SMOOTH TRANSITION FROM FIXED EFFECTS TO MIXED EFFECTS MODELS IN MULTI-LEVEL REGRESSION MODELS

MARÍA JOSÉ LOMBARDÍA
STEFAN SPERLICH

FUNDACIÓN DE LAS CAJAS DE AHORROS
DOCUMENTO DE TRABAJO
Nº 374/2008

De conformidad con la base quinta de la convocatoria del Programa de Estímulo a la Investigación, este trabajo ha sido sometido a evaluación externa anónima de especialistas cualificados a fin de contrastar su nivel técnico.

Smooth Transition from Fixed Effects to Mixed Effects Models in Multi-level regression Models

María José Lombardía
Departamento de Estadística e Investgación Operativa
Universidad de Santiago de Compostela
15782 Santiago de Compostela, SPAIN

Stefan Sperlich
Georg-August Universität Göttingen
Institut für Statistik und Ökonometrie
Platz der Göttinger Sieben 5
D - 37073 Göttingen

Abstract

We introduce a semiparametric extension of multi-level regression models that includes mixed and fixed effects models as its two extreme cases. In some practical cases, one could consider the fixed effects model as an over parametrized model without modeling but just plugging in dummies. In other words, it suffers from "too many parameters but too little model". The mixed effects model tries to overcome this by using just random effects and therefore has "too few parameters but too much model", where "too much model" refers to the necessary assumptions made. We propose including a nonparametric term that allows the practitioner to shift the model smoothly between these extremes, depending on its data and underlying problem. Thereby, the smoothing parameter serves as its switcher. We will show that so we can filter out possible dependency between covariates and random effects. We further provide consistent bootstrap procedures for possible inference and to analyze prediction power. The positive implications of using this model are highlighted in particular for small area statistics and econometrics. This is underlined by simulation studies and a real data application.[1]

Keywords and Phrases: Mixed effects models, semiparametric multilevel models, small area statistics, bootstrap inference, partial linear models.

[1] The authors gratefully acknowledge the financial supported of the Spanish "Direccíón General de Investigación del Ministerio de Ciencia y Tecnología", MTM2005-00820, FUNCAS, and of the Xunta de Galicia PGIDIT06PXIB207009R.

1 Introduction and Motivation

Mixed effects models are quite popular in many areas of statistics with panel data analysis being maybe the most popular one. More recently they have also attracted a good amount of attention in small area statistics, see Ghosh and Rao (1994), Rao (2003), Jiang and Lahiri (2006) for reviews of small-area techniques. The same holds still for the analysis of any kind of panel data, see e.g. Laird and Ware (1982) and Diggle, Heagerty, Liang and Zeger (2002) as typical examples. We can find these methods now equally well in biomedical, forestal or agricultural, economic and social science studies. Although the different research areas favor different names like small area statistics, multi-level (regression) models or simply mixed effects models, the statistical problems of modelling, estimation and testing are basically the same; the differences arise mainly in the subsequent inferences. For example, in panel data analysis they are often just a remedy to account for the heterogeneity over the cross sectional samples when the data is short in time; in small area statistics they shall improve the prediction of area level parameters, or in econometrics improve the prediction of macro indices from micro-data; and finally they have become popular in economics for doing data matching, i.e. to impute a certain factor for the individuals in the sample of interest with the aid of a different (auxiliary) sample.

More recently now, they have entered the nonparametric world, see Ruppert, Wand and Carroll (2003), Wand (2003), Hamilton (2001), Opsomer, Claeskens, Ranalli, Kauermann and Breidt (2005), Tutz (2001), and Tutz and Reithinger (2007). In applied statistics, semiparametric Bayesian approaches are often used in combination with (penalized) splines, series or MCMC (random field) estimators, see Fahrmeir, Kneib, and Lang (2004), Kneib and Fahrmeir (2006) and Fahrmeir and Lang (2001) among others. In maybe most of the literature on semiparametric models, the idea has always been to separate the nonparametric function into a deterministic (fixed effects) and a random part (random effects) so that the smoothing parameter of a spline estimator could be written in terms of the variances of the random effects and the error term. Considerations of additional random effects have mainly been limited to Bayesian approaches and/or computational questions. However, until recently, asymptotic theory has been missing for estimation in semiparametric mixed models. Lombardía and Sperlich (2006) introduced an estimation procedure for generalized partial linear mixed effects models, a semiparametric model specification test with a bootstrap procedure, and provided asymptotic theory for all these methods.

For response $Y_d \in \Re$ with covariates $\boldsymbol{X}_{dj} \in \Re^p$ including the intercept, the classical generalized linear Mixed Effects Model (MEM) with known link g(·) can be written as

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

with $\boldsymbol{Z}_{dj} \subseteq \boldsymbol{X}_{dj}$ of dimension ρ, $\beta \in \Re^p$ the fixed effect, and $\boldsymbol{u}_d \in \Re^\rho$ the i.i.d. unobservable random effect with mean zero and unknown variances-covariance matrix σ_u^2. This has to be estimated. We have sample size $n = \sum_{d=1}^{D} n_d$, where D is the number of areas (domains or groups) with the typical assumption of $D \to \infty$ at rate $O(n)$. An essential but crucial assumption for the existing ethodology is that $\boldsymbol{u}_d$ is independent from $\boldsymbol{X}_{dj}$ and that $g(\cdot)$ is known. Note that, if g is the identity, model (1) includes the nested-error ($\boldsymbol{Z}_{dj} = 1$ and $\boldsymbol{u}_d \in \Re$), the random regression coefficient ($\boldsymbol{Z}_{dj} = \boldsymbol{X}_{dj}$), and the Fay-Herriot model (only area specific information), see Prasad & Rao (1990) for a summary. Let us concentrate on the nested error model, which is very popular in practice.

For this response and covariates, the corresponding Fixed Effects Model (FEM) is

$$E\left[Y_{dj}|\boldsymbol{X}_{dj}\right] \;=\; g\left\{\boldsymbol{X}_{dj}^{t}\boldsymbol{\beta}+c_{d}\right\}, \quad d=1,\ldots,D;\; j=1,\ldots,n_{d}, \tag{2}$$

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

For a better understanding of the main idea, recall first two striking facts. The MEM is often motivated as follows: for not over parameterizing one models the area effect by random effects. This seems to beat any other parametric model when predicting. This is often due to the fact that in the moment of prediction one adds the predicted random effect to the total prediction. The additional variance of the prediction caused by assuming this effect to be random, is only slightly larger than the variance of a fixed effect estimate based on small samples. Additionally, the modelling of the new variance structure allows for a more efficient estimation of $\boldsymbol{\beta}$ using a feasible generalized least squares estimator. However, this gain can easily be a self-deception because it might improve prediction in the mean but under the quite unrealistic assumption of independence between area effects and the covariates as well as the individual (i.e. error) effects. Thus, even if MEM leads to a better sample fit, it does at the cost of biased estimates. Furthermore, it does not at all contribute to a better understanding of the underlying process. Finally, a method to do valid inference is not available. Indeed, all available methods for testing or prediction intervals are clearly inconsistent if the assumption of independence is violated.

We studied about 25 applications from the mixed-models literature where the random effect $\boldsymbol{u}_d$ 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 little credible. Clearly, this causes endogeneity giving inconsistent estimates for $\boldsymbol{\beta}$ and awful performance for out of sample predictors. In other words, it does not satisfy neither the needs of econometric (or other) modelling nor of small area prediction.

We therefore propose to use a flexible modelling of area effects that allows the practitioner (or its model, respectively) to move smoothly from a MEM (1) without area specific covariates over a Semiparametric Mixed Effects Model (SMEM) with area specific effects (3) to a FEM (2). In order to illustrate the ideas and comments before said, we will study an example with interest in economy and small area prediction. We consider the tourist expenditures in each county of Galicia (which is divided in 53 counties). Galicia is a region in the Northwest of Spain, and as with the rest of the country, tourism is one of the most important economic factors. This survey of tourism contains information about at least 10 tourists per county including average expenditure per day and several characteristics of the individuals. Note that our dependent variable *expenditure* means total expenditure including accommodation, food, purchases, travel, leisure activities, and other miscellaneous.

The rest of the paper is organized as follows. In Section 2 we develop the flexible modelling of area effects. The smooth transition is achieved by relaxing more and more the smoothness assumption on the semiparametric (area specific) impact: we start with the highest smoothness (a constant) giving a random effects model, and end up in no smoothness (interpolation of the area effects) giving a fixed effects model. This way we resolve all problems at once: we model and thus explain the area or group effect, can dispose of the "independence-assumption" problem, this way obtaining consistent estimates and valid inference; all this without losing the advantages of MEM and without running into the problems we would face in a FEM. Our method can be seen as an extension of Partial Linear Models (PLM); including random effects we are able to more efficiently estimate both the parametric and the nonparametric part. We will see why this approach can always be used to filter out possible dependence between the individual covariates and the random effects. Additionally, Lombardía and Sperlich (2006) provide methods of statistical inference which can be applied to this model. We introduce and compare the different models together with possible estimation and bootstrap procedures. We also discuss model extensions and how we overcome the (in)dependence problem. In Section 3 we illustrate with a simulation study the use of our method highlighting also the arguments which motivate our modelling idea. Finally, in Section 4 we come back to the real data example.

2 The Semiparametric Transition Model

2.1 The Model

In the following, area (specific) effects refer to random and/or fixed effects, respectively. We start from model (1) considering the nested-error model (i.e. $\boldsymbol{Z}_{dj} = 1$ and $\boldsymbol{u}_d \in \Re$). Let now $\boldsymbol{W} \in \Re^q$ denote the specific area covariates. Then, if any information $\boldsymbol{W}$ about the areas is available, we suggest using the following semiparametric mixed effects model (SMEM):

$$E\left[Y_{dj}|\boldsymbol{X}_{dj}, \boldsymbol{W}_d, u_d\right] \;=\; g\left\{\boldsymbol{X}_{dj}^t\boldsymbol{\beta} + \eta_v(\boldsymbol{W}_d) + u_d\right\} \;, \qquad (3)$$

where $\eta_v : \Re^q \to \Re$ is an unknown nonparametric function with a given a "switcher" v. We introduce the condition $E[\eta_v(\boldsymbol{W})] = 0$ for identification. Let us think of v as a smoothness parameter so that, for example, for kernel estimates we set the bandwidth to $h = v \cdot n^{-2/(4+q)}$. Then, for one extreme we have $\eta_0(\boldsymbol{W_d}) = c_d$ with $\sum_{d=1}^{D} c_d = 0$ and on the other extreme $\eta_\infty(\boldsymbol{W}_l) = 0$. In the first case η_0 catches the area effect completely so that we get $u_d = 0$ for all d and we indeed have a FEM (2), whereas for $h = \infty$ we obtain a MEM (1). Then, v is indeed a switcher between the models:

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

and for $0 < v < \infty$

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

For ease of presentation we set $\boldsymbol{\theta}$ for the variance components and $\boldsymbol{\delta} = (\boldsymbol{\beta}, \boldsymbol{\theta})$ for the unknown parameters of the model. Set $Var[Y_{dj}| \boldsymbol{X}_{dj}, W_d, u_d] = \sigma_e^2$ and $\sigma_u^2 = Var[u_d]$ for all $d = 1,\ldots,D$, $j = 1,\ldots, n_d$, and $\boldsymbol{\theta} = (\sigma_e^2, \sigma_u^2)$.

Note that this is not a data adaptive approach in the common sense because the smoothness is given by the practitioner. On the other hand, if for example inference or prediction intervals are the objective, to fix v in advance is not a restriction in practice, as for pre-estimated smoothness parameter valid inference is unfeasible anyway. Certainly, bandwidth selection procedures for kernel smoother are thinkable also for model (3) but would follow an idea of modelling different from what we focus on.

In practice, often some of the $\boldsymbol{X}$ are not independent of the area effects. Imagine they are correlated with some of the $\boldsymbol{W}$. It is clear then that estimation and prediction from model (1) gives biased results whereas those from model (3) will be unbiased. One may argue that, in order to handle this problem, model

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

would work as well.

