TESTING THE PARAMETRIC VS THE SEMIPARAMETRIC GENERALIZED MIXED EFFECTS MODELS

MARÍA JOSÉ LOMBARDÍA
STEFAN SPERLICH

FUNDACIÓN DE LAS CAJAS DE AHORROS
DOCUMENTO DE TRABAJO
Nº 294/2006

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.

ISBN: 84-89116-07-5

Testing the Parametric vs the Semiparametric Generalized Mixed Effects Models

María José Lombardía
Departamento de Estadística e Investigación Operativa
Universidad de Santiago de Compostela
15782 Santiago de Compostela
mjoselc@usc.es

Stefan Sperlich
Georg-August Universität Göttingen
Institut für Statistik und Ökonometrie
Platz der Göttinger Sieben 5, 37073 Göttingen
stefan.sperlich@wiwi.uni-goettingen.de

Abstract

The paper presents a study of the generalized partially linear model including random effects in its linear part. For these kinds of models we propose an estimator combining likelihood approaches for mixed effects models with kernel methods. Following the methodology of Härdle et al. (1998), we introduce different tests that allow us to choose between a parametric and the semiparametric mixed effects model. Along these lines we also discuss some bootstrap procedures to simulate the critical values. We prove consistency and give asymptotic theory for all our methods. Finally, a simulation study and a real data application are provided in order to demonstrate the feasibility and the excellent behaviour of our methods. [1]

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

[1] The authors gratefully acknowledge the financial supported of the Spanish "Dirección General de Investigación del Ministerio de Ciencia y Tecnología", SEJ2004-04583/ECON, MTM2005-00820 and of the Xunta de Galicia PGIDIT03PXIC20702PN. We further thank Jean Opsomer and two anonymous referees for their helpful discussion.

1. Introduction

Model-based small area estimation has received considerable attention in the last two decades in both the public and private sectors. The term “small area” is commonly used to denote a small geographical area, such a county, a census division, state, etc. It may also describe a “small domain”, i.e. a small subpopulation such as a specific group of people classified by age and sex or their race; or it even refers to climatic clusters. It is recognized that direct survey estimates for small areas are likely to yield unacceptably large standard errors due to the sample size. Traditionally, small area estimation relies on linear mixed effects model, relating the responses of interest in the small areas to each other and to covariates. Mixed models are suitable for small area estimation because of their flexibility in effectively combining different sources of information and explaining different sources of errors. These models typically incorporate area-specific random effects to explain the variations in information from different areas not explained by the fixed effects part of the model. In the exclusively model-based framework some interesting research has been done using Bayes methodology and using frequentist methodology, see Ghosh and Rao (1994), Rao (2003) and Jiang and Lahiri (2006) for a thorough review of different small-area estimation techniques.

Nowadays, in some countries, small area inference is an important statistical tool. Since 2003, EUROSTAT requires that EU States use it to provide statistics for their small areas (provinces, districts, departments, etc.). In the United States it is used by, among others, the Census Bureau’s Small Area Income and Poverty Estimates Program, the Bureau of Labor Statistics’ Local Area Unemployment Statistics Program, and the National Agricultural Statistics Service’s County Estimates Program. It is not surprising that small area statistics has attracted a lot of attention in statistical research. For example, the U.S. Department of Agriculture publishes annual estimates of farm real-estate values for 48 states based on the Agricultural and Land Values Survey (ALVS), which is characterized by its low response rate. For that reason, Pfefferman and Barnard (1991) studied efficient ways of combining the auxiliary data (e.g. from the Agriculture Census) with ALVS to improve estimates of farm land value. Battese et al. (1988) proposed model-based county estimates of crop acreage using remote sensing satellite data as auxiliary information, a method today adopted by the U.S. National Agricultural Service. To illustrate our methodology we are going to apply it to predict the number of hectares of forest in 53 *comarcas* (administrative regions) of Galicia (Northwest of Spain). This methodology has the advantage of being inexpensive and easily implemented.

