, administrative area or districts, villages or even large families, genetic groups or species.

More recently, mixed effects models have entered the nonparametric world; see Opsomer, Claeskens, Ranalli, Kauermann and Breidt (2008), Hamilton (2001), and Tutz (2001). In applied statistics, semiparametric Bayesian approaches are often used in combination with (penalized) splines, series or random field estimators; see Adebayo and Fahrmeir (2005) or Kneib and Fahrmeir (2006) among others. However the asymptotic theory for estimation in semiparametric mixed models was developed only recently. Lin and Carroll (2001) and Lombardía and Sperlich (2008) introduced an estimation procedure for generalized partial linear mixed effects models, specification tests with bootstrap procedures, and provided asymptotic theory for these methods.

Thus, for a more flexible modeling we may also allow some covariates to enter the model nonparametrically. To ease the notation let us call these variables $\boldsymbol{T} \in \mathbb{R}^q$ and be different from the variables $\boldsymbol{X}$ which enter the model linearly. Then we have a generalized Partial Linear mixed effects Model (PLM) of the form

$$E\left[Y_{dj}|\boldsymbol{X}_{dj},\boldsymbol{T}_{dj}\right] \;=\; g\left\{\boldsymbol{X}_{dj}^{t}\boldsymbol{\beta}+\gamma(\boldsymbol{T}_{dj}+\boldsymbol{Z}_{dj}^{t}\boldsymbol{u}_{d})\right\}, \quad d=1,\ldots,D;\; j=1,\ldots,n_d \tag{2}$$

with a nonparametric function $\gamma : I\!\!R^q \rightarrow I\!\!R$.

The MEM is often motivated by the fact that it allows for efficient estimation of the fixed part, but makes also use of the random effects for prediction. This seems to outperform other parametric models in predicting and efficient estimation. When predicting, the additional variance that results from assuming this effect to be random, is only slightly larger than the variance of a fixed effect estimate based on small samples, and this deficiency is easily compensated by the efficient estimation of $\boldsymbol{\beta}$. However, this improved prediction in the mean is illusory if the somewhat unrealistic assumption of independence between area effects and the covariates, as well as the unobserved individual effects, is not met. Thus, even when a MEM leads to a better sample fit, it does so at the cost of producing biased estimates, and consequently bad out-of-sample prediction. Furthermore, methods to do valid inference have not yet been developed. All the currently available methods for testing or prediction intervals are clearly inconsistent if the assumption of independence is violated. This deficiency is not shared by the Fixed Effects Model (FEM)

$$E\left[Y_{dj}|\boldsymbol{X}_{dj}\right] \;=\; g\left\{\boldsymbol{X}_{dj}^{t}\boldsymbol{\beta}+c_d\right\}, \quad d=1,\ldots,D;\; j=1,\ldots,n_d, \tag{3}$$

with c_d being an area (domain or group) specific fixed effect without the assumption of independence from the individual effects $\boldsymbol{X}_{dj}$.

We studied many applications where the random effects represented the effect of either a region, a climate type, a socio-economic group or the proband group in biostatistics. In almost all cases the independence assumption was hardly credible. This causes endogeneity giving inconsistent estimates for $\boldsymbol{\beta}$ and potentially woeful out-of-sample prediction performance. The affirmation, for the purpose of estimation the FEM, and for prediction the MEM, would be the right model is unfortunately wrong. For example, the FEM does not allow to include covariates which do not or hardly vary with i (*time* in panel data) for given d. A prediction with MEM when the unrealistic independence assumption is violated performs only well for in-sample prediction, and parameter estimates can not be interpreted.

We therefore propose to use a flexible modeling of area effects that allows to change continuously from a MEM (eqn 1) without area specific covariates to a Semiparametric Mixed Effects Model (SMEM) with a smooth area specific mean and a random effect (eqn 4), up to the other extreme, an FEM (eqn 3). We bridge the gap between FEM, MEM and PLM by a flexible modeling of area effects. The transition from MEM to SMEM and FEM is achieved by progressively relaxing the smoothness assumption on the semiparametric area specific impact: we start with the highest degree of smoothness (a constant) yielding to a random effects model, and end up with the lowest degree (interpolation of the area effects) yielding a fixed effects model. This way one can resolve all problems at once: model, and thus explain, the area or group effect, and dispose of the "independence assumption" problem. One obtains consistent estimates and valid inference. This is achieved without loosing the advantages of MEM, and without running into the problems we would face in a FEM. It should be emphasized that it nests MEM, FEM, and PLM. Consequently, it outperforms them all in estimation and prediction.

2 The Semiparametric Mixed Effects Model

In the following, area-specific effects can be either random or fixed. For the ease or presentation we concentrate on the nested-error model, i.e. (1) with $\boldsymbol{Z}_{dj} = 1$ and $u_d \in I\!R$. Let $\boldsymbol{W} \in I\!R^q$ denote some continuous area-specific covariates. We define the semiparametric mixed effects model (SMEM) as

$$E\left[Y_{dj}|\boldsymbol{X}_{dj}, \boldsymbol{W}_d, u_d\right] \;=\; g\left\{\boldsymbol{X}_{dj}^t\boldsymbol{\beta} + \eta_v(\boldsymbol{W}_d) + u_d\right\} \;, \tag{4}$$

where $\eta_v : I\!R^q \rightarrow I\!R$ is an unknown nonparametric function with a given "slider" v. The area effect is split in a conditional mean and a random effect. We set $E[\eta_v(\boldsymbol{W})] = 0$ for identification such that $\boldsymbol{X}_{dj}^t\boldsymbol{\beta}$ includes the intercept. We can think of v also as a smoothness parameter so that, for example, for kernel estimates we set the bandwidth to $h = v \cdot n^{-2/(4+q)}$. Then, at one extreme we have $\eta_0(\boldsymbol{W}_d) = c_d$ with $\sum_{d=1}^D c_d = 0$ and, at the other extreme, $\eta_\infty(\boldsymbol{W}_l) = 0$. In the first case η_0 captures the area effect completely via the conditional mean so that we get $u_d = 0$ for all d and a FEM (3), whereas for $h = \infty$ we obtain a MEM (1), where the area effect is regarded as a purely random effect. Finally, when h or v are between 0 and ∞ but $\sigma_u = 0$, then we face a (generalized) PLM. We see that v acts as a slider in the sense that by varying v we obtain a continuum of models between the two extremes:

$$v = 0: \;\; E\left[Y_{dj}|\boldsymbol{X}_{dj}\right] = g\left\{\boldsymbol{X}_{dj}^t\boldsymbol{\beta} + c_d\right\} \quad \leftrightarrow \quad v = \infty: \;\; E\left[Y_{dj}|u_d, \boldsymbol{X}_{dj}\right] = g\left\{\boldsymbol{X}_{dj}^t\boldsymbol{\beta} + u_d\right\}$$

and for $0 < v < \infty$

$$E\left[Y_{dj}|\boldsymbol{X}_{dj}, \boldsymbol{W}_d, u_d\right] = g\left\{\boldsymbol{X}_{dj}^t\boldsymbol{\beta} + \eta(\boldsymbol{W}_d) + u_d\right\}.$$

Let $\boldsymbol{\theta}$ be the vector of all variance and covariance parameters, including the elements of $\sigma_u^2 = Var[u_d]\ \forall d$, and set $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\theta})$. For example, in the case of a partial linear semi mixed effects model and $\rho = 1$, one typically assumes $Var[Y_{dj}|\boldsymbol{X}_{dj}, \boldsymbol{W}_d, u_d] = \sigma_e^2 \vartheta(\boldsymbol{X}_{dj}, \boldsymbol{W}_d)$ with known function $\vartheta(\cdot)$ and so $\boldsymbol{\theta} = (\sigma_u^2, \sigma_e^2)$.