Unfortunately, the dependence structure between $\boldsymbol{X}$ and the area is usually much more complex and not limited to a simple linear relation with the available area information $\boldsymbol{W}$. However, in practice the relation can indeed always be described sufficiently via $\psi(\boldsymbol{W})$ for a particular, though unknown function ψ as long as $\boldsymbol{W}$ varies continuously over the different areas. For illustration, imagine the relation between $\boldsymbol{X}_{dj}$ and the area d is summarized in $\boldsymbol{J}_d$, 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 an artificial function ψ so that $\psi(\boldsymbol{W_d}) = \boldsymbol{J}_d + \boldsymbol{V}_d$ with $\boldsymbol{V}_d$ defined implicitly as the residual. A particular boundary case is $\boldsymbol{W_d} \equiv \boldsymbol{J}_d$ with ψ the identity. Another particular is when ψ simply assigns $\boldsymbol{J}_d$ to $\boldsymbol{W_d}$ for each $d = 1,\ldots,D$, i.e. interpolation. Such an assignment is possible as long as $\boldsymbol{W}$ moves continuously over the D areas. Recall that η_v is a nonparametric function with appropriate smoothness v. Then, for an implicitly defined φ_ω we get

$$\begin{aligned}
E\left[Y_{dj}|\boldsymbol{X}_{dj},\boldsymbol{W}_d,u_d\right] &= g\left[\boldsymbol{X}_{dj}^t\boldsymbol{\beta}+\eta_v(\boldsymbol{W}_d)+u_d\right] \\
&= g\left[\boldsymbol{X}_{dj}^t\boldsymbol{\beta}+\varphi_\omega\left\{\psi(\boldsymbol{W}_d)\right\}+u_d\right] \\
&= g\left[\boldsymbol{X}_{dj}^t\boldsymbol{\beta}+\varphi_\omega\left\{\boldsymbol{J}_d+\boldsymbol{V}_d\right\}+u_d\right]\ ,
\end{aligned} \tag{5}$$

where φ_ω is again a nonparametric function with a smoothness parameter ω which depends on v and the smoothness of ψ or, vice verse, v depends on ω and ψ. From (5) we see clearly that this model does not suffer from dependency between $\boldsymbol{X}_{dj}$ and u_d , i.e. endogeneity of $\boldsymbol{X}_{dj}$. Consequently, in practice where we only face finite samples, η_v can perfectly filter out the endogeneity.

Now it is clear why in the SMEM we can always filter out possible dependence between the covariates and the random effect: it just depends on the choice of $\boldsymbol{W}$ and v. In econometric words, $\psi(\boldsymbol{W}_d)$ can serve here as a proxy. For example, in the "worst case", $\boldsymbol{W} \perp \boldsymbol{J}$, we need to set $v = 0$ as ψ cannot feature any smoothness. What then actually happens is that the SMEM becomes a FEM without random effect and thus without any independence problem.

2.2 Estimation

There exist plenty of estimation procedures for MEM, and more recently some for mixed effects models allowing for a semiparametric impact of $\boldsymbol{X}$ when g is the identity. Most of them are Bayesian methods, some combined with penalized splines, some with MCMC methods, see references in the Introduction. For the estimation of our model (3) many alternatives are possible. Lombardía and Sperlich (2006) propose a smoothed maximum likelihood based on the so called integral approach and describe also the so called penalized quasi likelihood method for maximizing the posterior mode of density $f(\boldsymbol{\beta}, \boldsymbol{u}|Y, \boldsymbol{W}, \boldsymbol{X}\,;\, \eta, \boldsymbol{\theta})$, see for example Breslow and Clayton (1993). The first method is easier to tackle from a conceptual point of view (asymptotic theory is straight forward for this one), whereas the second one is more popular in practice. Note that the implementation is combined with ideas of Vilar-Fernández and Francisco-Fernández (2002). When we apply the integral approach we concentrate on the following marginal density

$$f(Y_{dj}|\boldsymbol{W}_d, \boldsymbol{X}_{dj}; \eta, \boldsymbol{\delta}) \;=\; \int f(Y_{dj}|\boldsymbol{u}, \boldsymbol{W}_d, \boldsymbol{X}_{dj}; \eta, \boldsymbol{\beta}, \sigma_e^2) p(\boldsymbol{u}; \boldsymbol{\sigma}_u) d\boldsymbol{u}. \tag{6}$$

To estimate the parametric part we take directly the logarithm and get the (unsmoothed) likelihood

$$l(\boldsymbol{Y}; \eta, \boldsymbol{\delta}) \;=\; \sum_{d=1}^{D}\sum_{j=1}^{n_d} \log f(Y_{dj}|\boldsymbol{W}_d, \boldsymbol{X}_{dj}; \eta, \boldsymbol{\delta}) \ . \tag{7}$$

For estimating the nonparametric function, fix a point $\boldsymbol{w}_0$ and construct an empirical counterpart of $E[\log f(Y \mid \boldsymbol{W}, \boldsymbol{X}; \eta, \boldsymbol{\delta})| \boldsymbol{W} = \boldsymbol{w}_0]$ which, in terms of kernel function $K_{\boldsymbol{h}}(\cdot)$, could be

$$l_s(\boldsymbol{Y}; \eta, \boldsymbol{\delta}) = \sum_{d=1}^{D}\sum_{j=1}^{n_d} K_{\boldsymbol{h}}\left(\boldsymbol{w}_0 - \boldsymbol{W}_d\right) \log f(Y_{dj}|\boldsymbol{W}_d, \boldsymbol{X}_{dj}; \eta(\boldsymbol{w}_0), \boldsymbol{\delta}) \ . \tag{8}$$

Here, $K_{\boldsymbol{h}}(\cdot)$, is a q-dimensional product kernel, and $\boldsymbol{h} = (h_1,...,h_q)$ the corresponding bandwidth vector. Often, l_s is called the smoothed likelihood function.

Applying now (twice - as the variance components are unknown, too) the idea of profiled likelihood estimation one gets from Lombardía and Sperlich (2006), compare also with Lin and Carroll (2006), that under some rather common smoothness conditions we have

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

b) defining $h_{\text{prod}} = \prod_{j=1}^{q} h_j$ 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 $Var_\eta(\boldsymbol{w}_0) = \dfrac{\int K(\boldsymbol{w})^2 d\boldsymbol{w}}{p_W(\boldsymbol{w}_0)\ E\left[\frac{\partial}{\partial \eta_v} l\left(Y; \eta_v, \boldsymbol{\delta}_0\right)^2 | \boldsymbol{W} = \boldsymbol{w}_0\right]}$.

From Maity, Ma, and Carroll (2007) we even conclude that using these estimates will produce efficient predictors for forecasting area-specific means.

2.3 Extensions

When $\boldsymbol{W}$ is of higher dimension ($q > 3$), in practice one usually would like to break down the dimension by modelling η_v in a separable, maybe additive way

$$\eta_v(\boldsymbol{W}_d) = \sum_{k=1}^{q} \eta_v^k(W_{d,k}), \qquad \eta_v^k : I\!R \to I\!R \quad \forall\ k \tag{9}$$

or as a single index function $\eta_v\ (\boldsymbol{W}_d^t \boldsymbol{\delta})$, $\boldsymbol{\delta} \in \Re^q$ an unknown parameter vector. These extensions are mostly straightforward (concerning both implementation and asymptotic theory) when using the approach of Härdle, Huet, Mammen, and Sperlich (2004), respectively of Carroll, Fan, Gijbels, and Wand (1997).

In this context, we would like to recall the discussion in the Introduction. There, we referred e.g. to Fahrmeir et al. (2001,2004) and Tutz (2001) which combined Bayesian approaches with splines to estimate nonparametric additive mixed effects models in different applied problems. Although one could implement estimators for the above introduced models also with P-splines and additivity as in (9), we have disregard that possibility in this article for two reasons. Firstly, asymptotic theory for P-splines is not well developed, and especially not for $q > 1$. Secondly, while for kernel based estimators it is clear how to put into practice $v \to 0$, for P-splines we have to move 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 about the proper proportion of λ and K. In practice this makes a smooth transition from FEM to MEM much less convenient than it is for kernels.

Another extension is to allow for a nonparametric impact of $\boldsymbol{X}$, i.e.

$$E\left[Y_{dj}|\boldsymbol{X}_{dj}, \boldsymbol{W}_d, u_d\right] \;=\; g\left\{m(\boldsymbol{X}_{dj}) + \eta_v(\boldsymbol{W}_d) + u_d\right\} \;, \tag{10}$$

where $m(\cdot)$ is a nonparametric unknown function. In practice, one could model $m(\cdot)$ additively as a sum of p one-dimensional nonparametric functions, maybe with interactions, compare Sperlich, Tjøstheim, and Yang (2002). In case one would model both m and η_v additively we get a generalized additive mixed effects model. For g = identity, i.e. with $\boldsymbol{X}_{dj} = (X_{dj,1}, \ldots, X_{dj,p})^t$, $\boldsymbol{W}_d = (W_{d,1}, \ldots, W_{d,q})^t$, m^l and η_v^k being one dimensional nonparametric functions for all $l = 1, \ldots, p$ and $k = 1, \ldots, q$;

$$E\left[Y_{dj}|\boldsymbol{X}_{dj}, \boldsymbol{W}_d, u_d\right] \;=\; \sum_{l=1}^{p} m^l(X_{dj}^l) + \sum_{k=1}^{q} \eta_v^k(W_d^k) + u_d. \tag{11}$$

There exists a good set of applied research using Bayesian methods for the estimation, compare discussion in the Introduction. Quite recently, Roca Pardiñas and Sperlich (2007) introduced a weighted smooth backfitting estimator which can handle - at least from a computational point of view – the estimation of different kind of generalized structured models like (11) with m and η_v having additive, single index, interaction or other forms. For more semiparametric models to which extensions are thinkable, see Härdle, Müller, Sperlich, and Werwatz (2004).

2.4 Bootstrap Inference

We propose here a parametric bootstrap method for further inference. The algorithm is as follows. Assuming that the distribution of $Y_{dj}|(\boldsymbol{X}_{dj}, \boldsymbol{W}_d, u_d) \sim F(\mu_{dj}, \sigma_e^2)$ and the distribution of the random effects $u_d \sim G(0, \sigma_u^2)$ are both known, then:

Step 1: From sample, calculate the estimator $\hat{\boldsymbol{\delta}} = (\hat{\boldsymbol{\beta}}, \hat{\sigma}_u^2, \hat{\sigma}_e^2)$.

Step 2: Take a pilot bandwidth g, and calculate $\hat{\eta}_g(\cdot)$. This is done in order to try to reproduce the same bias incurred in the real world.

Step 3: Generate u_d^* from $G(0, \hat{\sigma}_u^2)$, $d = 1, \ldots, D$.

Step 4: Given the auxiliar variables $\boldsymbol{X}_{dj}$, $\boldsymbol{W}_d$ and the bootstrap random effect u_d^*, draw Y_{dj}^* from F($\mu_{dj}^*, \hat{\sigma}_e^2$), with $\mu_{dj}^* = g\{\boldsymbol{X}_{dj}^t\hat{\boldsymbol{\beta}} + \hat{\eta}_g(\boldsymbol{W}_d) + u_d^*\}$.

Step 5: From the bootstrap sample $\{Y_{dj}^{*}, \boldsymbol{X}_{dj}, \boldsymbol{W}_{d}\}$ and the estimation algorithm before, we can get the bootstrap estimators $\hat{\boldsymbol{\delta}}^{*} = (\hat{\boldsymbol{\beta}}^{*}, \sigma_{u}^{*2}, \sigma_{e}^{*2})$ and $\hat{\eta}_{v}^{*}$.

The consistency can be concluded from Lombardía and Sperlich (2006) where a similar procedure has been proposed for a different model. It is based on the consistency of the estimators of the model parameters and the behaviour of $\hat{\eta}_{g}$ (see Härdle and Marron, 1991). To the assessment of uncertainty about parameter values $\boldsymbol{\delta} = (\boldsymbol{\beta}, \sigma_{u}^{2}, \sigma_{e}^{2})$ and the nonparametric functions we can then derive confidence intervals and construct also prediction intervals.

Finally, note that the impact of η_v can be tested for significance with the statistics introduced in Lombardía and Sperlich (2006). It is easy to see that the asymptotic theory developed there carries over to our model.