Prasad and Rao (1990) discussed three of the most frequently considered models in small area statistics giving one closed linear mixed effects formula, see below. Similar to the simple linear models, the linear mixed effects models have been extended to generalized mixed effect models, defined by

$$G\left(E[Y_{dj} \mid u_d, X_{dj}]\right) = X_{dj}^t \beta + Z_{dj}^t u_d, \quad d = 1,\ldots,D;\ j = 1,\ldots,n_d, \tag{1}$$

where D is the number of small areas and $n = \sum_{d=1}^{D} n_d$ the sample size, with n_d being the small area size. For unit or individuum j of area d, $Y_{dj} \in ¡$ is the dependent

variable, $X_{dj} \in ¡^{p}$ the observable regressors (including the constant 1), β its coefficients (the fixed effects), $Z_{dj} \subseteq X_{dj}$ of dimension ñ , and $u_d \in ¡^{ñ}$ an unobservable area-specific effect. In practice, this is treated as a random effect and only its variance will be estimated. The reasoning can be different but the main, practical aspect, is that most of the n_d are small and D is large. A necessary assumption in the existing methodology is that u_d is independent from X_{dj}. This assumption is also made here but studied in detail in Lombardía and Sperlich (2006b). The function $G(\cdot)$ is the (known) link function and $g(\cdot)$ its unique inverse. When G is the identity, then model (1) includes the three classical versions: the nested-error regression (with $Z_{dj} = 1$, $u_d \in ¡$), the random regression coefficient (with $Z_{dj} = X_{dj}$), and the Fay-Herriot model (with only area-specific information, i.e. $X_{dj} = X_{di} \ \forall i, j$). Prasad and Rao (1990) condensed them in one single model of type (1) and indicated formulas for common estimators of β and linear predictors. However, the random regression coefficient has been treated in much less detail and is probably also the one least used.

Due to the effectiveness of mixed models, in particular the nested-error regression model, they are broadly applied in different fields of statistical analysis, including biology, medical research and surveys. Examples and overviews of this vast topic are provided by Searle et al. (1982), Breslow and Clayton (1993), Lee and Nelder (1996), Malec et al. (1997), Ghosh et al. (1998), McCulloch and Searle (2001). Further examples and explanations, but in particular different approaches to the nontrivial problem of implementation are provided, among others, by Fahrmeir and Tutz (2001), Diggle et al. (1998), and most recently by Skrondal and Rabe-Hesketh (2005). They have been extended to nonlinear parametric mixed models; see Kuhn and Lavielle (2005).

Quite recently, mixed effect models have entered the world of non- and semiparametric statistics. A first step was to separate the nonparametric functional into a deterministic (fixed effects) and a random part (random effects). Then, the smoothing parameter of a spline or sieve estimator can be written in terms of the variances of the random effect and error term; further extensions followed immediately, for more details see Ruppert et al. (2003) or Wand (2003). Similar ideas appear in the random fields theory, see Hamilton (2001). So far this research concentrates mainly on the challenging development of feasible algorithms for non- and semiparametric mixed models using spline methods. Kneip et al. (2005) provide a series estimator for a partial linear model with time varying individual effects; while Verbyla et al. (1999) study longitudinal data from designed experiments with smoothing splines. However, in most of the cases, asymptotic theory is missing. The same holds for theory based suggestions of model specification tests in (generalized) mixed models. Finally, to our knowledge mixed models have not been combined with kernel smoothing methods although the major part of the existing asymptotic theory for non- and semiparametric statistics is based on kernel smoothing methods.

This article intends to show how the combination of kernel based methods and mixed effects models can open a huge variety of statistical methods for the analysis, e.g. small area problems. To this aim we consider a generalized partially linear model (see e.g. Severini and Staniswalis, 1994) but include random effects, i.e.

$$G\left(E[Y_{dj} \mid u_d, T_{dj}, X_{dj}]\right) = m(T_{dj}) + X_{dj}^t\beta + Z_{dj}^t u_d, \quad d = 1,\ldots,D;\ j = 1,\ldots,n_d. \tag{2}$$

with Y_{dj}, u_d, X_{dj}, Z_{dj}, β as above, and $T_{dj} \in ¡^q$ further covariates with a nonparametric impact function $m(\cdot): ¡^q \to ¡$. The random effects u_d are supposed to be i.i.d. with mean zero and a ñ×ñ matrix of co-variances σ_u^2. We denote by

$$\mu_{dj} = E[Y_{dj} \mid u_d, T_{dj}, X_{dj}] = g\left(m(T_{dj}) + X_{dj}^t\beta + Z_{dj}^t u_d\right)$$