When selecting the smoothing parameter the empirical researcher needs to keep in mind that goodness of fit is not the only objective; an important consideration in choosing v is to achieve (approximately)

$$Cov_v\left[u\ ,\ \{\boldsymbol{X}|\eta_v(\boldsymbol{W})\}\right] = 0\ . \tag{5}$$

Let us discuss this issue more in detail. In practice it is often the case that some of the $\boldsymbol{X}$ are correlated with area. If they are correlated with $\boldsymbol{W}$ then, clearly, estimation and prediction based a MEM (1) will yield biased results, whereas those based on a SMEM (4) will be unbiased. It might seem that the model

$$E\left[Y_{dj}|\boldsymbol{X}_{dj}, \boldsymbol{W}_d, u_d\right] = g\left\{\boldsymbol{X}_{dj}^t\boldsymbol{\beta} + \boldsymbol{W}_d^t\boldsymbol{\alpha} + u_d\right\}\ , \tag{6}$$

would also solve the problem, but this is not the case in general. That's because the dependence structure between $\boldsymbol{X}$ and the area is often complex, and not limited to a simple linear relation with the available area information $\boldsymbol{W}$. Fortunately, in practice the relation can be described sufficiently well via $\psi(\boldsymbol{W})$ for an unknown function ψ, as long as $\boldsymbol{W}$ varies continuously over the different areas. Imagine the relation between $\boldsymbol{X}_{dj}$ and the area d is summarized in some (possibly latent) variable $\boldsymbol{J}_d$, i.e. $\boldsymbol{X}_{dj} = \boldsymbol{J}_d + \widetilde{\boldsymbol{X}}_{dj}$ with $\widetilde{\boldsymbol{X}}_{dj}$ being independent of any area effect. Our claim is that we can always find a function ψ fulfilling $\psi(\boldsymbol{W}_d) = \boldsymbol{J}_d + \boldsymbol{V}_d$

with $\boldsymbol{V}_d$ defined implicitly. In the case $\boldsymbol{W}_d \equiv \boldsymbol{J}_d$, ψ is the identity. Another particular case is when ψ simply assigns $\boldsymbol{J}_d$ to $\boldsymbol{W}_d$ for each $d = 1, \ldots, D$, i.e. doing interpolation. Recall that η_v is a nonparametric function with appropriate smoothness v. Then, for an implicitly defined φ_ω we get

$$\begin{aligned} E\left[Y_{dj}|\boldsymbol{X}_{dj}, \boldsymbol{W}_d, u_d\right] &= g\left[\boldsymbol{X}_{dj}^t\boldsymbol{\beta} + \eta_v(\boldsymbol{W}_d) + u_d\right] = \\ = g\left[\boldsymbol{X}_{dj}^t\boldsymbol{\beta} + \varphi_\omega\left\{\psi(\boldsymbol{W}_d)\right\} + u_d\right] &= g\left[\boldsymbol{X}_{dj}^t\boldsymbol{\beta} + \varphi_\omega\left\{\boldsymbol{J}_d + \boldsymbol{V}_d\right\} + u_d\right] \ , \end{aligned} \tag{7}$$

where φ_ω is again a nonparametric function with a smoothness parameter ω. This one depends on v and the smoothness of ψ or, vice verse, v depends on ω and ψ. From (7) we see that this model does not suffer from dependency between $\boldsymbol{X}_{dj}$ and u_d, i.e. endogeneity of $\boldsymbol{X}_{dj}$. Consequently, η_v can filter out the endogeneity in practice. In econometric terms, $\eta_v(\boldsymbol{W}_d)$ can be regarded as a nonparametric proxy. Even if $\boldsymbol{W}$ alone is a poor proxy, we just need to set $v \approx 0$. With that choice our SMEM reverts to (almost) an FEM, and is therefore free of endogeneity.

3 Estimation and Asymptotic Behavior

We will analyze the statistical properties by different means. First, we give estimation procedures and summarize their asymptotic behavior. Then, we will study the finite sample performance in contexts where multi-level models are used. For brevity we concentrate on cross sectional data.

There exist alerady some estimation procedures for semiparametric multi-level models (Lin and Carroll, 2001), and quite recently for MEM with semiparametric impact of $\boldsymbol{X}$ (see Opsomer, Claeskens, Ranalli, Kauermann, and Breidt, 2008). There also exist several estimation procedures based on Bayesian approaches, some combined with penalized splines, some with MCMC methods; see references in Section 1. Lin and Carroll (2001), and Lombardía and Sperlich (2008) proposed profile likelihood estimators. Lin and Carroll (2001) tackled correlated responses but without explicit random effects. Lombardía and Sperlich (2008) accounted for the dependence structure in generalized MEM, but only when estimating β.

Here we propose to also take account of the covariance structure when estimating the nonparametric part using ideas of Vilar-Fernández and Francisco-Fernández (2002) or Lin and Carroll (2001)for likelihood based estimators. For the latter we maximize a log-likelihood say $l(\cdot;\eta,\boldsymbol{\delta})$. When we apply the integral approach, this is based on the marginal density

$$f(Y_{dj}|\boldsymbol{W}_d,\boldsymbol{X}_{dj};\eta,\boldsymbol{\delta}) \;=\; \int f(Y_{dj}|\boldsymbol{u},\boldsymbol{W}_d,\boldsymbol{X}_{dj};\eta,\boldsymbol{\beta},\sigma_e^2)p(\boldsymbol{u};\boldsymbol{\sigma}_u)d\boldsymbol{u}, \qquad (8)$$

where $p(\cdot)$ is the density of the random effects. When we apply the posterior mode approach (easier to calculate by using EM-algorithms), see Fahrmeir and Tutz (2001), we maximize the logarithm of

$$\prod_{d=1}^{D}\prod_{j=1}^{n_d} f(Y_{dj}|\boldsymbol{u}_d,\boldsymbol{W}_d,\boldsymbol{X}_{dj};\eta,\boldsymbol{\delta})\prod_{d=1}^{D}p(\boldsymbol{u}_d;\boldsymbol{\sigma}_u)\ . \qquad (9)$$

To estimate η_v at a fixed point $\boldsymbol{w}_0$, one works with an empirical counterpart of its condition expectation, i.e. $E\left[l(Y;\eta,\boldsymbol{\delta})|\boldsymbol{W}=\boldsymbol{w}_0\right]$. We conclude from Lombardía and Sperlich (2008)that under rather common smoothness conditions on f and η_v

a) $\sqrt{n}(\hat{\boldsymbol{\delta}}-\boldsymbol{\delta}) \xrightarrow{d} N\left(0, I_{\boldsymbol{\delta}}^{-1}\right)$, where $I_{\boldsymbol{\delta}}$ is the (marginal) Fisher information matrix of $l(\cdot)$.

b) defining $h_{prod}=\prod_{j=1}^{q}h_j$ and $h_{max}=\max_{1\le j\le q}h_j$, $\boldsymbol{w}_0$ being from the interior of the support of $\boldsymbol{W}$, and $p_W(\cdot)$ its density function, then

$$\sqrt{nh_{prod}}\left(\hat{\eta}_v(\boldsymbol{w}_0) - \eta_v(\boldsymbol{w}_0) - B_\eta(\boldsymbol{w}_0)\right) \xrightarrow{d} \mathrm{N}\left(0, Var_\eta(\boldsymbol{w}_0)\right),$$

with bias $B_\eta(\boldsymbol{w}_0) = O(h_{max}^2)$

and variance $Var_\eta(\boldsymbol{w}_0) = \frac{\int K(\boldsymbol{w})^2 d\boldsymbol{w}}{p_W(\boldsymbol{w}_0)\ E\left[\frac{\partial}{\partial \eta_v} l\left(Y;\eta_v,\boldsymbol{\delta}_0\right)^2 | \boldsymbol{W}=\boldsymbol{w}_0\right]}$.

While typically it is assumed that the likelihood is correctly specified, this is not necessary for consistency, see Jiang and Lahiri (2006) for references. As can be seen from b), asymptotically, h_{max} has to go to zero and so do all the other bandwidths, but such that $nh_{max} \to \infty$. From Maity, Ma, and Carroll (2007) we conclude that using these estimates produce efficient predictors for forecasting area-specific means.