3 Simulation Studies

In this section we show the results of a large simulation study, which studies all points made in the motivation of this work. We consider a data generating process such as the independence assumption of the MEM is violated. And in this context we compare the estimation of the model parameters in the MEM versus SMEM, the (expected) mean squared errors of $\hat{\eta}_{v}$ and $\hat{\boldsymbol{\beta}}$ resulting from SMEM and PLM, the prediction power of MEM versus SMEM for in-sample and out-of-sample individuals, and the prediction power for the area-level parameters. Finally, we check the proposed bootstrap procedure.

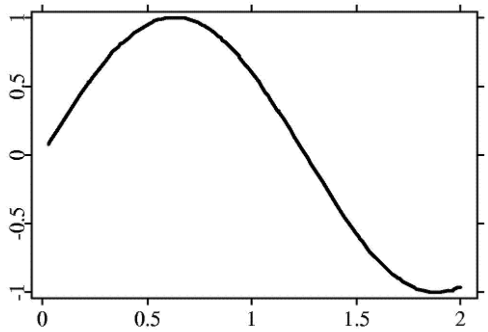
3.1 The Data Generating Process

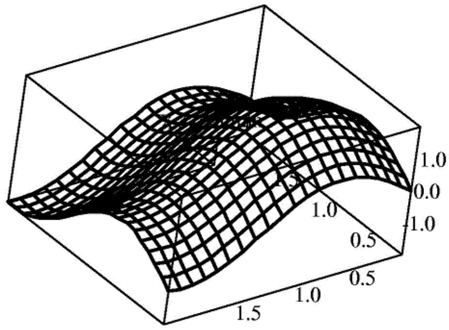
Consider the data generating process

$$Y_{dj} = \beta_0 + \boldsymbol{X}_{dj}^{t}\boldsymbol{\beta} + \sum_{k=1}^{q} \sin(2.5W_{d,k}) + u_d + \epsilon_{dj} \ , \tag{12}$$

where $(X_{dj,1}, X_{dj,2})^{t} = \boldsymbol{X}_{dj} \in \Re^{2}$. For all k is $W_{d,k} \sim U[0, 2]$ i.i.d., $u_d \sim N(0, \sigma_{u}^{2})$ i.i.d, and $\varepsilon \sim N(0, \sigma_{e}^{2})$ i.i.d. For $i = 1, 2$ we have created $X_{dj,i} = 0.8 \cdot O_{\mathrm{dj,i}} + 0.5 \sum_{k=1}^{q} W_{d,k}^{2}$ 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 will study the performance of our method for $q = 1$ and $q = 2$. 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 $q = 2$; $i, k = 1, 2$. In Figure 1 are plotted the impacts of the systematic area effect $\sum_{k=1}^{q} \sin(2.5W_{d,k})$ for $q = 1, 2$. Note that for the sake of illustration we have concentrated here on the canonical link for the normal distribution, i.e. g = *identity*.

Figure 1: The systematic area effect η for q=1 (left) and q=2 (right).



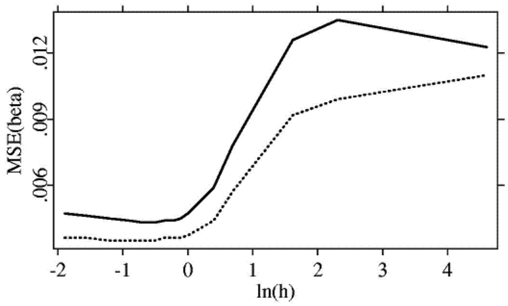


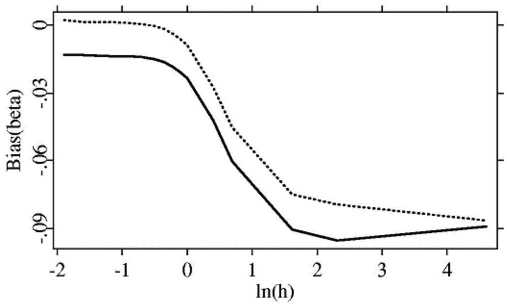
In the following we present simulation results for n = 250, d = 50, where data was generated from model (12) with different $\boldsymbol{\theta}$ = (σ_e^2, σ_u^2). All results shown here are based on 250 simulation runs. The procedures are implemented in FORTRAN90. Actually, the implementations of FMEM, MEM, SMEM and PLM are in a way so that they are nested algorithms to guarantee a fair comparison.

3.2 Results

Due to the data generating process (12) the independence assumption of the MEM is violated, therefore the estimator $\hat{\boldsymbol{\beta}}$ will be inconsistent, but it is not clear how the mean squared error (MSE) and bias will change with the bandwidth $0 \leq h \leq \infty$. Our first simulation has been performed to study MSE and bias of $\hat{\boldsymbol{\beta}}$ as a function of v, respectively ln(h), see Figure 2. The underlying data generating process is (12) with q = 1 and $\boldsymbol{\theta}$ = (0.5, 0.25). The figure may make one believe that v = 0, i.e. the FEM, would be a good choice. Notice that, this could be a delusion as this choice implicates the admissibility of a fixed effects model, a conclusion that can only be drawn from the practical context.

Figure 2: MSE and Bias of **β** estimates for β_1 and β_2 as functions of bandwidth h when q=1 and **θ**=(0,5,0.25).





Next, we compare the results of our SMEM

$$E\left[Y_{dj}|\boldsymbol{X}_{dj}, \boldsymbol{W}_d, u_d\right] = \beta_0 + \boldsymbol{X}_{dj}^t\boldsymbol{\beta} + \eta_v(\boldsymbol{W}_d) + u_d$$

for a fixed bandwidth h = 0.5, with the classical MEM ($h = \infty$ and $\sigma_u^2 \neq 0$)

$$E\left[Y_{dj}|\boldsymbol{X}_{dj}, u_d\right] = \beta_0 + \boldsymbol{X}_{dj}^t\boldsymbol{\beta} + u_d,$$

and with the PLM ($0 < h < \infty$ and $\sigma_u^2 = 0$)

$$E[Y_{dj}|\boldsymbol{X}_{dj}, \boldsymbol{W}_d] = \boldsymbol{X}_{dj}^t\boldsymbol{\beta} + \beta_0 + \eta_v(\boldsymbol{W}_d) .$$

The bias and variance for $\hat{\boldsymbol{\beta}}$ are compared in Table 1. Results are given for models with different $\boldsymbol{\theta}$ and q. We see that the bias can become quite serious when the independence assumption does not hold - something that happens quite often in practice, recall the discussion in the Introduction. At the same time the variance is not larger (even somewhat smaller) in the SMEM despite the nonparametric estimate applied in this model. Before studying this point more in detail recall that we make use of the variance estimates $\hat{\boldsymbol{\theta}}$ in both models, MEM and SMEM. Therefore, we also will have a look at the estimates of $\boldsymbol{\theta}$, see Table 2 where we give biases and mean squared errors (MSE) for σ_e^2 and σ_u^2. Note that $Var[\eta(\boldsymbol{W})] \approx 0.5067$ for $q = 1$, and 1.013 for $q = 2$. Here, a fair comparison is not really possible because in the MEM, part of the variation due to η will be assigned to σ_u^2 erroneously.

Table 1: Bias and variances of β-estimates. The SMES refers to h=0,5.

Model	σ_e^2	σ_u^2	$q=1$ bias($\hat\beta_1$)	bias($\hat\beta_2$)	var($\hat\beta_1$)	var($\hat\beta_2$)	$q=2$ bias($\hat\beta_1$)	bias($\hat\beta_2$)	var($\hat\beta_1$)	var($\hat\beta_2$)
	.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 2: Bias and MSEs of σ^2-estimates. The SMES refers to h=0,5.

Model	σ_e^2	σ_u^2	$q=1$ bias($\hat\sigma_e^2$)	bias($\hat\sigma_u^2$)	MSE($\hat\sigma_e^2$)	MSE($\hat\sigma_u^2$)	$q=2$ bias($\hat\sigma_e^2$)	bias($\hat\sigma_u^2$)	MSE($\hat\sigma_e^2$)	MSE($\hat\sigma_u^2$)
	.1	.0	.0003	.4470	.0001	.2038	-.0001	.8946	.0001	.8259
		.25	-.0003	.4574	.0001	.2256	-.0005	.9088	.0001	.8771
M		.5	-.0005	.4606	.0001	.2457	-.0008	.9161	.0001	.9209
E	.5	.0	.0177	.3174	.0027	.1087	.0257	.5807	.0042	.3687
M		.25	.0055	.3636	.0022	.1534	.0111	.6659	.0031	.4971
		.5	.0016	.3845	.0021	.1875	.0051	.7131	.0028	.5925
	.1	.0	-.0207	.0156	.0006	.0003	-.0272	.0220	.0010	.0005
S		.25	-.0036	-.0144	.0001	.0026	-.0062	-.0754	.0002	.0078
M		.5	-.0022	-.0390	.0001	.0104	-.0030	-.1691	.0001	.0361
E	.5	.0	-.0511	.0297	.0045	.0010	-.0762	.0244	.0089	.0007
M		.25	-.0150	-.0223	.0024	.0043	-.0235	-.0987	.0038	.0129
		.5	-.0097	-.0490	.0022	.0138	-.0112	-.1953	.0031	.0479

Although Vilar Fernández and Francisco Fernández (2002) considered a different context in their paper, it is clear from their theorems that our SMEM (3) will also be more efficient than common PLM estimators when estimating η_v (and, in our context, $\hat{\beta}$). For the numerical (not the asymptotic) performance this might even be true if the random effects are zero, since when estimating model (3), $\hat{\sigma}_u^2$ will correct for a possible over- or undersmoothing of the impact of $\mathbf{W}$ and vice versa. In Table 3 are compared the (expected) mean squared errors of $\hat{\eta}_v$ and $\hat{\beta}$ resulting from SMEM and PLM for h = 0.5. The *Expected Mean Squared Error* (EMSE) of $\hat{\eta}_v$ is defined by $E[\{\eta_v(\mathbf{W}) - \hat{\eta}_v(\mathbf{W})\}^2]$. The results are supporting our expectations: the SMEM clearly outperforms the PLM.

Table 3: EMSEs of and $\hat{\eta}_v$ and $\hat{\beta}$ for PLM and SMEM when the bandwidth is h=0.5.

Model	σ_e^2	σ_u^2	$q=1$ EMSE($\hat{\eta}_v$)	$q=1$ MSE($\hat{\beta}_1$)	$q=1$ MSE($\hat{\beta}_2$)	$q=2$ EMSE($\hat{\eta}_v$)	$q=2$ MSE($\hat{\beta}_1$)	$q=2$ MSE($\hat{\beta}_2$)
	.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

Another argument used in favor of MEM is their presumable prediction power, important for data matching when imputing factors for individuals and in small area statistics to predict area-, or say macro-, level parameters. Therefore, the next simulation study (see Table 4) compares the prediction power of MEM versus SMEM for all: in-sample prediction, out-of-sample prediction, for individuals, and for area-levels.

Here the used in-sample prediction risk measure is simply the average over the 250 simulations runs of the mean squared error, we denote this measure by ASE for *Averaged Squared Errors*

$$ASE = \frac{1}{250}\sum_{repl=1}^{250} MSE^{repl}$$

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 $Y_{dj} = \hat{\beta}^t \mathbf{X}_{dj} + \hat{\beta}_0 + \hat{\eta}_v(\mathbf{W}_d) + \hat{u}_d$ for the SMEM. Note that $\hat{\beta}$ and $\hat{u}$ are certainly different for the two models.

Now, our first out-of-sample risk is the *Mean Squared Prediction Error* for two particular $\mathbf{x}$, called $MSE(\hat{Y}_l, \hat{Y}_s)$ for $\mathbf{X}_l = \left(\frac{2q}{3} + 2.5, \frac{2q}{3} + 2.5\right)$, respectively $\mathbf{X}_s = \left(\frac{2q}{3} - 2.5, \frac{2q}{3} - 2.5\right)$, each in a different but fixed area.

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 a good job with respect to the mean squared error. In contrast, for small and moderate sample size nonparametric methods like we use them for the estimation of η in our SMEM can have an awful numerical performance. Nevertheless, the results in Table 4 show that our model clearly outperforms the MEM, surprisingly even in the ASE, and by far in the out-of-sample prediction. Note additionally that valid inference with the MEM is hardly possible for our data generating processes as the so far available methods are typically model based and therefore ``model biased".

Table 4: The average mean squared error (ASE) of the inside-sample predictors, and the $MSE(\hat{Y}_l, \hat{Y}_s)$ when SMEM is estimated with h=0.5.