the conditional expectation. Under this setup, let Y be the $n\times 1$ vector with elements Y_{dj}, X be the $p\times n$ matrix with rows X_{dj}, $U = \text{diag}\{u_d, d = 1,\ldots,D\}$ the $(ñD)\times D$ block matrix of random effects, and $Z = \text{diag}\{Z_d, d = 1,\ldots,D\}$ a $(ñD)\times n$ block matrix with blocks $Z_d = (Z_{d1},\ldots,Z_{dn_d})$. Note that for the most popular nested-error model this simplifies to $U = u$ with $u = (u_1,\ldots,u_D)^t$ and $Z = \text{diag}\{1_{n_d}^t, d = 1,\ldots,D\}$, where 1_{n_d} denotes a column vector of ones with size n_d. Then, in matrix notation, the model is $m(T) + X^t\beta + Z^tU$. Further we define the σ_{dj}^2s as the elements of the diagonal matrix $Var[Y \mid u, T, X]$, and set $V = Var[Y \mid T, X]$.

We will first introduce an estimation procedure for the model including asymptotic theory. Then, for the parametric part, confidence intervals and testing procedures can be derived directly. In contrast, for the nonparametric part $m(\cdot)$ statistical inference is much more sophisticated in theory and practice. The same holds for prediction intervals, where a lot of research is still going on even for purely parametric models. Therefore, a sensible first step is to check whether such an effort is justified. This means testing $m(\cdot)$ for significant nonlinearity. An extension to test significant deviations of $m(\cdot)$ from a fixed polynomial or just a constant is obvious. An additional strong motivation for our test is that inference in small areas is model based. Therefore, model selection and validation play a vital role in the model-based inference. Seriously misspecified models lead to erroneous conclusions. However, hypothesis testing in the general mixed model framework for small areas inference has hardly been investigated. Jiang et al. (2001) studied a generalization of the Pearson's χ^2 goodness-of-fit test, which is applied to a real data example with geographically small areas. Zhu and Fung (2004) investigated a test for heteroscedasticity within the framework of semiparametric mixed models, illustrated with the analysis of a longitudinal study.

We introduce a test for the parametric null hypothesis

$$H_0: \quad m(T) = c + T^t\gamma \quad \text{vs} \quad H_1: \quad m(T) \neq c + T^t\gamma \tag{3}$$

for any γ and c, i.e. a generalized linear mixed effects model versus the semiparametric alternative (2). Such a test was introduced by Härdle et al. (1998), among others, for the case when there are no random effects. It turns out that their theory carries over to our mixed effects model. This also holds true for the (nonparametric) bootstrap we will use to obtain reasonable critical values for the test statistic. Extensions to related bootstrap tests in these kind of models, e.g. like those proposed most recently by Härdle et al. (2004) or Rodríguez-Póo et al. (2004) are also

obvious.

The rest of the paper is organized as follows. In Section 2 we introduce the estimators for the semiparametric model (2), i.e. the parametric counterpart of the null hypothesis H_0, together with its asymptotic properties. In Section 3 we first introduce an estimator of the parametric model that is convoluted with a kernel and will be used in the test statistic to account for the bias present in the estimate of the semiparametric alternative. Then, we will introduce and discuss tests statistics and adequate bootstrap procedures, again together with its asymptotic behaviours. A simulation study in Section 4 shows the excellent behaviour of the tests even for moderate sample sizes. A real data application illustrates its feasibility in practical problems. Section 5 contains the conclusion. The lists of technical assumptions and definitions can be found in the Appendix.