But how do we get such an estimator in practice? Fahrmeir and Tutz (2001) discuss computational expensive algorithms for nonlinear links $g(\cdot)$ but in else somewhat simpler context. To understand better the efficient estimation of η, let us first consider $g = identity$ and introduce the following notation: define $\mathbf{1} := \mathrm{diag}\{\mathbf{1}_{n_d}\}_{d=1}^D$, $\boldsymbol{Y} = (Y_{1,1}, Y_{1,2}, \ldots, Y_{D,(n_D-1)}, Y_{D,n_D})^t$, a vector of i.i.d. zero-mean errors $\boldsymbol{\epsilon}$, and $\boldsymbol{X} \in I\!R^{n\times p}$, $\boldsymbol{u} \in I\!R^{D\times\rho}$, $\boldsymbol{Z} \in I\!R^{n\times(D\rho)}$, $\boldsymbol{W} \in I\!R^{D\times q}$. Then our model (4) writes as

$$\boldsymbol{Y} \quad = \quad \boldsymbol{X\beta} + \mathbf{1}\ \eta_v(\boldsymbol{W}) + \boldsymbol{Zu} + \boldsymbol{\epsilon}\ , \tag{10}$$

where $\eta_v(\boldsymbol{W})$ means row-wise application of $\eta_v : I\!R^q \to I\!R$. We propose profiled likelihood neglecting the integral approach as we are not only interested in the estimation of $\boldsymbol{\delta}$ but also in the prediction of the random effects. We want to account for the covariance structure when estimating the nonparametric functions, similar to Vilar Fernández and Francisco Fernández (2002). More specific, we nest the iteration steps of feasible profile likelihood estimation in parametric MEM (see e.g. Chapter 6 in Rao, 2003) with the ones of profile likelihood estimation in partial linear regression estimation . Let $\hat{H}$ be the so-called smoothing hat matrix of a Nadaraya-Watson kernel smoother around $\boldsymbol{W}$, with quartic kernel and bandwidth h. Let $\bar{K}_h = \text{diag}\{\mathbf{1}_{n_d}\ K_h(\boldsymbol{W}_d - w)\}_{d=1}^{D}$, $w \in I\!R^q$, and $K_h(\omega) = \prod_{j=1}^{q} \frac{1}{h} K(\omega_j/h)$, $\omega \in I\!R^q$, a product kernel. Then, in the $(k+1)$th iteration we calculate

$$
\begin{aligned}
\boldsymbol{\beta}^{(k+1)} &= \left\{(\boldsymbol{X} - \hat{H}\boldsymbol{X})^t V^{-1} \boldsymbol{X}\right\}^{-1} (\boldsymbol{X} - \hat{H}\boldsymbol{X})^t V^{-1} \left\{\boldsymbol{Y} - \mathbf{1}\eta_v^{(k)}(\boldsymbol{W})\right\}, \\
V &= \boldsymbol{I}_n \sigma_e^{2(k)} + \boldsymbol{Z}\sigma_u^{2(k)}\boldsymbol{Z}^t, \qquad A^{-1} = \boldsymbol{I}_D + \boldsymbol{Z}^t \frac{1}{\sigma_e^{2(k)}} \boldsymbol{Z}\sigma_u^{2(k)}, \\
\eta_v^{(k+1)}(w) &= \left(\mathbf{1}_n^t V^{-1/2} \bar{K}_h V^{-1/2} \mathbf{1}_n\right)^{-1} \mathbf{1}_n^t V^{-1/2} \bar{K}_h V^{-1/2} \left(\boldsymbol{Y} - \boldsymbol{X}^t \boldsymbol{\beta}^{(k+1)}\right), \\
\boldsymbol{u}^{(k+1)} &= \sigma_u^{2(k)} \boldsymbol{Z}^t V^{-1} \left(\boldsymbol{Y} - \boldsymbol{X}^t \boldsymbol{\beta}^{(k+1)} - \mathbf{1}\eta_v^{(k+1)}(\boldsymbol{W})\right), \\
\boldsymbol{\epsilon}^{(k+1)} &= \boldsymbol{Y} - \boldsymbol{X}^t \boldsymbol{\beta}^{(k+1)} - \mathbf{1}\eta_v^{(k+1)}(\boldsymbol{W}) - \boldsymbol{Z}^t \boldsymbol{u}^{(k+1)}, \\
\sigma_e^{2(k+1)} &= \frac{1}{n} \boldsymbol{Y}^t \boldsymbol{\epsilon}^{(k+1)}, \text{ and } \sigma_u^{2(k+1)} = \frac{1}{D}\left(\boldsymbol{u}^{(k+1)^t} \boldsymbol{u}^{(k+1)} + \sigma_u^{2(k)} tr(A)\right).
\end{aligned}
$$

For the case when $g \neq identity$, this algorithm can be completed by including a local scoring step to (re-)linearize the estimation problem.

We conclude with three remarks. First, the extensions of SMEM either to allow for a non- or semiparametric modeling of the impact of $\boldsymbol{X}$ or to impose some structure on η_v like additivity are quite straight and therefore not discussed here. Second, as mentioned in the introduction, the use of P-splines has become quite popular in the context of mixed effects models, and some of the just enumerated extensions may be easier to implement in splines than with kernels. However, while it is clear how to achieve $v \to 0$ for kernel based estimators, in the case of P-splines we have to change both, the penalizing coefficient (usually denoted as λ) and the number of knots K. To guarantee a *smooth transition* one would also need additional information to select an appropriate combination of λ and K. In practice this makes a smooth and continuous transition from FEM to MEM much less convenient. Third, Hall and Maiti (2006) have introduced nonparametric estimation of mean-squared prediction error, and Lombardía and Sperlich (2008) methods for testing in semiparametric mixed effects models. These methods directly carry over to the here introduced SMEM, and can therefore be used for further inference. Both introduce bootstrap methods used in the following.

4 Empirical Evidence and Illustration

For $(X_{dj,1}, X_{dj,2})^t = \boldsymbol{X}_{dj} \in I\!R^2$ consider model (10) with $\eta_v(\boldsymbol{W}) = \sum_{k=1}^q \sin(2.5W_{d,k})$, where $W_{d,k} \sim U[0,2]$ i.i.d.$\forall$ k, $u_d \sim N(0, \sigma_u^2)$ i.i.d., and $\epsilon \sim N(0, \sigma_e^2)$ i.i.d. For $i = 1, 2$ we generated $X_{dj,i} = 0.8 \cdot O_{dj,i} + 0.5 \sum_{k=1}^q W_{d,k}^2$ with $O_{dj,i} \sim N(0, 1)$ i.i.d. Further, $\beta_0 = 1.5$, $\beta_1 = 1.5$, and $\beta_2 = 1$. We consider only data generating processes (DGP hereafter) with $q = 1$ or $q = 2$. This gives $Var[X_{dj,i}] \approx 1$ with $Corr[X_{dj,i}, W_d] \approx 0.29$ for $q = 1$, and $Var[X_{dj,i}] \approx 1.35$ with $Corr[X_{dj,i}, W_{d,k}] \approx 0.25$ for $q = 2$; $i, k = 1, 2$. Simulation results are based on $n = 250$, $d = 50$ with different $\boldsymbol{\theta} = (\sigma_u^2, \sigma_e^2)$, and all results given refer to 250 simulation runs. The implementations of FMEM, MEM, SMEM and PLM are nested algorithms to guarantee a fair comparison. We use quartic kernels to estimate η and will henceforth speak of bandwidth h and slider v synonymously. In our simulations the convergence criterion was a change of $\hat{\boldsymbol{\delta}}$ by less than 0.01% in the Euclidian norm. This was reached after no more than about ten iterations for the real, and about five for the simulated data. In Figure 1 are plotted the impacts of the systematic area effect for $q = 1, 2$. The sine function has been chosen for two reasons: first, trigonometrical functions are known to be hard to estimate by the common nonparametric methods; second, this way the dependence structure between $\boldsymbol{X}$ and the area impact is particularly complex.