Model			$q=1$			$q=2$		
	σ_e^2	σ_u^2	ASE	MSE$(\hat{Y}_l, \hat{Y}_s)$		ASE	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

When predicting area-level parameter, MEM is expected to perform reasonably well compared to SMEM regardless of possible violation of the independence assumption. In the next simulation study we predict the mean parameter for each area $d = 1,\dots,50$. Two slightly different parameters at the area level: (i) $\mu_d = E\left[\overline{Y}_{d\bullet} \mid X, W\right]$, this is assuming that the number of population units in the dth area is large; and (ii) $\overline{Y}_{d\bullet} = \sum_{j=1}^{N_d} y_{dj} / N_d$, assuming a superpopulation regression model of the form (12) for the N_d population units in the dth area. When considering (ii), the best linear unbiased estimator of $\overline{Y}_{d\bullet}$ is given by

$$\hat{\overline{Y}}_{d\bullet} = f_d \overline{y}_s + (1 - f_d)\hat{\mu}_{dr},$$

where $f_d = n_d / N_d$, $\overline{y}_s$ is the average of the in-sample values and $\hat{\mu}_{dr}$ is the estimator of μ_d with the mean of the $\mathbf{X}_{dj}$ for the $(N_d - n_d)$ non sampled units. Therefore we only concentrate here on the situation when we need to predict *Y* for some individuals of which *X* is available.

We performed two simulation runs, each with 250 replications. In both cases data was generated from the model (12) with q=2 but with 10 observations X_{dj} for each area whereas Y_{dj} was observed only for the first 5 individuals j in each area. In the first case all X_{dj} were randomly drawn (d = 1,...,50 and j = 1,...,10) and in the second we set

$$X_{d6} = (-1, -1),\ X_{d7} = (0.16, 0.16),\ X_{d8} = (1.33, 1.33), \qquad (13)$$
$$X_{d9} = (2.5, 2.5),\ \text{and}\ X_{d10} = (3.67, 3.67).$$

Note that (-1,0.16,1.33,2.5,3.67) are approximately

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

for each element X of $\mathbf{X}$, with σ_X denoting its standard deviation unconditionally from the area. The values Y_{dj} were generated with its corresponding X_{dj} (j = 1,...,5). As before, we show only results with bandwidth h = 0.5 throughout, although in simulation studies not shown here the SMEM with somewhat larger bandwidths $h \approx 0.7$ outperformed the MEM even more, depending on the model.

The results are given in form of box-plots which show the distributions of the D = 50 mean squared errors for different (σ_e^2, σ_u^2) in the data generating process. In each plot the box-1 and box-2 are refered to SMEM and MEM respectively, with X_{d6} to X_{d10} taken randomly; box-3 and box-4 are refered to SMEM and MEM respectively, with X_{d6} to X_{d10} as in (13). For better illustration we skipped the extreme large mean squared errors for the MEM (about 2 to 5% of the data).

Given the particular simulation model, the mean squared errors are generally quite small, especially when predicting (ii) $\overline{Y}_{d\bullet}$, where 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. When turning to the prediction of (i) $\mu_d = E[\overline{Y}_{d\bullet} \mid \mathbf{X}, \mathbf{W}]$, then the mean squared errors have to increase significantly, representing half in-sample and half out-of-sample prediction errors. We already have discussed and seen in Table 4, compare the ASE, that the advantage of the SMEM over the MEM for in-sample prediction can become fairly small though always visible. When now looking at the box-plots given in Figure 4, recalling that we always have to compare box-plot 1 with 2 and box-plot 3 with 4, the superiority of the SMEM over MEM even in this exercise is surprisingly strong. For the prediction of both parameters, $\overline{Y}_{d\bullet}$ and μ_d, we see that the most area effects can be captured by instruments $\mathbf{W}$.

Figure 3: Mean squared error distributions over the 50 predicted area level parameter $\overline{Y}_{d\bullet}$ for different data generating (σ_e^2, σ_u^2). Boxes 1 and 2 refer to MEM and SMEM respectively, with X_{d6} to X_{d10} random; Boxes 3 and 4 refer to MEM and SMEM respectively, with X_{d6} to X_{d10} fixed.

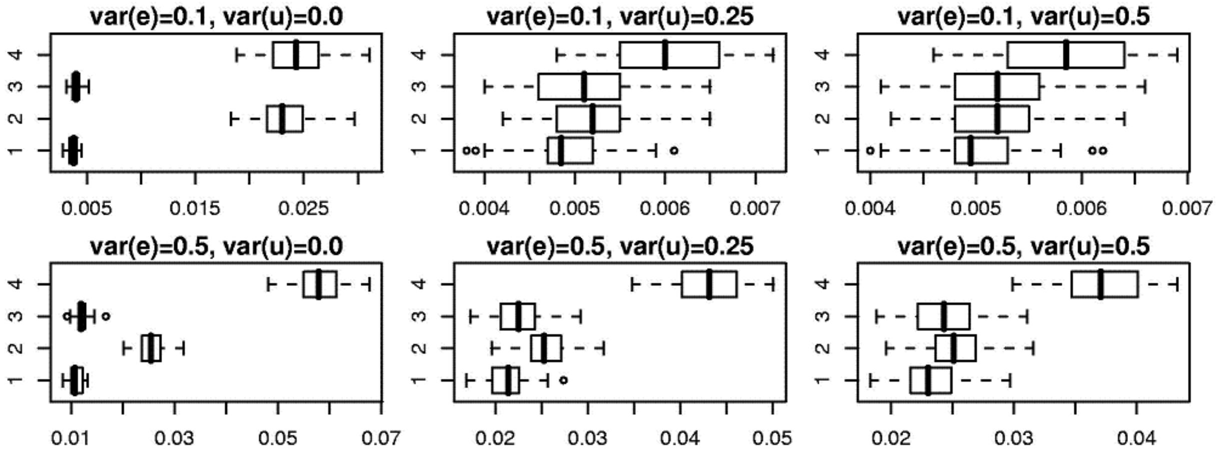
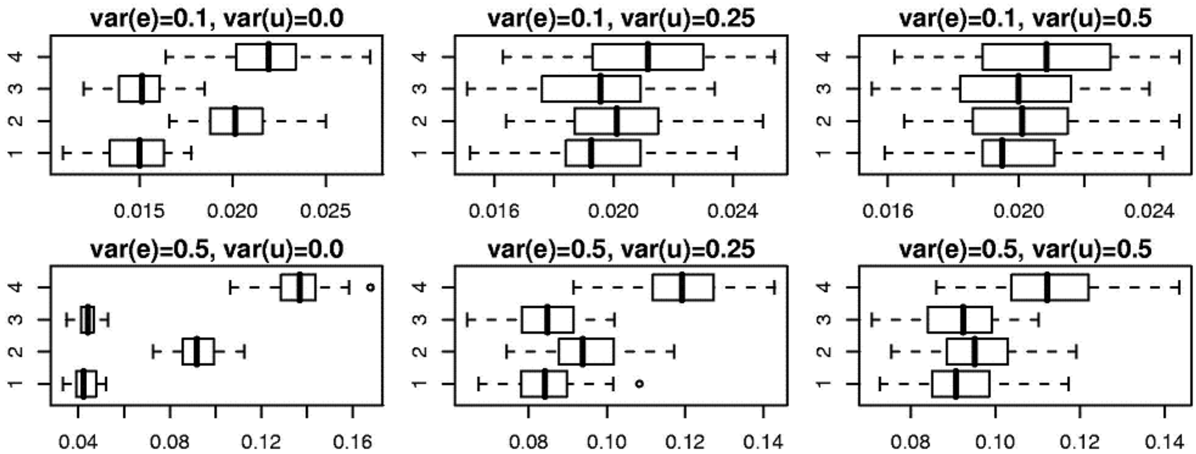


Figure 4: Mean squared error distributions over the 50 predicted area level parameter μ_d for different data generating (σ_e^2, σ_u^2). Boxes 1 and 2 refer to MEM and SMEM respectively, with X_{d6} to X_{d10} random; Boxes 3 and 4 refer to MEM and SMEM respectively, with X_{d6} to X_{d10} fixed.



Finally, we would like to check in a small simulation study whether the proposed bootstrap procedure works. We consider the data generating process as in Section 3.1 with $q = 1$. We have done only 100 simulation runs with 200 bootstrap replicates being aware that this will give only a rough approximation. The results are given in Table 5. Following the recommendation of Härdle and Marron (1991), the bootstrap model was constructed using a pilot bandwidth g greater than h; concretely, all bootstrap samples were generated with a bandwidth $g = 1.1\ h$. We also tried other values for g obtaining rather similar results. As can be seen, the results confirm that the bootstrap procedure can serve as a reasonable tool for doing inference in our SMEM. Further simulations not shown here revealed that the bootstrap does a quite good job for estimating the variance but sometimes, i.e. not for all cases, problems to catch the bias, a not surprising but rather expected finding. We therefore admit that further research on improving the bootstrap performance might be desirable, e.g. to give better bias estimates, but this is beyond the scope of this paper.

Table 5: Bootstrap approximations (B) of actual mean squared errors (O) for β, θ, and our individual predictors. Estimates and predictions are calculated in the SMEM, dimension q = 1, with bandwidth h = 0,5.

	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

3.3. Concluding Remarks

The simulation study has confirmed all the points made in the introduction and the motivation of our procedure. Firstly, our SMEM allows for a smooth transformation from FEM to MEM. Secondly, under the wrong assumption of independence the parameter vector β is estimated with a clear bias in the MEM. Also the variance of $\hat{\beta}$ is larger than the one in our semiparametric alternative. Similar statements certainly hold true for $\hat{\theta} = (\hat{\sigma}_e^2, \hat{\sigma}_u^2)$. Thirdly, our proposal also clearly outperforms FEM and PLM. However, recall that - asymptotically - such a conclusion is not surprising as our model nests FEM, PLM and MEM. This is exactly the strength of our innovation.

Fourthly, due to the nature and the construction of the MEM made rather for prediction of area parameters then for estimating individual effects, one might have expected strong biases for the model parameters, see Tables 1 and 2. However, the simulations show that even for prediction of area parameters the results are better for the SMEM in both out-of-sample and in-sample prediction.

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

4 A Real Data Example

4.1 Application to a Tourism Survey in Galicia

For Galicia like for the rest of Spain, tourism is one of the most important economic factors. 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, in particular the level of the so-called *comarcas* of which 53 exist in Galicia. Obviously, to receive reliable information about a tourists expenditure is cumbersome and expensive, and one is happy with observing, i.e. interviewing in detail maybe 10 individuals per comarca. A peculiarity of Galicia is the famous pilgrim trails to Santiago de Compostela, in particular the so-called *French trail*. For example, in the *holy year* of 2004 about 180000 pilgrims visited Galicia, and in particular