2. Estimating the Semiparametric Model

Our aim is to study the relationship between $Y \in ¡$ and a set of explanatory variables $(T, X) \in ¡^{q+p}$, taking the random effects into account. To streamline the presentation we restrict to $\sigma_{dj}^2 = \sigma_e^2 \ \forall \ d = 1,\ldots,D, j = 1,\ldots,n_d$ and set $\theta = (vec(\sigma_u^2), \sigma_e^2)$, also called the variance components. Here, $vec(\sigma_u^2)$ means that if σ_u^2 is of dimension larger than 1×1, then we stack the elements into a vector. Heteroscedasticity is a straight forward extension but would so complicate the presentation as to obscure our basic ideas. In small area statistics, one typically assumes $D \to \infty$ at rate $O(n)$. Suppose we have a sample of $n = \sum_{d=1}^{D} n_d$ replicates $\{(Y_{dj}, T_{dj}, X_{dj})\}_{j=1,d=1}^{n_d,D}$. The conditional distribution of Y given the random effects u, T, and X is from the family of densities $\{f(Y \mid u, T, X; m, \delta) : \delta \in \Delta, m(T) \in M\}$ with M and Δ compact sets, where $\delta = (\beta, \theta)$. The observations are taken independently, (T, X) are taken from compact sets $\aleph \subset ¡^{p}$ and $\Im \subset ¡^{q}$ respectively, and are independent from the random effect. Finally, $p(u; \sigma_u^2)$ denotes the density of the random effects.

We will consider the profiled likelihood approach by Breslow and Clayton (1993). For this purpose, assume $m(\cdot)$ as being known and concentrate on the estimation of δ. They derived a penalized quasi likelihood (PQL) that is based on this criterion

$$f(\beta, u \mid Y, T, X; m, \theta) = \frac{\prod_{d=1}^{D}\prod_{j=1}^{n_d} f(Y_{dj} \mid u_d, T_{dj}, X_{dj}; m, \delta)\prod_{d=1}^{D} p(u_d; \sigma_u^2)}{\int\int \prod_{d=1}^{D}\prod_{j=1}^{n_d} f(Y_{dj} \mid u_d, T_{dj}, X_{dj}; m, \delta)\, p(u_d; \sigma_u^2) du d\beta}, \tag{4}$$

where it is enough to maximize the numerator, see Fahrmeir and Tutz (2001). This they modified to a profiled likelihood to get simultaneous estimates for the variance components. González-Manteiga et al. (2005) also started with the PQL but estimated the variance components from a linearized version of the generalized linear model, going back to the Schall (1991) idea. Following these ideas, one considers the log-likelihoods

$$\varphi_1(Y;m,\delta)=\sum_{d=1}^{D}\sum_{j=1}^{n_d}\log f(Y_{dj}\mid u_d,T_{dj},X_{dj};m,\delta), \tag{5}$$

$$\varphi_2(u;\sigma_u^2)=\sum_{d=1}^{D}\log p(u_d;\sigma_u^2) \tag{6}$$

and $$\varphi(Y,u;m,\delta)=\varphi_1(Y;m,\delta)+\varphi_2(u;\sigma_u^2). \tag{7}$$

To obtain an estimator of the nonparametric part we first have to fix a point t_0 on which we aim to estimate function $m(\cdot)$. We again consider (7) and take the empirical counterpart of

$$E\left[\log f(Y\mid u,T,X;m,\delta)+\log p(u;\sigma_u^2)\mid T=t_0\right]$$

which is, in terms of kernels $K_h(\cdot)$,

$$\varphi_s(Y;m,\delta)=\sum_{d=1}^{D}\sum_{j=1}^{n_d}K_h\left(t_0-T_{dj}\right)\log f(Y_{dj}\mid u_d,T_{dj},X_{dj};m(t_0),\delta)+\varphi_2(u;\sigma_u^2) \tag{8}$$

with $K_h(\cdot)$ a q-dimensional product kernel, and $h=(h_1,\ldots,h_q)$ the corresponding bandwidth vector. This is also called the smoothed likelihood function. Note that the conditioning on T has no impact on φ_2; that is why we do not convolute that part with the kernel function. Because T is assumed to be independent of u, one might also use the simplified smoothed likelihood by skipping $\varphi_2(u_d;\sigma_u^2)$ from φ_s:

$$\varphi_{ss}(Y;m,\delta)=\sum_{d=1}^{D}\sum_{j=1}^{n_d}K_h\left(t_0-T_{dj}\right)\log f(Y_{dj}\mid u_d,T_{dj},X_{dj};m(t_0),\delta). \tag{9}$$

The version based on the Breslow and Clayton (1993) approach may be the most popular, because of its availability and easy handling. In SAS for example it is implemented as procedure GLIMMIX. However, in the last few years the integral method, see (10), has also gained popularity; see e.g. procedure NLMIXED in SAS. For the integral method we could have defined likelihoods (5) to (9) based on the likelihood

$$f(Y\mid T,X;m,\delta)=\int f(Y_{dj}\mid u,T_{dj},X_{dj};m,\delta)\,p(u;\sigma_u^2)du. \tag{10}$$

Taking the logarithm again provides us with a possible objective function; see e.g. McCulloch and Searle (2001) for the parametric case.