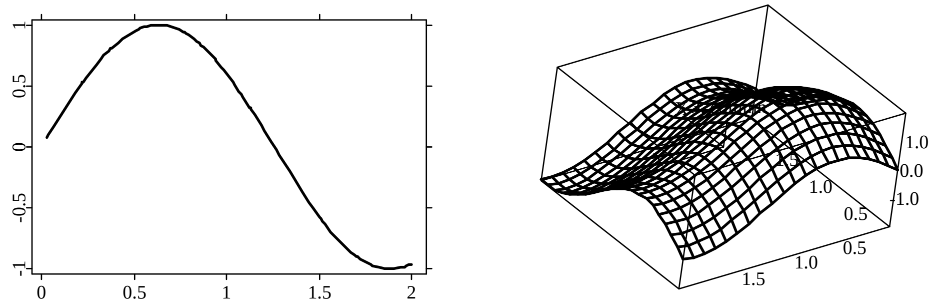

Figure 1: The systematic area effect $\eta_v(\boldsymbol{W})$ for $q = 1$ (left) and $q = 2$ (right).

Bias and Mean Squared Errors of $\hat{\boldsymbol{\beta}}$ and $\hat{\eta}_v$

As the independence assumption of the MEM is violated, the estimator $\hat{\boldsymbol{\beta}}$ will be inconsistent, but it is not clear how mean squared error (MSE) and bias will change with bandwidths $0 \leq h \leq \infty$; they are given as functions of $\ln(h)$ in Figure 2. In our DGP we set $q = 1$ and $\boldsymbol{\theta} = (0.25, 0.5)$.

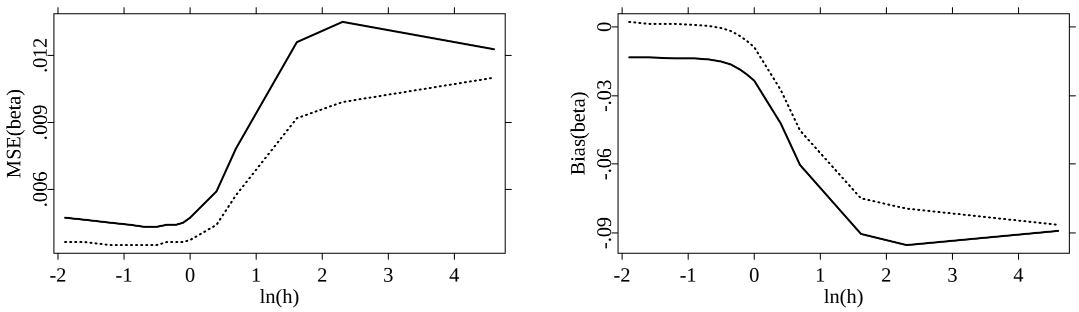

Figure 2: MSE($\hat{\boldsymbol{\beta}}$) and Bias($\hat{\boldsymbol{\beta}}$) for β_1 (solid) and β_2 (dotted) as functions of bandwidth h when $q = 1$ and $\boldsymbol{\theta} = (0.25, 0.5)$.

We compare SMEM first with MEM, afterwards with PLM. Biases and variances for $\hat{\boldsymbol{\beta}}$ are summarized in Table 1 for models with different $\boldsymbol{\theta}$ and q. Obviously, the bias can become quite severe when the independence assumption does not hold. At the same time the variance is not larger in the SMEM, despite the fact that nonparametric estimation is applied. Before studying this point in greater detail, recall that we make use of the variance estimates $\hat{\boldsymbol{\theta}}$ in both MEM and SMEM. Therefore, we also examined the estimates of $\boldsymbol{\theta}$. However, a fair comparison is not possible because, in the MEM, part of the variation due to η will be erroneously assigned to σ_u^2, such that the SMEM always beats the MEM by far.

Model			$q = 1$				$q = 2$			
	σ_e^2	σ_u^2	bias($\hat{\beta}_1, \hat{\beta}_2$)		var($\hat{\beta}_1, \hat{\beta}_2$)		bias($\hat{\beta}_1, \hat{\beta}_2$)		var($\hat{\beta}_1, \hat{\beta}_2$)	
	.1	.0	-.0315	-.0305	.0009	.0008	-.0327	-.0294	.0008	.0008
		.25	-.0215	-.0203	.0008	.0007	-.0263	-.0230	.0008	.0007
M		.5	-.0166	-.0154	.0008	.0007	-.0221	-.0189	.0008	.0007
E	.5	.0	-.1336	-.1311	.0046	.0040	-.1617	-.1533	.0051	.0050
M		.25	-.0893	-.0864	.0043	.0036	-.1206	-.1124	.0043	.0042
		.5	-.0690	-.0662	.0042	.0035	-.0987	-.0908	.0041	.0040
	.1	.0	-.0045	-.0033	.0007	.0007	-.0072	-.0046	.0007	.0007
S		.25	-.0028	-.0015	.0008	.0007	-.0036	-.0008	.0008	.0007
M		.5	-.0025	-.0013	.0008	.0007	-.0030	-.0002	.0008	.0007
E	.5	.0	-.0084	-.0057	.0032	.0033	-.0136	-.0093	.0032	.0031
M		.25	-.0074	-.0042	.0038	.0035	-.0104	-.0049	.0037	.0034
		.5	-.0067	-.0035	.0040	.0036	-.0086	-.0028	.0038	.0034

Table 1: Bias and variance of $\hat{\boldsymbol{\beta}}$. The SMEM refers to $h = 0.5$.

Although Vilar Fernández and Francisco Fernández (2002) considered a quite different model, it is clear from their results that our SMEM will be more efficient than common PLM estimators of η_v, and, in our context, of $\hat{\boldsymbol{\beta}}$. For the numerical performance this is even true if $\sigma_u^2 = 0$, since when estimating model (4), our random effects with estimate $\hat{\sigma}_u^2$ will correct for a possible over- or undersmoothing of the impact of $\boldsymbol{W}$ and vice versa. In Table 2 are compared the expected mean squared errors (EMSE hereafter) of $\hat{\eta}_v$ and $\hat{\boldsymbol{\beta}}$ resulting from SMEM and PLM for $h = 0.5$, and of $\hat{\eta}_v$, defined by $E\left[\{\eta_v(\boldsymbol{W}) - \hat{\eta}_v(\boldsymbol{W})\}^2\right]$. The results support our expectation, namely that SMEM also outperforms the PLM.

Model	σ_e^2	σ_u^2	$q=1$ EMSE($\hat{\eta}_v$)	$q=1$ MSE($\hat{\beta}_1, \hat{\beta}_2$)		$q=2$ EMSE($\hat{\eta}_v$)	$q=2$ MSE($\hat{\beta}_1, \hat{\beta}_2$)	
	.1	.0	.0150	.0007	.0007	.0344	.0009	.0008
		.25	.0340	.0022	.0021	.0933	.0021	.0018
P		.5	.0529	.0037	.0035	.1520	.0033	.0028
L	.5	.0	.0247	.0033	.0034	.0640	.0035	.0032
M		.25	.0440	.0048	.0046	.1231	.0048	.0043
		.5	.0630	.0063	.0060	.1818	.0060	.0053
	.1	.0	.0148	.0007	.0007	.0320	.0007	.0007
S		.25	.0311	.0008	.0007	.0842	.0008	.0007
M		.5	.0474	.0008	.0007	.1371	.0008	.0007
E	.5	.0	.0246	.0033	.0033	.0630	.0034	.0032
M		.25	.0419	.0039	.0035	.1159	.0038	.0034
		.5	.0585	.0040	.0036	.1681	.0038	.0034

Table 2: Expected mean squared errors of $\hat{\eta}_v$ and $\hat{\boldsymbol{\beta}}$ for PLM and SMEM($h = 0.5$).

Prediction Power in and out of Samples

A most often mentioned argument in favor of MEMs is their superior prediction power, which is important for data matching when imputing factors for individuals, and in small area statistics to predict area-, or say macro-, level parameters. The next simulation study (see Table 3) was designed to assess this claim, by comparing the prediction power of MEMs and SMEMs for (a) in-sample prediction, (b) out-of-sample prediction, (c) individuals, and (d) area-levels.