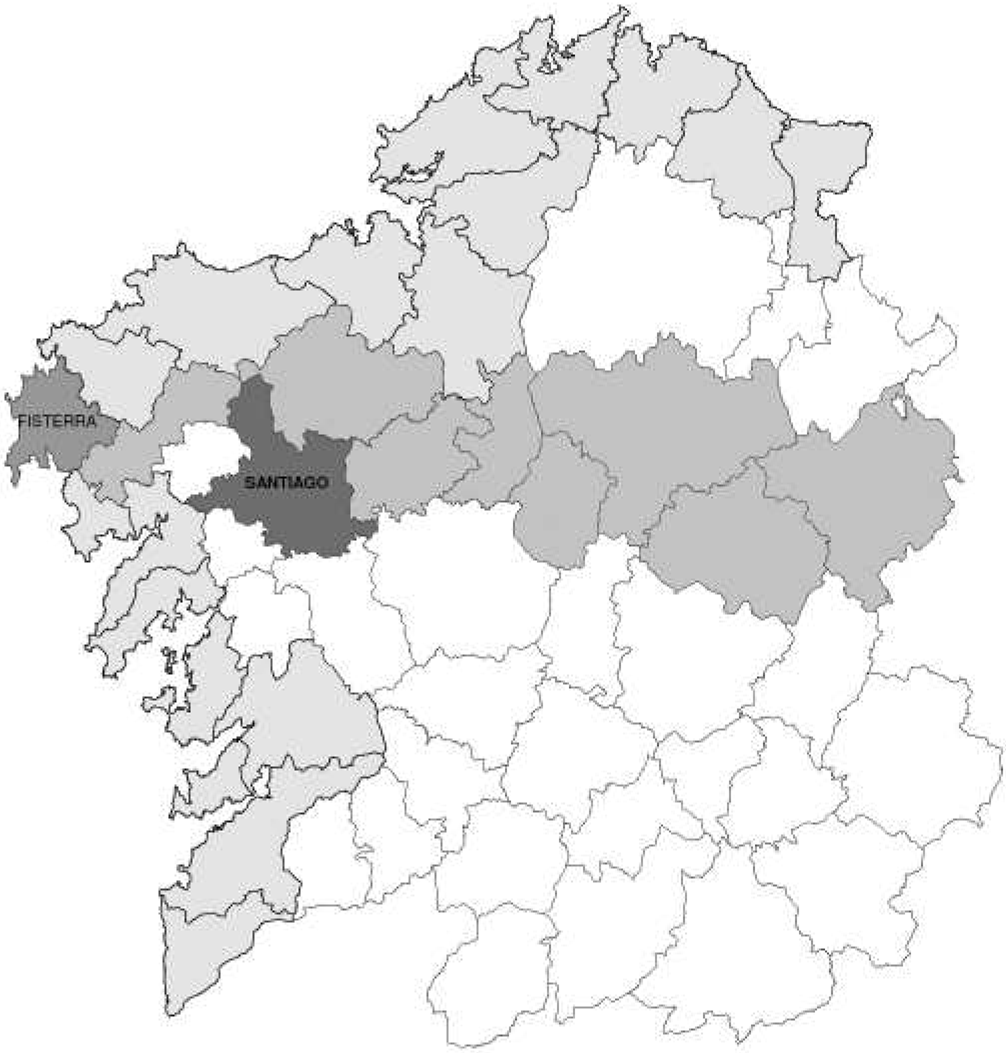
Santiago[2]. Another tourist attraction is certainly the coast. Figure 5 shows the map of the comarcas of Galicia; with all comarcas having coast in light grey and the comarcas that the French trail passes through in grey. The comarca with name *Fisterra* has both peculiarities, and *Santiago* is the pilgrim center and capital. Finally, rural tourism is growing in the sparsely populated areas.

We make use of a survey organized by the University of Santiago de Compostela in 2004. This survey contains information of 10 tourists per comarca including average expenditure per day and several characteristics of the individuals. Note that our dependent variable expenditure means total expenditure including accommodation, food, purchases, travel, leisure activities, and other miscellaneous. We have selected the set of variables described in Table 6.

Other variables of the comarcas which at first glimpse should be important like the index of tourism, the index of bars and restaurants, and the index of economic activity (all the three being monetary quantities) have been studied but finally excluded from the model. This has various reasons. Firstly, all three indices have two sources for endogeneity: measurement error and simultaneity. In some regions of Spain bars and restaurants claim sales of about the same amount tourists report to consume in bars and restaurants. So either the restaurants and bars under-report or the tourists over-report or residents do not consume in bars and restaurants. Similar problems occur for the economic activity, e.g. in Spain the construction branch has an important impact on the Gross Product but it is commonly believed that alone in this sector more than 30% of the real turnover is {paid cash in hand}, that is without VAT (value-added tax). The simultaneity is evident. The exclusion is also 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%). At the same time, this indicates that lpopd is a good instrument for the indices ``tourism" and ``bars and restaurants".

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.

[2] For details see statistics at www.santiago-today.com/santiago_article.cfm?art_id=302.



The inclusion of dummies in the nonparametric function η_v is not a problem but alternatively a partial linear modelling could be considered. However, we observed interaction between lpopd, ftrail so that we preferred to disregard that alternative. Next we performed a simple linear regression of each covariate of the comarcas on the individual characteristics to study the dependence structure. For lpobd we got $R^2 = .063$, for ftrail $R^2 = .179$, and for coast $R^2 = .076$. This indicates dependence so that the independence assumption necessary for MEM is violated.

Table 6: Descriptive statistics: mean, standard deviation, and median.

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 ≤ age ≤ 65	.7057	.4557	1.000
single	= 1 if single	.4094	.4917	.0000
child	= 1 if children ≤ 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

Firstly we give the coefficient estimates for five different smoothing parameter, including 0 and 1000, together with the bootstrap estimates of the standard errors, see Table 7. The bandwidths have been $h = h_c \sigma_W$ with $h_c = 0$ (giving a FEM), 0.2, 0.4, 0.8, and 1000 (giving a MEM); where σ_W is the vector of standard deviations for the comarca covariates. In the bootstrap we used g = 1.1h as the pilot bandwidth for the pre-estimation, and 400 bootstrap replications.

Table 7: Coefficient and variance estimates with their bootstrap standard errors.

$h_c =$	0.0		0.2		0.4		0.8		1000	
Var.	$\hat{\beta}$	S.E.	$\hat{\beta}$	S.E.	$\hat{\beta}$	S.E.	$\hat{\beta}$	S.E.	$\hat{\beta}$	S.E.
sex	-.0284	.0447	-.0147	.0437	-.0242	.0448	-.0264	.0399	-.0327	.0409
age1	.2428	.1199	.2201	.1196	.2249	.1188	.2226	.1047	.2271	.1017
age2	.2665	.1080	.1916	.0972	.1978	.0966	.2069	.0857	.2145	.0879
single	-.0402	.0613	-.0877	.0609	-.0745	.0568	-.0667	.0548	-.0543	.0526
child	.0003	.0565	-.0190	.0565	-.0166	.0501	-.0103	.0475	-.0164	.0486
ngal	.2288	.0560	.2237	.0582	.2377	.0565	.2428	.0468	.2474	.0471
educ	.0648	.0487	.0514	.0509	.0481	.0454	.0514	.0404	.0517	.0410
stud	-.2219	.1011	-.2134	.1045	-.2212	.0963	-.2173	.0816	-.2312	.0829
self	.0809	.0786	.1316	.0825	.1288	.0740	.1173	.0719	.1131	.0721
pilgr	-.7004	.0910	-.7188	.0936	-.6926	.0752	-.6802	.0654	-.6918	.0683
family	-.1798	.0544	-.1304	.0527	-.1478	.0487	-.1565	.0466	-.1689	.0443
stay	-.0047	.0014	-.0042	.0015	-.0045	.0014	-.0045	.0012	-.0044	.0011
$\hat{\sigma}_u^2$	.0001	.0141	.0108	.0165	.0297	.0171	.0424	.0133	.0650	.0123
$\hat{\sigma}_e^2$	.2545	.0000	.2763	.0016	.2937	.0067	.2770	.0042	.2830	.0050

Table 8: For FEM, the parametric estimates (parametric orthogonal least squares regression) of the standard errors (O) and the bootstrap approximation (B).

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

For the FEM, see the columns refer to $h_c = 0.0$ in Table 7, we can compare the results with a parametric orthogonal least squares regression to check the robustness of our implementation when applying it to real data. The coefficient estimates coincide perfectly with the numbers given in Table 7 and we get an $R^2 = .3908$. The parametric estimates of the standard errors deviate only slightly from the bootstrap analogs, see Table 8 which is organized like Table 7. We conclude once more that the bootstrap approximation seems to work reasonable well. However, have in mind that also the bootstrap estimates certainly are model based estimates. So if it is used for example to approximate the mean squared error of a predictor in the MEM although the independence assumption is violated, then it is clear that the bootstrap will strongly under estimate the true mean squared error.

Next, the aim is to predict

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

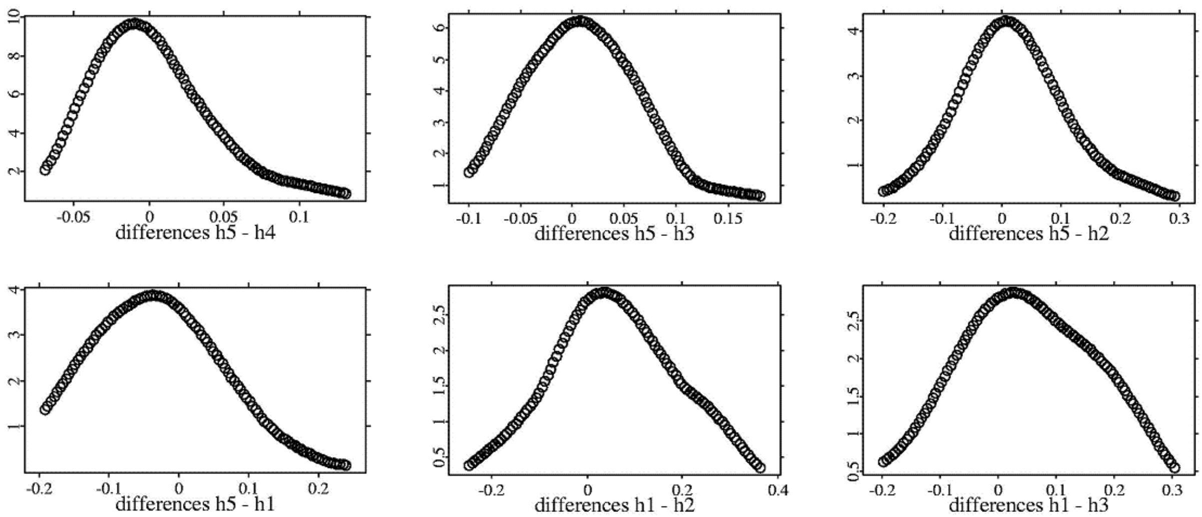
or $\widetilde{Y}_{d\bullet} := \bar{Y}_{d\bullet} - \bar{Y}_{\bullet\bullet}$. We have done this with the same bandwidth and the same bootstrap as above. While the coefficient estimates in Table 7 seem not to change a lot with the chosen model, the estimation of these parameters do. For half of the comarcas *d*, their $\widetilde{Y}_{d\bullet}$ is changing significantly with the chosen model; some of the values tripled when changing from one model to the other, others changed signs, etc. For illustration issues let us denote by $0 = h_1 < h_2 < \ldots < h_5 = 1000$ the five different smoothing parameter. We are interested in differences between the MEM h_5 and the SMEM (h_2, h_3, h_4), respectively the FEM (h_1), and between the FEM and the most flexible SMEM (h_2). First, note that for $\bar{Y}_{\bullet\bullet}$ we have

	h_1	h_2	h_3	h_4	h_5
$\widehat{\bar{Y}_{\bullet\bullet}}$	3.8166	3.8076	3.7993	3.7825	3.6922

For an easier comparison of the different predictions of $\widetilde{Y}_{d\bullet}$ we have plotted in Figure 6 the differences between predictors based on different models, i.e. based on the different h_j. We can see from these graphs that the differences are quite serious and the bootstrap estimates of the standard deviations (not shown them here) reveal that many changes are significant.

Finally, let us simply look at he number of sign changes of $\widetilde{Y}_{d\bullet}$ when switching from model to model. When we express the models simply in terms of h_l (l = 1,2,3,4,5), we observe the following number of sign changes ($\widetilde{Y}_{d\bullet} < 0$ to $\widetilde{Y}_{d\bullet} > 0$ or vice versa) for the 53 comarcas:

Figure 6: Densities of the differences $\widetilde{Y}_{d\bullet}(h_j) - \widetilde{Y}_{d\bullet}(h_k)$ for different j,k=1,2,3,4,5



h_5 to h_4	h_5 to h_3	h_5 to h_2	h_5 to h_1	h_1 to h_2
4	6	6	4	8

This brings to evidence once again that possible dependence between X and the area effect matters a great deal. Conclusions like political decisions, which are based on $\widetilde{Y}_{d\bullet}$ can therefore be strongly misleading if they rely on simple mixed effects models, or highly vague if based on simple fixed effects models (due to the large variance).

4.2 Concluding Remarks

Summarizing, we have got an idea that the model choice not only matters, an inappropriate choice can give rather misleading estimates and predictions. The bootstrap estimates of the standard error decrease clearly with increasing bandwidth giving the smallest values for the MEM. Unfortunately, we know that for the MEM the model bias is the most serious problem but is not captured by the parametric bootstrap. Not surprisingly, we conclude that our SMEM arms the practitioner with a powerful tool but cannot overcome the bias-variance trade-off inherit in all statistical problems. Further research will be directed into studying a possible data adaptive choice of the switcher. However, in practice this choice will often be - at least to some extent - subject to the empirical researchers aims and interests, and it is not clear what would be the most opportune objective function for such a bandwidth choice.