To estimate the whole model (2) we have to combine the existing fully parametric likelihood approach for random effect models with a semiparametric regression. Therefore we continue with the profiled likelihood idea. Let m_δ denote a least favorable curve – for definition see Appendix – in M to take into account the nuisance

parameter $m(\cdot)$ when estimating δ. Then the $\hat{\delta}$ that maximizes $\varphi(Y,u;m_\delta,\delta)$ is a consistent and asymptotically efficient estimator for δ. Because m_δ is unknown, Severini and Wong (1992) (*hereinafter SW92*) established sufficient conditions for an estimator $\hat{m}_\delta$, their so called *Conditions NP*, also in the Appendix, so that maximizing $\varphi(Y,u;\hat{m}_\delta,\delta)$ yields a consistent and asymptotically efficient estimator for δ. They also provide conditions to guarantee that maximizing the smoothed log likelihood, compare above, is such a valid estimator. Rodríguez-Póo et al. (2003) (*hereinafter RSV03*) have extended the results of SW92 such that they apply to our context with $\varphi_s(Y;m,\delta)$ being the smoothed log likelihood in question. Summarizing, we get

Procedure A.

1. For a value t_0 and fixed δ estimate $m(t_0)$ as the solution of the problem

$$\hat{m}_\delta = \underset{\{m\in M\}}{argmax}\,\varphi_s(Y;m,\delta).$$

2. Estimate δ by

$$\delta = \underset{\{\delta\in\Delta\}}{argmax}\,\varphi(Y,u;\hat{m}_\delta,\delta). \tag{11}$$

3. Notice that the Breslow and Clayton (1993) approach predicts the u_d for all d simultaneously, whereas the integral approach, compare (10), does not.
4. With the estimates obtained in steps 1 and 2, set finally $\hat{m} = \hat{m}_\delta$.

Clearly, the first step is to get estimators for the least favourable curve. For many likelihoods this yields the Nadaraya Watson smoother. Other estimators are thinkable; the one given here corresponds to the one proposed by SW92, Severini, Staniswalis (1994), and RSV03 (each one in a different context). We conclude

Lemma 1. *Suppose $D\to\infty$ at rate $O(n)$ and assumptions [A.1] - [A.3], [B.1] - [B.3] and [N.1] - [N.2] from the Appendix hold. Then, maximizing the smoothed likelihood as given in step 1 of Procedure A provides a valid estimator of the least favourable curve.*

Denote the true parameter by δ_0. Given a *valid estimator* $\hat{m}\delta$, a consequence of SW92 and RSV03 is:

Corollary 1. *Let δ be the log likelihood estimate as given in step 2 of Procedure A. Then, under the assumptions of Lemma 1, as $n=\sum_{d=1}^{D} n_d$ tends to infinity*

$$\sqrt{n}(\delta-\delta_0)\xrightarrow[d]{}N\left(0,I_\delta^{-1}\right),$$

where I_δ is the (marginal) Fisher information matrix, i.e. is equal to

$$E_T\{E_{X,u,T}\left[\frac{\partial}{\partial\delta}l\left(Y;m,\delta\right)\frac{\partial}{\partial\delta^t}l\left(Y;m,\delta\right)\right]-E_{X,u}\left[\frac{\partial}{\partial\delta}l\left(Y;m,\delta\right)\frac{\partial}{\partial m}l\left(Y;m,\delta\right)|\,T\right]$$

$$\times E_{X,u}\left[\left(\frac{\partial}{\partial m} l(Y;m,\delta)\right)^2 | T\right]^{-1} E_{X,u}\left[\frac{\partial}{\partial m} l\left(Y;m,\delta\right)\frac{\partial}{\partial \delta^t} l\left(Y;m,\delta\right) | T\right]\}, \tag{12}$$

where $E_W[\cdot]$ *is the expectation with respect the variable* W. *Further,* $l\left(Y;m,\delta\right) = \log f(Y \mid u,T,X;m,\delta) + \log p(u;\sigma_u^2)$, *and*

$$\frac{\partial}{\partial \delta} l\left(Y;m,\delta\right) = \left(\frac{\partial}{\partial \beta_1} l\left(Y;m,\delta\right), \text{L}\ , \frac{\partial}{\partial \beta_p} l\left(Y;m,\delta\right), \frac{\partial}{\partial \sigma_u^2} l\left(Y;m,\delta\right), \frac{\partial}{\partial \sigma_e^2} l\left(Y;m,\delta\right)\right)^t.$$