Our in-sample prediction risk measure is simply the average over the 250 simulation runs of the mean squared error. We denote this measure by ASE (averaged squared errors); it is defined by $ASE = \frac{1}{250}\sum_{repl=1}^{250} MSE_{repl}$, where $MSE = \frac{1}{n}\sum_{d=1}^{D}\sum_{j=1}^{n_d}(Y_{dj} - \hat{Y}_{dj})^2$, with $\hat{Y}_{dj} = \hat{\boldsymbol{\beta}}^t \boldsymbol{X}_{dj} + \hat{\beta}_0 + \hat{u}_d$, the so-called feasible EBLUP for the MEM, and $\hat{Y}_{dj} = \hat{\boldsymbol{\beta}}^t \boldsymbol{X}_{dj} + \hat{\beta}_0 + \hat{\eta}_v(\boldsymbol{W}_d) + \hat{u}_d$ for the SMEM. Note that $\hat{\boldsymbol{\beta}}$ and $\hat{u}$ are certainly different for the two models.

Our out-of-sample risk is the mean squared prediction error for two particular $\boldsymbol{X}$, called $\text{MSE}(\hat{Y}_l, \hat{Y}_s)$: for $\boldsymbol{X}_l = (\frac{2q}{3} + 2.5, \frac{2q}{3} + 2.5)$, respectively $\boldsymbol{X}_s = (\frac{2q}{3} - 2.5, \frac{2q}{3} - 2.5)$, each in a different but fixed area.

Model			$q = 1$			$q = 2$		
	σ_e^2	σ_u^2	ASE	$\text{MSE}(\hat{Y}_l, \hat{Y}_s)$		ASE	$\text{MSE}(\hat{Y}_l, \hat{Y}_s)$	
	.1	.0	.0208	.0780	.1286	.0214	.0995	.1806
		.25	.0205	.0562	.0816	.0211	.0775	.1360
M		.5	.0205	.0486	.0652	.0210	.0663	.1126
E	.5	.0	.1004	.7701	1.659	.1179	1.625	3.344
M		.25	.0987	.4634	.8589	.1093	.9843	1.966
		.5	.0988	.3540	.5947	.1063	.7199	1.406
	.1	.0	.0103	.0368	.0261	.0154	.0393	.0576
S		.25	.0194	.0352	.0409	.0198	.0424	.0566
M		.5	.0200	.0363	.0413	.0203	.0426	.0557
E	.5	.0	.0251	.1152	.1012	.0461	.1565	.2294
M		.25	.0797	.1515	.1829	.0862	.1976	.2785
		.5	.0894	.1651	.1958	.0936	.2056	.2810

Table 3: The average squared error (ASE) of the inside-sample predictors, and the mean squared error of $\hat{Y}_l$ and $\hat{Y}_s$ when SMEM is estimated with $h = 0.5$.

Due to the nature of the MEM which basically fits the area effect with random coefficients, it is clear that an in-sample prediction will always do well in terms of the mean squared error. In contrast, for small and moderate sample sizes nonparametric methods such as we use to estimate η in our SMEM can have very poor numerical performance. Nevertheless, the results in Table 3 show that SMEMs clearly outperform the MEM, even in the ASE, and substantially in the out-of-sample prediction. Recall that valid inference with MEM is hardly possible for this DGP as the available methods are typically model based (*model biased*) and therefore inconsistent.

Calculation of Area Parameter (Macro Indices)

When calculating macro indices, also called prediction of area-level parameters in small area statistics, MEM is expected to perform reasonably well. As this is partly an in-sample prediction problem of aggregates, it should easily compete with SMEM regardless of possible violation of the independence assumption. Let us predict, for each area $d = 1, ..., 50$, two indices (parameters at the area level): (i) $\mu_d = E[\bar{Y}_{d\bullet}|\boldsymbol{X}_d, \boldsymbol{W}_d, u_d]$, this is assuming that the number of population units in the dth area is large; and (ii) $\bar{Y}_{d\bullet} = \sum_{j=1}^{N_d} y_{dj}/N_d$, assuming a super-population regression model of the form (10) for the N_d population units in the dth area. When considering (ii), the best linear unbiased estimator of $\bar{Y}_{d\bullet}$ is given by

$$\hat{\bar{Y}}_{d\bullet} = f_d \; \bar{y}_{sd} + (1 - f_d)\hat{\mu}_{nd},$$

where $f_d = n_d/N_d$, $\bar{y}_{sd}$ is the average of the in-sample values and $\hat{\mu}_{nd}$ is the predictor of μ_d for the $(N_d - n_d)$ non-sampled units. Therefore we consider now the situation in which we wish to predict Y for some individuals for whom X is available.

We performed two simulation runs with $q = 2$, but with 10 observations X_{dj} for each area, whereas Y_{dj} was observed only for the first 5 individuals. In a first run all X_{dj} were randomly drawn ($d = 1, ..., 50$ and $j = 1, ..., 10$). In a second run we set

$$X_{d6} = (-1, -1),\; X_{d7} = (0.16, 0.16),\; X_{d8} = (1.33, 1.33), \qquad (11)$$
$$X_{d9} = (2.5, 2.5),\; \text{and}\; X_{d10} = (3.67, 3.67)$$

for all areas. They are independent of the area effect u_d but $(-1, 0.16, 1.33,$

2.5, 3.67) represent (approximately), for each element X of $\boldsymbol{X}$,

$$(E[X] - 2\sigma_X,\ E[X] - \sigma_X,\ E[X],\ E[X] + \sigma_X,\ E[X] + 2\sigma_X)$$

with σ_X denoting its standard deviation unconditionally from the area. Each value Y_{dj} was generated with its corresponding X_{dj} ($j = 1, ..., 5$). As before, we show only results for bandwidth $h = 0.5$. We note, however, that in our simulations using $h \approx 0.7$, the SMEM outperformed the MEM by an even greater extent.

The results are given in form of box-plots which show the distributions of the $D = 50$ mean squared errors for different data generating process. In each plot the box-1 and box-2 refer to SMEM and SEM respectively, with X_{d6} to X_{d10} taken randomly; box-3 and box-4 refer to SMEM and MEM respectively, with X_{d6} to X_{d10} as in (11). For clarity of illustration the displays do not show the extreme (large) mean squared errors for the MEM (between 2 to 5% of the points).

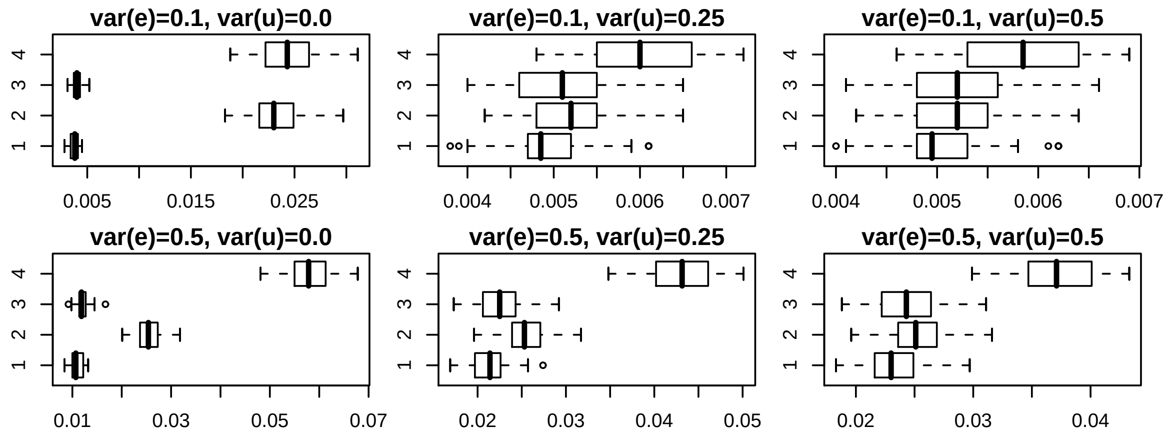

Figure 3: Predicting $\bar{Y}_{d\bullet}$: Mean squared error distributions over the 50 predicted area level parameter $\bar{Y}_{d\bullet}$ for different (σ_e^2, σ_u^2). Box 1 refers to SMEM and box 2 to MEM, with X_{d6} to X_{d10} randomly; Box 3 refers to SMEM and box 4 to MEM, with X_{d6} to X_{d10} as in (11).