References

BATTESE, G.E., R.M. HARTER AND W.A. FULLER. (1988) An Error Components Model or Prediction of County Crop Areas Using Survey and Satellite Data. *Journal of the American Statistical Association*, **83**: 28-36.

CARROLL, R.J., J. FAN, I. GIJBELS, AND M.P. WAND (1997) Generalized Partially Linear Single-index Models. *Journal of the American Statistical Association*, **92**: 477-489.

DIGGLE, P.J., P. HEAGERTY, K-L. LIANG AND S. ZEGER (2002) *Analysis of Longitudinal Data*, 2nd ed. Oxford University Press.

FAHRMEIR, L. AND S. LANG (2001) Bayesian Inference for Generalized Additive Mixed Models Based on Markov Random Field Priors. *Applied Statistics, JRSSC*, **50**: 201-220.

FAHRMEIR, L., T. KNEIB AND S. LANG (2004) Penalized Structured Additive Regression for Space-Time Data: A Bayesian Perspecitve. *Statistica Sinica*, **14**: 715-745.

FAY, R.E. AND R.A. HERRIOT (1979) Estimates of Income for Small Places. An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association*, **74**:
69-277.

GHOSH, M. AND J.N.K. RAO (1994) Small Area Estimation: an Appraisal. *Statistical Science*, **9**: 55-93 (with discussion).

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

HÄRDLE, W., S. HUET, E. MAMMEN, AND S. SPERLICH (2004) Bootstrap Inference in Semiparametric Generalized Additive Models. *Econometric Theory*, **20**: 265-300.

HÄRDLE, W. AND J.S. MARRON (1991) Bootstrap simultaneous error bars for non-parametric regression. *Annals of Statistics*, **19**: 778-796.

HÄRDLE, W., M. MÜLLER, S. SPERLICH, AND A. WERWATZ (2004) *Nonparametric and Semiparametric Models.* Springer Verlag Berlin Heidelberg.

JIANG, J. AND P. LAHIRI (2006) Mixed Model Prediction and Small Area Estimation *Test*, **15**: 1-96.

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

LAIRD, N.M. AND J.H.WARE (1982) Random-Effects Models for Longitudinal Data. *Biometrics*, **38**: 963-974.

LIN, X. AND R. CARROLL (2006) Semiparametric Estimation in General Repeated Measures Problems. *Journal of the Royal Statistical Society, Series B*, **68**: 69-88.

LOMBARDÍA, M.J. AND S. SPERLICH (2006) Testing the Parametric vs the Semi-parametric Generalized Mixed Effects Models. *Working Paper n.294, FUNCAS, Spain. http* : *\\www:funcas:ceca:es=P ublicaciones=Documentos Trabajo:asp*

MAITY, A., Y. MA, AND R. CARROLL (2007) Efficient Estimation of Population-level Summaries in General Semiparametric Regression Models. *Journal of the American Statistical Association*, *to appear*.

OPSOMER, J., G. CLAESKENS, M.G. RANALLI, G. KAUERMANN, AND F.J. BREIDT (2005) Nonparametric Small Area Estimation Using Penalized Spline Regression. *Preprint, Universität Bielefeld.*

PRASAD, N.G.N. AND J.N.K. RAO (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.

ROBINSON, P. (1988) Root-N-Consistent Semiparametric Regression. *Econometrica*, **56**: 931-954.

ROCA PARDINAS, J. AND S. SPERLICH (2007) Efficient Estimation of Generalized Structured Models. *Working Paper, Georg-August Universität Göttingen*

RUPPERT, D., M.P. WAND, AND R.J. CARROLL (2003) *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.

SPECKMAN, P. (1988) Kernel Smoothing in Partial Linear Models. *J.Royal Statistical Society, Ser.B*, **50**: 413 - 436.

SPERLICH, S., D. TJOTHEIM, AND L. YANG (2002) Nonparametric Estimation and Testing of Interaction in Additive Models. *Econometric theory*, **18**, 197-251.

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

TUTZ, G. AND F. REITHINGER (2007) Flexible Semiparametric Mixed Models. *Statistics in Medicine*, to appear.

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

WAND, M.P. (2003) Smoothing and Mixed Models. *Computational Statistics*, **18**: 223-249.

FUNDACIÓN DE LAS CAJAS DE AHORROS

DOCUMENTOS DE TRABAJO

Últimos números publicados

159/2000 Participación privada en la construcción y explotación de carreteras de peaje
Ginés de Rus, Manuel Romero y Lourdes Trujillo

160/2000 Errores y posibles soluciones en la aplicación del *Value at Risk*
Mariano González Sánchez

161/2000 Tax neutrality on saving assets. The spahish case before and after the tax reform
Cristina Ruza y de Paz-Curbera

162/2000 Private rates of return to human capital in Spain: new evidence
F. Barceinas, J. Oliver-Alonso, J.L. Raymond y J.L. Roig-Sabaté

163/2000 El control interno del riesgo. Una propuesta de sistema de límites riesgo neutral
Mariano González Sánchez

164/2001 La evolución de las políticas de gasto de las Administraciones Públicas en los años 90
Alfonso Utrilla de la Hoz y Carmen Pérez Esparrells

165/2001 Bank cost efficiency and output specification
Emili Tortosa-Ausina

166/2001 Recent trends in Spanish income distribution: A robust picture of falling income inequality
Josep Oliver-Alonso, Xavier Ramos y José Luis Raymond-Bara

167/2001 Efectos redistributivos y sobre el bienestar social del tratamiento de las cargas familiares en el nuevo IRPF
Nuria Badenes Plá, Julio López Laborda, Jorge Onrubia Fernández

168/2001 The Effects of Bank Debt on Financial Structure of Small and Medium Firms in some European Countries
Mónica Melle-Hernández

169/2001 La política de cohesión de la UE ampliada: la perspectiva de España
Ismael Sanz Labrador

170/2002 Riesgo de liquidez de Mercado
Mariano González Sánchez

171/2002 Los costes de administración para el afiliado en los sistemas de pensiones basados en cuentas de capitalización individual: medida y comparación internacional.
José Enrique Devesa Carpio, Rosa Rodríguez Barrera, Carlos Vidal Meliá

172/2002 La encuesta continua de presupuestos familiares (1985-1996): descripción, representatividad y propuestas de metodología para la explotación de la información de los ingresos y el gasto.
Llorenc Pou, Joaquín Alegre

173/2002 Modelos paramétricos y no paramétricos en problemas de concesión de tarjetas de credito.
Rosa Puertas, María Bonilla, Ignacio Olmeda

174/2002 Mercado único, comercio intra-industrial y costes de ajuste en las manufacturas españolas.
José Vicente Blanes Cristóbal

175/2003 La Administración tributaria en España. Un análisis de la gestión a través de los ingresos y de los gastos.
Juan de Dios Jiménez Aguilera, Pedro Enrique Barrilao González

176/2003 The Falling Share of Cash Payments in Spain.
Santiago Carbó Valverde, Rafael López del Paso, David B. Humphrey
Publicado en "Moneda y Crédito" nº 217, pags. 167-189.

177/2003 Effects of ATMs and Electronic Payments on Banking Costs: The Spanish Case.
Santiago Carbó Valverde, Rafael López del Paso, David B. Humphrey

178/2003 Factors explaining the interest margin in the banking sectors of the European Union.
Joaquín Maudos y Juan Fernández Guevara

179/2003 Los planes de stock options para directivos y consejeros y su valoración por el mercado de valores en España.
Mónica Melle Hernández

180/2003 Ownership and Performance in Europe and US Banking – A comparison of Commercial, Co-operative & Savings Banks.
Yener Altunbas, Santiago Carbó y Phil Molyneux

181/2003 The Euro effect on the integration of the European stock markets.
Mónica Melle Hernández

182/2004 In search of complementarity in the innovation strategy: international R&D and external knowledge acquisition.
Bruno Cassiman, Reinhilde Veugelers

183/2004 Fijación de precios en el sector público: una aplicación para el servicio municipal de suministro de agua.
Mª Ángeles García Valiñas

184/2004 Estimación de la economía sumergida es España: un modelo estructural de variables latentes.
Ángel Alañón Pardo, Miguel Gómez de Antonio

185/2004 Causas políticas y consecuencias sociales de la corrupción.
Joan Oriol Prats Cabrera

186/2004 Loan bankers' decisions and sensitivity to the audit report using the belief revision model.
Andrés Guiral Contreras and José A. Gonzalo Angulo

187/2004 El modelo de Black, Derman y Toy en la práctica. Aplicación al mercado español.
Marta Tolentino García-Abadillo y Antonio Díaz Pérez

188/2004 Does market competition make banks perform well?.
Mónica Melle

189/2004 Efficiency differences among banks: external, technical, internal, and managerial
Santiago Carbó Valverde, David B. Humphrey y Rafael López del Paso

190/2004 Una aproximación al análisis de los costes de la esquizofrenia en españa: los modelos jerárquicos bayesianos
F. J. Vázquez-Polo, M. A. Negrín, J. M. Cavasés, E. Sánchez y grupo RIRAG

191/2004 Environmental proactivity and business performance: an empirical analysis
Javier González-Benito y Óscar González-Benito

192/2004 Economic risk to beneficiaries in notional defined contribution accounts (NDCs)
Carlos Vidal-Meliá, Inmaculada Domínguez-Fabian y José Enrique Devesa-Carpio

193/2004 Sources of efficiency gains in port reform: non parametric malmquist decomposition tfp index for Mexico
Antonio Estache, Beatriz Tovar de la Fé y Lourdes Trujillo

194/2004 Persistencia de resultados en los fondos de inversión españoles
Alfredo Ciriaco Fernández y Rafael Santamaría Aquilué

195/2005 El modelo de revisión de creencias como aproximación psicológica a la formación del juicio del auditor sobre la gestión continuada
Andrés Guiral Contreras y Francisco Esteso Sánchez

196/2005 La nueva financiación sanitaria en España: descentralización y prospectiva
David Cantarero Prieto

197/2005 A cointegration analysis of the Long-Run supply response of Spanish agriculture to the common agricultural policy
José A. Mendez, Ricardo Mora y Carlos San Juan

198/2005 ¿Refleja la estructura temporal de los tipos de interés del mercado español preferencia por la liquidez?
Magdalena Massot Perelló y Juan M. Nave

199/2005 Análisis de impacto de los Fondos Estructurales Europeos recibidos por una economía regional: Un enfoque a través de Matrices de Contabilidad Social
M. Carmen Lima y M. Alejandro Cardenete

200/2005 Does the development of non-cash payments affect monetary policy transmission?
Santiago Carbó Valverde y Rafael López del Paso

201/2005 Firm and time varying technical and allocative efficiency: an application for port cargo handling firms
Ana Rodríguez-Álvarez, Beatriz Tovar de la Fe y Lourdes Trujillo

202/2005 Contractual complexity in strategic alliances
Jeffrey J. Reuer y Africa Ariño

203/2005 Factores determinantes de la evolución del empleo en las empresas adquiridas por opa
Nuria Alcalde Fradejas y Inés Pérez-Soba Aguilar

204/2005 Nonlinear Forecasting in Economics: a comparison between Comprehension Approach versus Learning Approach. An Application to Spanish Time Series
Elena Olmedo, Juan M. Valderas, Ricardo Gimeno and Lorenzo Escot

205/2005 Precio de la tierra con presión urbana: un modelo para España
Esther Decimavilla, Carlos San Juan y Stefan Sperlich

206/2005 Interregional migration in Spain: a semiparametric analysis
Adolfo Maza y José Villaverde

207/2005 Productivity growth in European banking
Carmen Murillo-Melchor, José Manuel Pastor y Emili Tortosa-Ausina

208/2005 Explaining Bank Cost Efficiency in Europe: Environmental and Productivity Influences.
Santiago Carbó Valverde, David B. Humphrey y Rafael López del Paso

209/2005 La elasticidad de sustitución intertemporal con preferencias no separables intratemporalmente: los casos de Alemania, España y Francia.
Elena Márquez de la Cruz, Ana R. Martínez Cañete y Inés Pérez-Soba Aguilar