As can be observed from this result, the semiparametric estimator achieves the semiparametric efficiency bound, see Newey (1994). In practice, the asymptotic variance can be approximated with the aid of the Hessian matrix that one obtains as a by-product from the maximum likelihood estimation. Note further that our model assumptions do not contain restrictions on the dependence structure between X and T.

Remark 1. *If we assume a link* $g(\cdot)$ *being the identity function considering the nested-error regression model, i.e.*

$$Y_{dj} = m(T_{dj}) + X_{dj}^t \beta + u_d + \varepsilon_{dj}, \quad d = 1,\ldots,D,\ j = 1,\ldots,n_d;$$

with u_d *and* ε_{dj} *defined as before, we get for the variance of* $\hat{\beta}$ *the Fisher information*

$$\begin{aligned} I_\beta &= E_T\left\{X^t V^{-1} X - E[X \mid T]^t V^{-1} E[X \mid T]\right\} \\ &= E_T\left\{(X - E[X \mid T])^t V^{-1} (X - E[X \mid T])\right\} \end{aligned}$$

which equals the variance of Robinson (1988) in the simple partial linear model.

The above mentioned articles also provide us with the asymptotic distribution of the final nonparametric estimate. To this end, define $h_{prod} = \prod_{j=1}^q h_j$ and $h_{max} = \max_{1\le j\le q} h_j$. Then we can state,

Corollary 2. *Using the same conditions as in Corollary 1,* t_0 *being from the interior of the support of* T, $p_T(\cdot)$ *its density function, and* $n = \sum_{d=1}^D n_d$ *going to infinity, we have*

$$\sqrt{n h_{prod}}\left(\hat{m}(t_0) - m(t_0) - B_m(t_0)\right) \xrightarrow{d} \text{N}\left(0, Var_m(t_0)\right),$$

with $B_m(t_0) = O(h_{max}^2)$ *and*

$$Var_m\left(t_0\right) = \frac{\int K(t)^2 dt}{p_T(t_0) E\left[\frac{\partial}{\partial m} l\left(Y;m,\delta_0\right)^2 | T = t_0\right]}. \tag{13}$$

Alternatively, one could estimate the variance components by moment methods, and applying local polynomial expansions instead of a simple kernel convolution, see

Lombardía and Sperlich (2006a).

Remark 2. *If the conditional distribution of Y belongs to the exponential family, then*

$$E\left[\frac{\partial}{\partial m} l(Y;m,\delta_0)^2 \mid T=t_0\right] = E\left[g'(X^t\beta_0 + Z^t u + m(t_0))^2\right] Var\left[Y \mid u, t_0, X\right]^{-1},$$

which gives an asymptotic variance $Var_m(t_0)$ of the form

$$\int K(t)^2 dt p_T^{-1}(t_0) Var\left[Y \mid u, t_0, X\right] g'\left(X^t\beta_0 + Z^t u + m(t_0)\right)^{-2}.$$

Furthermore, the bias $B_m(t_0)$ then becomes

$$\nu_2(K)\left[\frac{1}{2}\sum_{j=1}^{q} h_j^2 \frac{\partial^2}{\partial t_j^2} m(t_0) + \frac{1}{p_T(t_0)}\sum_{j=1}^{q} h_j^2 \frac{\partial m}{\partial t_j}(t_0)\frac{\partial p_T}{\partial t_j}(t_0)\right] + o(h_{max}^2),$$

where $\nu_2(K)$ is implicitly defined by $\int tt^t K(t)dt = \nu_2(K) I$, with I being the identity matrix, compare Fan et al. (1995).