The mean squared errors are quite small, especially when predicting (ii) $\bar{Y}_{d\bullet}$, where

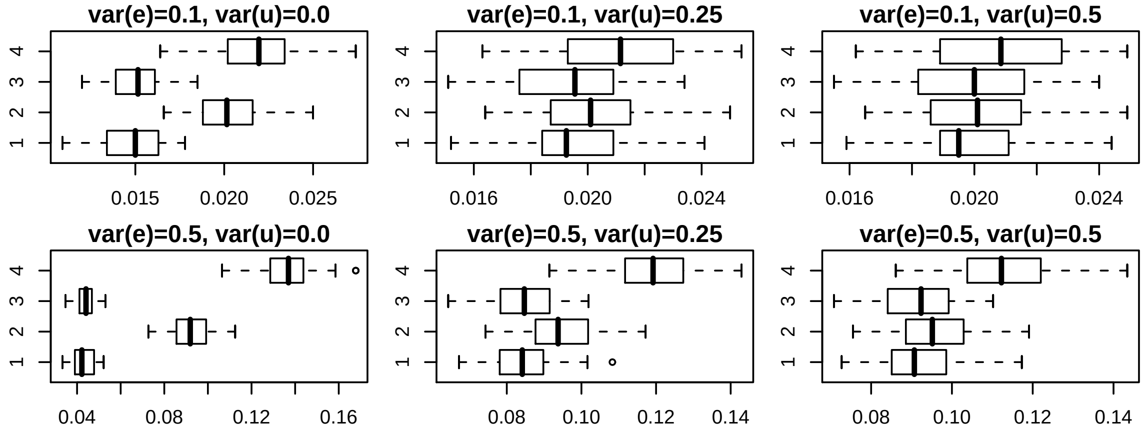

Figure 4: Predicting μ_d: Mean squared error distributions over the 50 predicted area level parameter μ_d for different (σ_e^2, σ_u^2). Box 1 refers to SMEM and box 2 to MEM, with X_{d6} to X_{d10} randomly; Box 3 refers to SMEM and box 4 to MEM, with X_{d6} to X_{d10} as in (11).

half of the information (Y_{d1} to Y_{d5}) is given. There, the differences between the prediction based on MEM compared to the prediction based on SMEM is restricted to the out-of-sample prediction. Consequently, as can be seen in all graphs of Figure 3, the SMEM outperforms MEM by far. Turning to the prediction of (i) $\mu_d = E[\bar{Y}_{d\bullet}|\boldsymbol{X}_d, \boldsymbol{W}_d, u_d]$, the mean squared errors have increased significantly, representing half in-sample and half out-of-sample prediction errors. The comparisons of the ASE in Table 3, indicate that the advantage of the SMEM over the MEM for in-sample prediction can be fairly small, though always visible. Thus, examining the box-plots given in Figure 4, it is surprising that the superiority of the SMEM over MEM is so marked.

Exactness of Bootstrap Approximates

Finally, we checked whether bootstrap procedure (Lombardía and Sperlich, 2008) works for a DGP with $q = 1$. We carried out only 100 simulation runs with 200 bootstrap replicates being aware that this will give only a rough approximation.

The results are given in Table 4. Following the recommendation of Härdle and Marron (1991), the bootstrap model was constructed using a pilot bandwidth g greater than h, ($g = 1.1\ h$). We also tried other values for g; the results were similar. As can be seen, the results confirm that the bootstrap procedure can serve as a reasonable tool for doing inference in our SMEMs. Additional simulations, not shown here, revealed that the bootstrap does reasonably well at estimating the variance but, in some cases, does less well in estimating the bias, which is not surprising, as this is generally the case for bootstrap in nonparametrics.

	σ_e^2	σ_u^2	MSE($\hat{\beta}_1$,	$\hat{\beta}_2$)	MSE($\hat{\sigma}_e^2$,	$\hat{\sigma}_u^2$)	MSE($\hat{Y}_l$,	$\hat{Y}_s$)
O	.1	.25	.0008	.0007	.0001	.0026	.0352	.0409
B	.1	.25	.0007	.0007	.0001	.0030	.0337	.0388
O	.5	.25	.0039	.0035	.0024	.0043	.1515	.1829
B	.5	.25	.0035	.0036	.0025	.0050	.1455	.1684

Table 4: Bootstrap approximations (**B**) of actual mean squared errors (**O**) for $\hat{\beta}$, $\hat{\boldsymbol{\theta}}$, and predictors ($\hat{Y}_l, \hat{Y}_s$). Estimates and predictions were calculated in the SMEM, dimension $q = 1$, with bandwidth $h = 0.5$.

5 Analyzing Tourist Expenditures in Galicia

We now apply of our model class in the context of small area statistics predicting average tourist expenditures in the 53 counties of Galicia, a region in the Northwest of Spain. As with the rest of the country, tourism is one of the most important sources of revenue. Therefore, official statistics and politics have a strong interest in acquiring information about the expenditure behavior of tourists. Presently, the Galician Statistical Institute (IGE) is focusing its efforts on extending their statistics to county level, and to the level of the so-called *comarcas* of which 53 exist in Galicia. Obviously, the task of obtaining reliable information about a

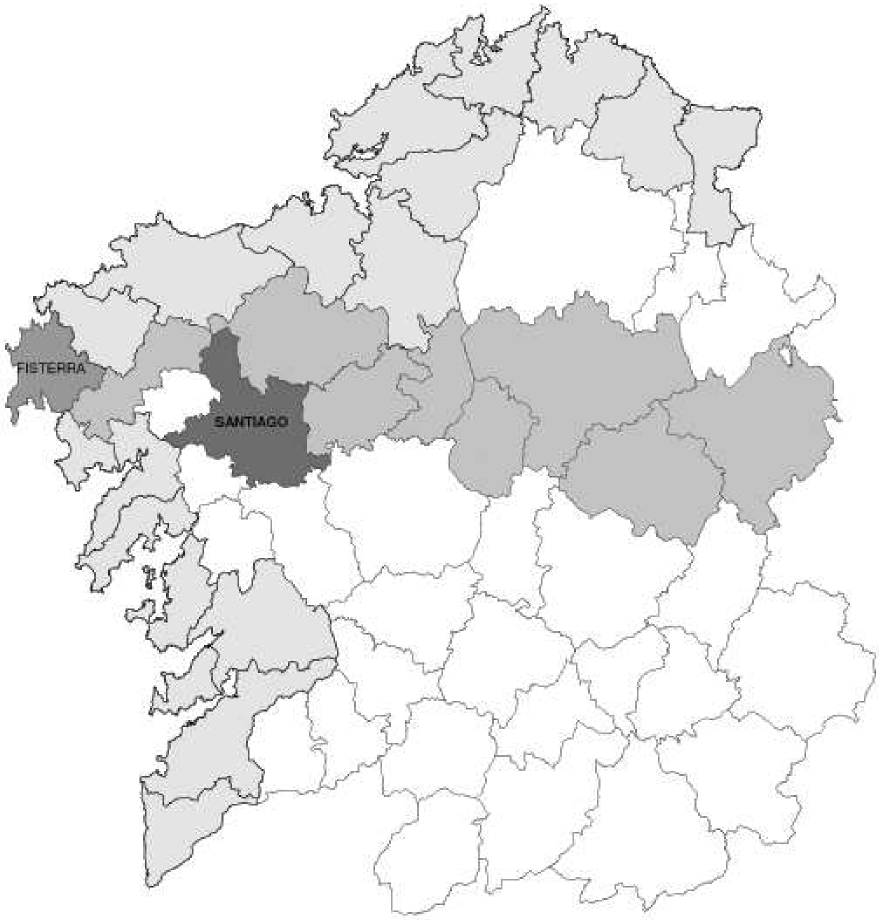

Figure 5: Map of Galicia, with all comarcas having coast are in light grey, comarcas the French trail passes through in grey. *Fisterra* has both, and *Santiago* is the pilgrim center and the capital.

tourists expenditure is cumbersome and expensive, and one must make do with modest sample sizes, i.e. interviewing in detail maybe 10 individuals per comarca. A peculiarity of Galicia is the famous pilgrim trails to Santiago de Compostela, especially the so-called *French trail*. For example, in the *holy year* of 2004 about 180000 pilgrims visited Galicia. Another tourist attraction is certainly the coast. Figure 5 shows the map of the comarcas of Galicia; with all comarcas having coast in light grey and the comarcas that the French trail passes through in grey. The comarca *Fisterra* has both peculiarities, and *Santiago* is the pilgrim center and capital. Also relevant is the fact that rural tourism is growing in the sparsely populated areas.