210/2005 Contribución de los efectos tamaño, book-to-market y momentum a la valoración de activos: el caso español.
Begoña Font-Belaire y Alfredo Juan Grau-Grau

211/2005 Permanent income, convergence and inequality among countries
José M. Pastor and Lorenzo Serrano

212/2005 The Latin Model of Welfare: Do 'Insertion Contracts' Reduce Long-Term Dependence?
Luis Ayala and Magdalena Rodríguez

213/2005 The effect of geographic expansion on the productivity of Spanish savings banks
Manuel Illueca, José M. Pastor and Emili Tortosa-Ausina

214/2005 Dynamic network interconnection under consumer switching costs
Ángel Luis López Rodríguez

215/2005 La influencia del entorno socioeconómico en la realización de estudios universitarios: una aproximación al caso español en la década de los noventa
Marta Rahona López

216/2005 The valuation of spanish ipos: efficiency analysis
Susana Álvarez Otero

217/2005 On the generation of a regular multi-input multi-output technology using parametric output distance functions
Sergio Perelman and Daniel Santin

218/2005 La gobernanza de los procesos parlamentarios: la organización industrial del congreso de los diputados en España
Gonzalo Caballero Miguez

219/2005 Determinants of bank market structure: Efficiency and political economy variables
Francisco González

220/2005 Agresividad de las órdenes introducidas en el mercado español: estrategias, determinantes y medidas de performance
David Abad Díaz

221/2005 Tendencia post-anuncio de resultados contables: evidencia para el mercado español
Carlos Forner Rodríguez, Joaquín Marhuenda Fructuoso y Sonia Sanabria García

222/2005 Human capital accumulation and geography: empirical evidence in the European Union
Jesús López-Rodríguez, J. Andrés Faíña y Jose Lopez Rodríguez

223/2005 Auditors' Forecasting in Going Concern Decisions: Framing, Confidence and Information Processing
Waymond Rodgers and Andrés Guiral

224/2005 The effect of Structural Fund spending on the Galician region: an assessment of the 1994-1999 and 2000-2006 Galician CSFs
José Ramón Cancelo de la Torre, J. Andrés Faíña and Jesús López-Rodríguez

225/2005 The effects of ownership structure and board composition on the audit committee activity: Spanish evidence
Carlos Fernández Méndez and Rubén Arrondo García

226/2005 Cross-country determinants of bank income smoothing by managing loan loss provisions
Ana Rosa Fonseca and Francisco González

227/2005 Incumplimiento fiscal en el irpf (1993-2000): un análisis de sus factores determinantes
Alejandro Estellér Moré

228/2005 Region versus Industry effects: volatility transmission
Pilar Soriano Felipe and Francisco J. Climent Diranzo

229/2005 Concurrent Engineering: The Moderating Effect Of Uncertainty On New Product Development Success
Daniel Vázquez-Bustelo and Sandra Valle

230/2005 On zero lower bound traps: a framework for the analysis of monetary policy in the 'age' of central banks
Alfonso Palacio-Vera

231/2005 Reconciling Sustainability and Discounting in Cost Benefit Analysis: a methodological proposal
M. Carmen Almansa Sáez and Javier Calatrava Requena

232/2005 Can The Excess Of Liquidity Affect The Effectiveness Of The European Monetary Policy?
Santiago Carbó Valverde and Rafael López del Paso

233/2005 Inheritance Taxes In The Eu Fiscal Systems: The Present Situation And Future Perspectives.
Miguel Angel Barberán Lahuerta

234/2006 Bank Ownership And Informativeness Of Earnings.
Víctor M. González

235/2006 Developing A Predictive Method: A Comparative Study Of The Partial Least Squares Vs Maximum Likelihood Techniques.
Waymond Rodgers, Paul Pavlou and Andres Guiral.

236/2006 Using Compromise Programming for Macroeconomic Policy Making in a General Equilibrium Framework: Theory and Application to the Spanish Economy.
Francisco J. André, M. Alejandro Cardenete y Carlos Romero.

237/2006 Bank Market Power And Sme Financing Constraints.
Santiago Carbó-Valverde, Francisco Rodríguez-Fernández y Gregory F. Udell.

238/2006 Trade Effects Of Monetary Agreements: Evidence For Oecd Countries.
Salvador Gil-Pareja, Rafael Llorca-Vivero y José Antonio Martínez-Serrano.

239/2006 The Quality Of Institutions: A Genetic Programming Approach.
Marcos Álvarez-Díaz y Gonzalo Caballero Miguez.

240/2006 La interacción entre el éxito competitivo y las condiciones del mercado doméstico como determinantes de la decisión de exportación en las Pymes.
Francisco García Pérez.

241/2006 Una estimación de la depreciación del capital humano por sectores, por ocupación y en el tiempo.
Inés P. Murillo.

242/2006 Consumption And Leisure Externalities, Economic Growth And Equilibrium Efficiency.
Manuel A. Gómez.

243/2006 Measuring efficiency in education: an analysis of different approaches for incorporating non-discretionary inputs.
Jose Manuel Cordero-Ferrera, Francisco Pedraja-Chaparro y Javier Salinas-Jiménez

244/2006 Did The European Exchange-Rate Mechanism Contribute To The Integration Of Peripheral Countries?.
Salvador Gil-Pareja, Rafael Llorca-Vivero y José Antonio Martínez-Serrano

245/2006 Intergenerational Health Mobility: An Empirical Approach Based On The Echp.
Marta Pascual and David Cantarero

246/2006 Measurement and analysis of the Spanish Stock Exchange using the Lyapunov exponent with digital technology.
Salvador Rojí Ferrari and Ana Gonzalez Marcos

247/2006 Testing For Structural Breaks In Variance Withadditive Outliers And Measurement Errors.
Paulo M.M. Rodrigues and Antonio Rubia

248/2006 The Cost Of Market Power In Banking: Social Welfare Loss Vs. Cost Inefficiency.
Joaquín Maudos and Juan Fernández de Guevara

249/2006 Elasticidades de largo plazo de la demanda de vivienda: evidencia para España (1885-2000).
Desiderio Romero Jordán, José Félix Sanz Sanz y César Pérez López

250/2006 Regional Income Disparities in Europe: What role for location?.
Jesús López-Rodríguez and J. Andrés Faíña

251/2006 Funciones abreviadas de bienestar social: Una forma sencilla de simultanear la medición de la eficiencia y la equidad de las políticas de gasto público.
Nuria Badenes Plá y Daniel Santín González

252/2006 "The momentum effect in the Spanish stock market: Omitted risk factors or investor behaviour?".
Luis Muga and Rafael Santamaría

253/2006 Dinámica de precios en el mercado español de gasolina: un equilibrio de colusión tácita.
Jordi Perdiguero García

254/2006 Desigualdad regional en España: renta permanente versus renta corriente.
José M.Pastor, Empar Pons y Lorenzo Serrano

255/2006 Environmental implications of organic food preferences: an application of the impure public goods model.
Ana Maria Aldanondo-Ochoa y Carmen Almansa-Sáez

256/2006 Family tax credits versus family allowances when labour supply matters: Evidence for Spain.
José Felix Sanz-Sanz, Desiderio Romero-Jordán y Santiago Álvarez-García

257/2006 La internacionalización de la empresa manufacturera española: efectos del capital humano genérico y específico.
José López Rodríguez

258/2006 Evaluación de las migraciones interregionales en España, 1996-2004.
María Martínez Torres

259/2006 Efficiency and market power in Spanish banking.
Rolf Färe, Shawna Grosskopf y Emili Tortosa-Ausina.

260/2006 Asimetrías en volatilidad, beta y contagios entre las empresas grandes y pequeñas cotizadas en la bolsa española.
Helena Chuliá y Hipòlit Torró.

261/2006 Birth Replacement Ratios: New Measures of Period Population Replacement.
José Antonio Ortega.

262/2006 Accidentes de tráfico, víctimas mortales y consumo de alcohol.
José Mª Arranz y Ana I. Gil.

263/2006 Análisis de la Presencia de la Mujer en los Consejos de Administración de las Mil Mayores Empresas Españolas.
Ruth Mateos de Cabo, Lorenzo Escot Mangas y Ricardo Gimeno Nogués.

264/2006 Crisis y Reforma del Pacto de Estabilidad y Crecimiento. Las Limitaciones de la Política Económica en Europa.
Ignacio Álvarez Peralta.

265/2006 Have Child Tax Allowances Affected Family Size? A Microdata Study For Spain (1996-2000).
Jaime Vallés-Giménez y Anabel Zárate-Marco.

266/2006 Health Human Capital And The Shift From Foraging To Farming.
Paolo Rungo.

267/2006 Financiación Autonómica y Política de la Competencia: El Mercado de Gasolina en Canarias.
Juan Luis Jiménez y Jordi Perdiguero.

268/2006 El cumplimiento del Protocolo de Kyoto para los hogares españoles: el papel de la imposición sobre la energía.
Desiderio Romero-Jordán y José Félix Sanz-Sanz.

269/2006 Banking competition, financial dependence and economic growth
Joaquín Maudos y Juan Fernández de Guevara

270/2006 Efficiency, subsidies and environmental adaptation of animal farming under CAP
Werner Kleinhanß, Carmen Murillo, Carlos San Juan y Stefan Sperlich

271/2006 Interest Groups, Incentives to Cooperation and Decision-Making Process in the European Union
A. Garcia-Lorenzo y Jesús López-Rodríguez

272/2006 Riesgo asimétrico y estrategias de momentum en el mercado de valores español
Luis Muga y Rafael Santamaría

273/2006 Valoración de capital-riesgo en proyectos de base tecnológica e innovadora a través de la teoría de opciones reales
Gracia Rubio Martín

274/2006 Capital stock and unemployment: searching for the missing link
Ana Rosa Martínez-Cañete, Elena Márquez de la Cruz, Alfonso Palacio-Vera and Inés Pérez-Soba Aguilar

275/2006 Study of the influence of the voters' political culture on vote decision through the simulation of a political competition problem in Spain
Sagrario Lantarón, Isabel Lillo, Mª Dolores López and Javier Rodrigo

276/2006 Investment and growth in Europe during the Golden Age
Antonio Cubel and Mª Teresa Sanchis

277/2006 Efectos de vincular la pensión pública a la inversión en cantidad y calidad de hijos en un modelo de equilibrio general
Robert Meneu Gaya

278/2006 El consumo y la valoración de activos
Elena Márquez y Belén Nieto

279/2006 Economic growth and currency crisis: A real exchange rate entropic approach
David Matesanz Gómez y Guillermo J. Ortega

280/2006 Three measures of returns to education: An illustration for the case of Spain
María Arrazola y José de Hevia

281/2006 Composition of Firms versus Composition of Jobs
Antoni Cunyat

282/2006 La vocación internacional de un holding tranviario belga: la Compagnie Mutuelle de Tram-ways, 1895-1918
Alberte Martínez López

283/2006 Una visión panorámica de las entidades de crédito en España en la última década.
Constantino García Ramos

284/2006 Foreign Capital and Business Strategies: a comparative analysis of urban transport in Madrid and Barcelona, 1871-1925
Alberte Martínez López

285/2006 Los intereses belgas en la red ferroviaria catalana, 1890-1936
Alberte Martínez López

286/2006 The Governance of Quality: The Case of the Agrifood Brand Names
Marta Fernández Barcala, Manuel González-Díaz y Emmanuel Raynaud

287/2006 Modelling the role of health status in the transition out of malthusian equilibrium
Paolo Rungo, Luis Currais and Berta Rivera

288/2006 Industrial Effects of Climate Change Policies through the EU Emissions Trading Scheme
Xavier Labandeira and Miguel Rodríguez

289/2006 Globalisation and the Composition of Government Spending: An analysis for OECD countries
Norman Gemmell, Richard Kneller and Ismael Sanz

290/2006 La producción de energía eléctrica en España: Análisis económico de la actividad tras la liberalización del Sector Eléctrico
Fernando Hernández Martínez

291/2006 Further considerations on the link between adjustment costs and the productivity of R&D investment: evidence for Spain
Desiderio Romero-Jordán, José Félix Sanz-Sanz and Inmaculada Álvarez-Ayuso

292/2006 Una teoría sobre la contribución de la función de compras al rendimiento empresarial
Javier González Benito

293/2006 Agility drivers, enablers and outcomes: empirical test of an integrated agile manufacturing model
Daniel Vázquez-Bustelo, Lucía Avella and Esteban Fernández