3. Testing the Parametric versus the Semiparametric Model

As discussed in detail in the Introduction, model specification testing in this context has at least two driving motivations: a) inference in small area statistics is model based so that a correct model specification is essential; b) semiparametric inference is much more complicated and expensive than the parametric one. For these reasons we will consider a test checking parametric versus semiparametric modelling. More specifically, we consider the testing problem $H_0 : m(t) = c + t^t\gamma$ vs. $H_1 : m(t) \neq c + t^t\gamma$. The testing of other parametric specifications is analogous. Our test statistic is based on an almost direct comparison of the semiparametric estimate with the corresponding estimate in the parametric model. First note that for this purpose it is enough to have

$$\sup_{t_0 \in \Im} \left|\hat{m}(t_0) - m(t_0)\right| = O_p\left(\sqrt{\frac{\log n}{n h_{prod}}}\right),$$

see Rodríguez-Póo et al. (2004). It follows from Rodríguez-Póo et al. (2003) that this holds for our estimator introduced in Section 2.

Let us consider the null hypothesis that

$$G(E[Y_{dj} \mid u_d, T_{dj}, X_{dj}]) = c + T_{dj}^t\gamma + X_{dj}^t\beta + Z_{dj}^t u_d.$$

Here, the estimation problem is purely parametric. Set in the following $\gamma_c^t = (c, \gamma^t)$, and denote the log-likelihood estimators for this model by $(\hat{\gamma}_c, \hat{u}, \hat{\sigma}^2)$. Following the arguments of Härdle et al. (1998), a direct comparison of $\hat{m}(T)$ with $\hat{c} + T^t\hat{\gamma}$ may be misleading, because $\hat{m}(\cdot)$ has a smoothing bias which is typically non negligible. To avoid this effect, we add a bias to $\hat{c} + T^t\hat{\gamma}$ that will compensate for the bias of $\hat{m}(T)$:

Procedure B.

1. We build the artificial data set: $\{\tilde{Y}_{dj}, T_{dj}, X_{dj}\}$ with

$$\tilde{Y}_{dj} = g(\tilde{c} + T_{dj}^t \tilde{\gamma} + X_{dj}^t \tilde{\beta} + Z_{dj}^t \tilde{u}),$$

the parametric fit of $\mu_{dj} = E[Y_{dj} \mid u_d, T_{dj}, X_{dj}]$.

2. Repeat only the nonparametric step from Procedure A replacing all parametric unknowns by their estimates $\tilde{\delta}$ (and eventually $\tilde{u}$). E.g., using the likelihood (8) or (9) one would set

$$\tilde{m}(t_0) = \underset{\{m \in M\}}{argmax}\, \varphi_s(\tilde{Y}; m, \tilde{\delta}). \tag{14}$$

3. The resulting estimators we use for the direct comparison with its semiparametric analogy are therefore $(\tilde{m}, \tilde{u}, \tilde{\delta})$.

Then, under $H_0 : m(t) = c + t^t\gamma$, one will get $|\tilde{m}(t) - [\tilde{c} + t^t\tilde{\gamma} + B_m(t)]| = o_p(1)$, where $B_m(t)$ is the bias of $\hat{m}(t)$, and therefore $|\hat{m}(t) - \tilde{m}(t)| = o_p(1)$.

A most traditional testing approach would be based on the likelihood ratio. But this test does not work here because $\hat{m}$ and $\tilde{\delta}$ were calculated with different likelihood functions (smoothed and unsmoothed functions), see Härdle et al. (1998). Instead let us consider the following weighted and unweighted squared differences:

$$R_w = \sum_{d=1}^{D} \sum_{j=1}^{n_d} H(\mathrm{m}(t_{dj}), \delta) \left[\hat{m}(t_{dj}) - \tilde{m}(t_{dj}) + X_{dj}^t (\hat{\beta} - \tilde{\beta}) \right]^2 \pi(t_{dj}), \tag{15}$$

or just

$$R = \sum_{d=1}^{D} \sum_{j=1}^{n_d} \left[\hat{m}(t_{dj}) - \tilde{m}(t_{dj}) + X_{dj}^t (\hat{\beta} - \tilde{\beta}) \right]^2 \pi(t_{dj}), \tag{16}$$

with $\pi(\cdot)$ being a weight function chosen by the empirical researcher and

$$H(m(t_{dj}), \delta) = \frac{\partial}{\partial m} l\left(Y_{dj}; m, \delta\right)^2.$$