We make use of an own survey organized by the University of Santiago de Compostela in 2004. It contains information on 10 tourists per comarca, including average expenditures per day and several characteristics of the individuals. Our dependent variable, expenditure, refers to total expenditure, including accommodation, food, purchases, travel, leisure activities, and miscellaneous expenditure. The presented study uses the set of variables described in Table 5.

The dependent variable				
lexp	ln of total expenditure per day & cap.	4.064	.6464	4.086
Variables of the individuals				
sex	= 1 if male	.4774	.4995	.0000
age1	= 1 if strictly younger than 29	.2340	.4233	.0000
age2	= 1 if 29 $\leq$ age $\leq$ 65	.7057	.4557	1.000
single	= 1 if single	.4094	.4917	.0000
child	= 1 if children $\leq$ 16 years old	.2792	.4486	.0000
ngal	= 1 if not from Galicia	.7453	.4357	1.000
educ	= 1 if academic	.4981	.5000	.0000
stud	= 1 if student	.1226	.3280	.0000
self	= 1 if self-employed	.1000	.3000	.0000
pilgr	= 1 if pilgrim	.1189	.3236	.0000
family	= 1 visit family, friends, etc.	.3868	.4870	.0000
stay	measured in days	16.74	17.71	10.00
Variables of the comarca				
lpopd	ln of population density	3.276	.8068	3.156
ftrail	= 1 if French pilgrim trail	.0440	.0913	.0000
coast	= 1 if coast	.0839	.1122	.0000

Table 5: Descriptive statistics: mean, standard deviation, and median.

We included all three area variables in η_v to account for interactions. Other variables of the comarcas, which might seem important, like the index of tourism, the index of bars and restaurants, and the index of economic activity were disregarded for a number of reasons. Firstly, they have two sources for endogeneity: measurement error and simultaneity. In some regions of Spain the sales reported by bars and restaurants are about the same amount that tourists claim to have consumed there. So, unless one believes that residents do not consume in bars and restaurants, either the owners under-report the sales or the tourists over-report their consumption. Similar problems occur for the economic activity, e.g. the construction branch has an important impact on the Gross National Product but it is widely believed that, alone in this sector, more than 30% of the real turnover is paid cash in hand without VAT (value-added tax). Simultaneity is also evident. Finally, the exclusion of these indices is justified by problems of multicollinearity: all three indices are strongly correlated (up to 99.4%) between each other and with population density (up to 98.8%). These correlations also indicate that lpopd is a good instrument for the indices "tourism" and "bars and restaurants".

The main interest is in predicting the mean expenditures, i.e.

$$\bar{Y}_{d\bullet} := \beta_0 + \frac{1}{n_d} \sum_{j=1}^{n_d} \boldsymbol{X}_{dj}^t \boldsymbol{\beta} + \eta_v(\boldsymbol{W}_d) + u_d \quad \text{for all } d = 1, \ldots, D,$$

or, preferably, the deviations $\tilde{Y}_{d\bullet} := \bar{Y}_{d\bullet} - \bar{Y}_{\bullet\bullet}$. For the sake of brevity we only report results on estimates with bandwidths $h = h_c \boldsymbol{\sigma}_W$ for $h_c = 0$ (giving a FEM), 0.4 (giving our SMEM), and 1000 (giving a MEM), where $\boldsymbol{\sigma}_W$ is the vector of standard deviations for the comarca covariates. While the coefficient estimates, discussed later, seem to not to change much for the different models, the estimates of the macro-parameters do. We found that, for half of the comarcas d, their $\tilde{Y}_{d\bullet}$ changes significantly with the chosen model; some of the values tripled when changing from one model to another, others changed signs, etc. In Figure 6 we see how the distribution of the predicted means changes smoothly (with v being the slider) from fixed to mixed effects models. For a better comparison we plotted all densities on the same scales. It is very clear that the predicted means are more spread for the (most flexible) FEM, and that the spread shrinks as the flexibility of the model is reduced, i.e. as we approach the MEM, at which point the spread is at its smallest. Not evident in these plots are the substantial changes in the predicted means of most of the 53 comarcas. The effect evident in Figure 6 could be caused simply by shrinking the predicted means towards the center.

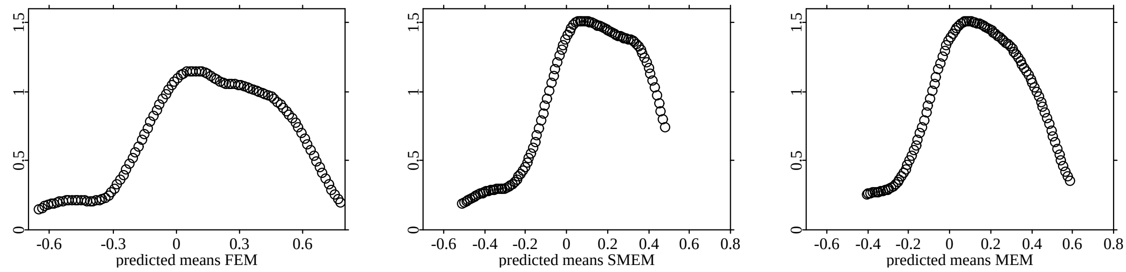

Figure 6: Densities of the predicted means $\widehat{\tilde{Y}_{d\bullet}}$ for the different models.

A different way to look at the changes when moving from FEM to MEM is given in Figure 7. From these graphs now we see that the differences are substantial, and the bootstrap estimates of the standard deviations (not shown) reveal that many changes are significant.

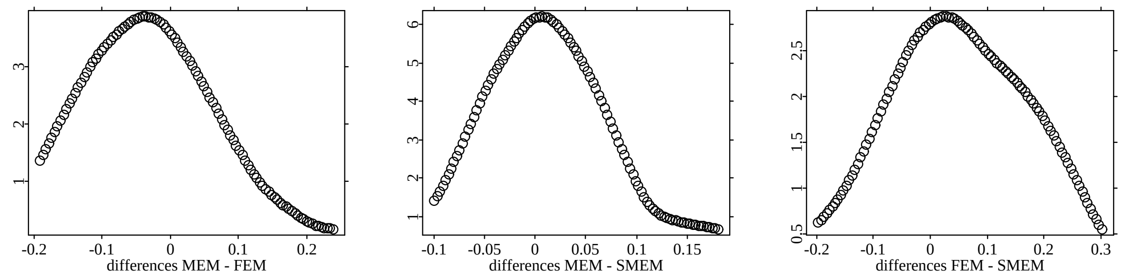

Figure 7: Densities of the differences $\widehat{\tilde{Y}_{d\bullet}}(MEM) - \widehat{\tilde{Y}_{d\bullet}}(FEM)$, $\widehat{\tilde{Y}_{d\bullet}}(MEM) - \widehat{\tilde{Y}_{d\bullet}}(SMEM)$, and $\widehat{\tilde{Y}_{d\bullet}}(FEM) - \widehat{\tilde{Y}_{d\bullet}}(SMEM)$.