294/2006 Testing the parametric vs the semiparametric generalized mixed effects models
María José Lombardía and Stefan Sperlich

295/2006 Nonlinear dynamics in energy futures
Mariano Matilla-García

296/2006 Estimating Spatial Models By Generalized Maximum Entropy Or How To Get Rid Of W
Esteban Fernández Vázquez, Matías Mayor Fernández and Jorge Rodriguez-Valez

297/2006 Optimización fiscal en las transmisiones lucrativas: análisis metodológico
Félix Domínguez Barrero

298/2006 La situación actual de la banca online en España
Francisco José Climent Diranzo y Alexandre Momparler Pechuán

299/2006 Estrategia competitiva y rendimiento del negocio: el papel mediador de la estrategia y las capacidades productivas
Javier González Benito y Isabel Suárez González

300/2006 A Parametric Model to Estimate Risk in a Fixed Income Portfolio
Pilar Abad and Sonia Benito

301/2007 Análisis Empírico de las Preferencias Sociales Respecto del Gasto en Obra Social de las Cajas de Ahorros
Alejandro Esteller-Moré, Jonathan Jorba Jiménez y Albert Solé-Ollé

302/2007 Assessing the enlargement and deepening of regional trading blocs: The European Union case
Salvador Gil-Pareja, Rafael Llorca-Vivero y José Antonio Martínez-Serrano

303/2007 ¿Es la Franquicia un Medio de Financiación?: Evidencia para el Caso Español
Vanesa Solís Rodríguez y Manuel González Díaz

304/2007 On the Finite-Sample Biases in Nonparametric Testing for Variance Constancy
Paulo M.M. Rodrigues and Antonio Rubia

305/2007 Spain is Different: Relative Wages 1989-98
José Antonio Carrasco Gallego

306/2007 Poverty reduction and SAM multipliers: An evaluation of public policies in a regional framework
Francisco Javier De Miguel-Vélez y Jesús Pérez-Mayo

307/2007 La Eficiencia en la Gestión del Riesgo de Crédito en las Cajas de Ahorro
Marcelino Martínez Cabrera

308/2007 Optimal environmental policy in transport: unintended effects on consumers' generalized price
M. Pilar Socorro and Ofelia Betancor

309/2007 Agricultural Productivity in the European Regions: Trends and Explanatory Factors
Roberto Ezcurra, Belen Iráizoz, Pedro Pascual and Manuel Rapún

310/2007 Long-run Regional Population Divergence and Modern Economic Growth in Europe: a Case Study of Spain
María Isabel Ayuda, Fernando Collantes and Vicente Pinilla

311/2007 Financial Information effects on the measurement of Commercial Banks' Efficiency
Borja Amor, María T. Tascón and José L. Fanjul

312/2007 Neutralidad e incentivos de las inversiones financieras en el nuevo IRPF
Félix Domínguez Barrero

313/2007 The Effects of Corporate Social Responsibility Perceptions on The Valuation of Common Stock
Waymond Rodgers , Helen Choy and Andres Guiral-Contreras

314/2007 Country Creditor Rights, Information Sharing and Commercial Banks' Profitability Persistence across the world
Borja Amor, María T. Tascón and José L. Fanjul

315/2007 ¿Es Relevante el Déficit Corriente en una Unión Monetaria? El Caso Español
Javier Blanco González y Ignacio del Rosal Fernández

316/2007 The Impact of Credit Rating Announcements on Spanish Corporate Fixed Income Performance: Returns, Yields and Liquidity
Pilar Abad, Antonio Díaz and M. Dolores Robles

317/2007 Indicadores de Lealtad al Establecimiento y Formato Comercial Basados en la Distribución del Presupuesto
Cesar Augusto Bustos Reyes y Óscar González Benito

318/2007 Migrants and Market Potential in Spain over The XXth Century: A Test Of The New Economic Geography
Daniel A. Tirado, Jordi Pons, Elisenda Paluzie and Javier Silvestre

319/2007 El Impacto del Coste de Oportunidad de la Actividad Emprendedora en la Intención de los Ciudadanos Europeos de Crear Empresas
Luis Miguel Zapico Aldeano

320/2007 Los belgas y los ferrocarriles de vía estrecha en España, 1887-1936
Alberte Martínez López

321/2007 Competición política bipartidista. Estudio geométrico del equilibrio en un caso ponderado
Isabel Lillo, Mª Dolores López y Javier Rodrigo

322/2007 Human resource management and environment management systems: an empirical study
Mª Concepción López Fernández, Ana Mª Serrano Bedia and Gema García Piqueres

323/2007 Wood and industrialization. evidence and hypotheses from the case of Spain, 1860-1935.
Iñaki Iriarte-Goñi and María Isabel Ayuda Bosque

324/2007 New evidence on long-run monetary neutrality.
J. Cunado, L.A. Gil-Alana and F. Perez de Gracia

325/2007 Monetary policy and structural changes in the volatility of us interest rates.
Juncal Cuñado, Javier Gomez Biscarri and Fernando Perez de Gracia

326/2007 The productivity effects of intrafirm diffusion.
Lucio Fuentelsaz, Jaime Gómez and Sergio Palomas

327/2007 Unemployment duration, layoffs and competing risks.
J.M. Arranz, C. García-Serrano and L. Toharia

328/2007 El grado de cobertura del gasto público en España respecto a la UE-15
Nuria Rueda, Begoña Barruso, Carmen Calderón y Mª del Mar Herrador

329/2007 The Impact of Direct Subsidies in Spain before and after the CAP'92 Reform
Carmen Murillo, Carlos San Juan and Stefan Sperlich

330/2007 Determinants of post-privatisation performance of Spanish divested firms
Laura Cabeza García and Silvia Gómez Ansón

331/2007 ¿Por qué deciden diversificar las empresas españolas? Razones oportunistas versus razones económicas
Almudena Martínez Campillo

332/2007 Dynamical Hierarchical Tree in Currency Markets
Juan Gabriel Brida, David Matesanz Gómez and Wiston Adrián Risso

333/2007 Los determinantes sociodemográficos del gasto sanitario. Análisis con microdatos individuales
Ana María Angulo, Ramón Barberán, Pilar Egea y Jesús Mur

334/2007 Why do companies go private? The Spanish case
Inés Pérez-Soba Aguilar

335/2007 The use of gis to study transport for disabled people
Verónica Cañal Fernández

336/2007 The long run consequences of M&A: An empirical application
Cristina Bernad, Lucio Fuentelsaz and Jaime Gómez

337/2007 Las clasificaciones de materias en economía: principios para el desarrollo de una nueva clasificación
Valentín Edo Hernández

338/2007 Reforming Taxes and Improving Health: A Revenue-Neutral Tax Reform to Eliminate Medical and Pharmaceutical VAT
Santiago Álvarez-García, Carlos Pestana Barros y Juan Prieto-Rodriguez

339/2007 Impacts of an iron and steel plant on residential property values
Celia Bilbao-Terol

340/2007 Firm size and capital structure: Evidence using dynamic panel data
Víctor M. González and Francisco González

341/2007 ¿Cómo organizar una cadena hotelera? La elección de la forma de gobierno
Marta Fernández Barcala y Manuel González Díaz

342/2007 Análisis de los efectos de la decisión de diversificar: un contraste del marco teórico "Agencia-Stewardship"
Almudena Martínez Campillo y Roberto Fernández Gago

343/2007 Selecting portfolios given multiple eurostoxx-based uncertainty scenarios: a stochastic goal programming approach from fuzzy betas
Enrique Ballestero, Blanca Pérez-Gladish, Mar Arenas-Parra and Amelia Bilbao-Terol

344/2007 "El bienestar de los inmigrantes y los factores implicados en la decisión de emigrar"
Anastasia Hernández Alemán y Carmelo J. León

345/2007 Governance Decisions in the R&D Process: An Integrative Framework Based on TCT and Knowledge View of The Firm.
Andrea Martínez-Noya and Esteban García-Canal

346/2007 Diferencias salariales entre empresas públicas y privadas. El caso español
Begoña Cueto y Nuria Sánchez- Sánchez

347/2007 Effects of Fiscal Treatments of Second Home Ownership on Renting Supply
Celia Bilbao Terol and Juan Prieto Rodríguez

348/2007 Auditors' ethical dilemmas in the going concern evaluation
Andres Guiral, Waymond Rodgers, Emiliano Ruiz and Jose A. Gonzalo

349/2007 Convergencia en capital humano en España. Un análisis regional para el periodo 1970-2004
Susana Morales Sequera y Carmen Pérez Esparrells

350/2007 Socially responsible investment: mutual funds portfolio selection using fuzzy multiobjective programming
Blanca Mª Pérez-Gladish, Mar Arenas-Parra , Amelia Bilbao-Terol and Mª Victoria Rodríguez-Uría

351/2007 Persistencia del resultado contable y sus componentes: implicaciones de la medida de ajustes por devengo
Raúl Iñiguez Sánchez y Francisco Poveda Fuentes

352/2007 Wage Inequality and Globalisation: What can we Learn from the Past? A General Equilibrium Approach
Concha Betrán, Javier Ferri and Maria A. Pons

353/2007 Eficacia de los incentivos fiscales a la inversión en I+D en España en los años noventa
Desiderio Romero Jordán y José Félix Sanz Sanz

354/2007 Convergencia regional en renta y bienestar en España
Robert Meneu Gaya

355/2007 Tributación ambiental: Estado de la Cuestión y Experiencia en España
Ana Carrera Poncela

356/2007 Salient features of dependence in daily us stock market indices
Luis A. Gil-Alana, Juncal Cuñado and Fernando Pérez de Gracia

357/2007 La educación superior: ¿un gasto o una inversión rentable para el sector público?
Inés P. Murillo y Francisco Pedraja

358/2007 Effects of a reduction of working hours on a model with job creation and job destruction
Emilio Domínguez, Miren Ullibarri y Idoya Zabaleta

359/2007 Stock split size, signaling and earnings management: Evidence from the Spanish market
José Yagüe, J. Carlos Gómez-Sala and Francisco Poveda-Fuentes

360/2007 Modelización de las expectativas y estrategias de inversión en mercados de derivados
Begoña Font-Belaire

361/2008 Trade in capital goods during the golden age, 1953-1973
Mª Teresa Sanchis and Antonio Cubel

362/2008 El capital económico por riesgo operacional: una aplicación del modelo de distribución de pérdidas
Enrique José Jiménez Rodríguez y José Manuel Feria Domínguez

363/2008 The drivers of effectiveness in competition policy
Joan-Ramon Borrell and Juan-Luis Jiménez

364/2008 Corporate governance structure and board of directors remuneration policies: evidence from Spain
Carlos Fernández Méndez, Rubén Arrondo García and Enrique Fernández Rodríguez

365/2008 Beyond the disciplinary role of governance: how boards and donors add value to Spanish foundations
Pablo De Andrés Alonso, Valentín Azofra Palenzuela y M. Elena Romero Merino

366/2008 Complejidad y perfeccionamiento contractual para la contención del oportunismo en los acuerdos de franquicia
Vanesa Solís Rodríguez y Manuel González Díaz

367/2008 Inestabilidad y convergencia entre las regiones europeas
Jesús Mur, Fernando López y Ana Angulo

368/2008 Análisis espacial del cierre de explotaciones agrarias
Ana Aldanondo Ochoa, Carmen Almansa Sáez y Valero Casanovas Oliva

369/2008 Cross-Country Efficiency Comparison between Italian and Spanish Public Universities in the period 2000-2005
Tommaso Agasisti and Carmen Pérez Esparrells

370/2008 El desarrollo de la sociedad de la información en España: un análisis por comunidades autónomas
María Concepción García Jiménez y José Luis Gómez Barroso

371/2008 El medioambiente y los objetivos de fabricación: un análisis de los modelos estratégicos para su consecución
Lucía Avella Camarero, Esteban Fernández Sánchez y Daniel Vázquez-Bustelo

372/2008 Influence of bank concentration and institutions on capital structure: New international evidence
Víctor M. González and Francisco González

373/2008 Generalización del concepto de equilibrio en juegos de competición política
Mª Dolores López González y Javier Rodrigo Hitos

374/2008 Smooth Transition from Fixed Effects to Mixed Effects Models in Multi-level regression Models
María José Lombardía and Stefan Sperlich