This clear evidence that the estimates obtained depend strongly on the model used.

This matters a great deal because, among other things, political decisions, are based on the estimates $\tilde{Y}_{d\bullet}$. As illustrated here, such estimates can be misleading (due to misspecification and endogeneity) if they are based on simple mixed effects models, or very imprecise, if based on simple fixed effects models. We have skipped here the changes in $\hat{\beta}$ for brevity; it is of less interest in this particular study, and our findings already highlight that the model estimates change indeed notably.

Finally, we examine the coefficient estimates, together with the bootstrap estimates of the standard errors, see Table 6. In the bootstrap we used $g = 1.1h$ as the pilot bandwidth for the pre-estimation, and 400 bootstrap replications. Recall that the bootstrap for the MEM is inconsistent here, so that the standard errors for the MEM are underestimated for MEMs. For $\hat{\boldsymbol{\beta}}$ we can hardly find significant differences between the three models in this application.

	FEM		*SMEM*		*MEM*	
	$\hat{\beta}$	S.E.	$\hat{\beta}$	S.E.	$\hat{\beta}$	S.E.
sex	-.0284	.0447	-.0242	.0448	-.0327	.0409
age1	.2428	.1199	.2249	.1188	.2271	.1017
age2	.2665	.1080	.1978	.0966	.2145	.0879
single	-.0402	.0613	-.0745	.0568	-.0543	.0526
child	.0003	.0565	-.0166	.0501	-.0164	.0486
ngal	.2288	.0560	.2377	.0565	.2474	.0471
educ	.0648	.0487	.0481	.0454	.0517	.0410
stud	-.2219	.1011	-.2212	.0963	-.2312	.0829
self	.0809	.0786	.1288	.0740	.1131	.0721
pilgr	-.7004	.0910	-.6926	.0752	-.6918	.0683
family	-.1798	.0544	-.1478	.0487	-.1689	.0443
stay	-.0047	.0014	-.0045	.0014	-.0044	.0011
$\hat{\sigma}_u^2$			.0297	.0171	.0650	.0123

Table 6: Coefficients estimates with their bootstrap standard errors.

For the FEM we can compare the results with a parametric orthogonal least squares regression to check the robustness of our implemented method. Note that the coefficient estimates coincided perfectly with the numbers given in Table 6, with an $R^2 = .3908$. More interesting is to compare the parametric estimates of the standard errors with our bootstrap estimates. Although they deviate slightly (numerically) from each other, see Table 7, we conclude that the bootstrap approximation works reasonable well (as we did in the simulation part). Nevertheless, it must be kept in mind that the bootstrap estimates are model based, and can easily mislead, as they typically do in MEMs.

	sex	age1	age2	single	child	ngal	educ	stud	self	pilgr	family	stay
O	.0461	.1207	.1022	.0623	.0579	.0591	.0510	.0989	.0812	.0893	.0564	.0015
B	.0447	.1199	.1080	.0613	.0565	.0560	.0487	.1011	.0786	.0910	.0544	.0014

Table 7: For FEM, the parametric estimates (parametric orthogonal least squares regression) of the standard errors (**O**) and the bootstrap approximation (**B**).

6 Conclusions

We have introduced a new class of semi-mixed effects models that combines fixed effects, mixed effects and partial linear models. Nesting these models it can benefit from the advantages each model offers, and at the same time mitigate or even avoid its shortcomings. Our SMEM allows for a smooth transition from FEM to MEM, i.e. our class contains the continuum between them including also the PLM. Under the wrong assumption of independence the model is estimated with a serious bias in the MEM, and the variance of the estimates is larger than that in our semiparametric alternative. Moreover, we do not only offer consistent estimators, but also outperform the nested models FEM, PLM and MEM by construction. That this holds also true for finite samples, is exactly the strength of the proposed class as it demonstrates that we have successfully combined the advantages of these models to find a compromise that avoids the pitfalls of each extreme.

Further, although the construction of the MEM would favor it's performance in calculating macro or area parameters (rather than in estimating individual effects), the simulations show that SMEM is superior in terms of both out-of-sample and in-sample prediction. It is clear that SMEM is always better for consistent estimation and modeling with respect to interpretability. The example of analyzing tourist expenditures underpins this finding.

Finally, a consistent bootstrap arms us with a valid and feasible procedure to do statistical inference. FEM and PLM based bootstrap will suffer from a large variance in practice, whereas the SMEM is consistent and has small variance. In contrast, applying bootstrap in MEM when the independence assumption is violated is inconsistent as it is based on a wrong model and therefore leads to wrong conclusions.

References

Adebayo, S.B. and Fahrmeir, L. (2005). Analysing Child Mortality in Nigeria with Geoadditive Discrete-time Survival Models. *Statistics in Medicine*, **24**, 709-728.

Diggle, P.J., Heagerty, P., Liang, K.-L., and Zeger, S. (2002). *Analysis of Longitudinal Data*, 2nd ed. Oxford University Press.

Elbers, C., Lanjouw, J.O., and Lanjouw, P. (2003). Micro-level Estimation of Poverty and Inequality. *Econometrica*, **71**, 355-364.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models.* Springer Series in Statistics, 2nd edition, Springer-Verlag, New York.

Fay, R.E. and Herriot, R.A. (1979). Estimates of Income for Small Places. An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association*, **74**, 269-277.

Ghosh, M., Nangia, N., and Kim, D.H. (1996). Estimation of Median Income of Four-Person Families: A Bayesian Time Series Approach. *Journal of the American Statistical Association*, **91**, 1423-1431.

Goldstein, H. (2003). *Multilevel Statistical Models*, London: Edward Arnold.

Hall, P. and Maiti, T. (2006). Nonparametric estimation of mean-squared prediction error in nested-error regression models. *Annals of Statistics*, **34**, 1733-1750.

Hamilton, J.D. (2001). A Parametric Approach to Flexible Nonlinear Inference. *Econometrica*, **69**, 537-573.

Härdle, W. and Marron, J.S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Annals of Statistics*, **19**, 778-796.

Jiang, J. and Lahiri, P. (2006). Mixed Model Prediction and Small Area Estimation *Test*, **15**, 1-96.

Kneib, T. and Fahrmeir, L. (2006). Structured Additive Regression for Multicategorical Space-Time Data: a Mixed Model Approach. *Biometrics*, **62**, 109-118.

LIN, X. AND CARROLL, R. (2001). Semiparametric Regression for Clustered Data Using Generalized Estimating Equations. *Journal of the American Statistical Association*, **96**, 1045-1056.

LOMBARDÍA, M.J. AND SPERLICH, S. (2008). Semiparametric Inference in Generalized Mixed Effects Models. *Journal of the Royal Statistical Society, Series B*, **70**, 913-930.

MAITY, A., MA, Y., AND CARROLL, R. (2007). Efficient Estimation of Population-level Summaries in General Semiparametric Regression Models with Missing Response. *Journal of the American Statistical Association*, **102**, 123-139.

OPSOMER, J., CLAESKENS, G., RANALLI, M.G., KAUERMANN, G., AND BREIDT, F.J. (2008). Non-parametric Small Area Estimation Using Penalized Spline Regression. *Journal of the Royal Statistical Society, Series B*, **70**, 265-286.

PRASAD, N.G.N. AND RAO, J.N.K. (1990). The Estimation of the Mean Squared Error of Small-Area Estimators. *Journal of the American Statistical Association*, **85**, 163-171.

RAO, J.N.K. (2003). *Small Area Estimation.* John Wiley and Sons, Inc., New-York.

TUTZ, G. (2001). Generalized Semiparametrically Structured Mixed Models. *Computational Statistics & Data Analysis*, **46**, 777-800.

VILAR FERNÁNDEZ, J.M. AND FRANCISCO FERNÁNDEZ, M. (2002). Local Polynomial Regression Smoothers with AR-error Structure. *Test*, **11**, 439-464.