In semiparametric, these kinds of test statistics are quite popular. On the one hand they try to imitate the likelihood ratio in the semiparametric world, see Härdle et al. (1998) and Müller (2001). On the other hand, all expressions are automatically calculated during the estimation procedure. Alternative statistics, either for incorporating the random effects or for simplifying the computational expense, are discussed in Lombardía and Sperlich (2006a). The covariances between the $\{\hat{m}(t_{dj}) - \tilde{m}(t_{dj})\}$ are asymptotically negligible for different observations of t_{dj}. Due to this fact, the asymptotic distribution of our test statistic can be concluded from Härdle, et al. (1998):

Corollary 3. *Under the hypothesis $m(t) = c + t^t\gamma$, the previous assumptions and [A.4], it holds that*

$$v^{-1}(R_w - b) \xrightarrow{d} N(0,1), \text{ where} \tag{17}$$

$b = h_{prod}^{-1} \int K(t)^2 dt E[\pi(T) p^{-1}(T)] + o(1)$ and $v^2 = 2h_{prod}^{-1} \int K^{(2)}(t)^2 dt E[\pi(T)^2 p^{-1}(T)]$. *Here,* $K^{(2)}$ *refers to a two-fold convolution of kernel* K.

If we skip the weighting with $H(\cdot)$ in the test statistic (16), bias and variance will become

$$b = h_{prod}^{-1} \int K(t)^2 dt E\left[E[H^{-1}(m(T), \delta_0)\pi(T) p^{-1}(T) \right] + o(1) \quad \text{and}$$

$$v^2 = 2h_{prod}^{-1} \int K^{(2)}(t)^2 dt E\left[H^{-2}(m(T), \delta_0)\pi(T)^2 p^{-1}(T) \right].$$

It is well known in non and semiparametric that in practice, even though we can substitute estimates for the unknowns, the asymptotics do not even come close to the real finite sample distribution, see e.g. Hjellvik et al. (1998). Mostly, bootstrap or subsampling is used to simulate the critical values. Therefore, here we introduce a wild bootstrap version, which was first proposed by Härdle and Mammen (1993) in nonparametric setups. Liu (1988) studied the wild bootstrap under regression models with non-i.i.d. observations, and Härdle et al. (2004) discussed intensively different bootstrap methods for doing inference in generalized partial linear additive models. Our procedure works as follows:

Procedure C.

1. From the sample, calculate a consistent estimator $\hat{\theta} = (\hat{\sigma}_u^2, \hat{\sigma}_e^2)$ of $\theta = (\sigma_u^2, \sigma_e^2)$.
2. Generate D independent copies of a vector $w_1 \in ¡^{ñ}$ with $E[w_1] = 0$ and $E[w_1^2] = 1_{ñ}$ with subexponential tails; that is, for a constant C_1 it holds that $E[exp\{|w_1|/C_1\}] \le C_1 1_{ñ}$ (c.f. [A.4]). Construct the vector $u^* = \sigma_u w_1$ such that the mean vector is zero and the variance covariance matrix is σ_u^2. Here, $\sigma_u = (\hat{\sigma}_u^2)^{1/2}$.
3. Generate n independent copies of w_2 such that $E[w_2] = 0$ and $E[w_2^2] = 1$ with subexponential tails (c.f. [A.4]). Set $e^* = \hat{\sigma}_e w_2$, which is independent of u^*, has mean vector zero, and variance $\hat{\sigma}_e^2$.
4. Under H_0 true, set

$$Y_{dj}^* = g\left(T_{dj}^t \gamma + X_{dj}^t \beta + Z_{dj}^t u_d^*\right) + e_{dj}^*, d = 1,\text{L}\ , D, j = 1,\text{L}\ , n_d.$$

5. Calculate the test statistic from the bootstrap sample (Y^*, X, T).

In the binary response Y_{dj} is a Bernoulli variable with parameter μ_{dj}. Hence, then it is reasonable to resample from the Bernoulli distribution with parameter $\hat{\mu}_{dj}$. E.g. for the special situation of the logistic semiparametric mixed model, we recommend the following parametric bootstrap:

Procedure C'.

1. From the sample, calculate a consistent estimator $\hat{\sigma}_u^2$ of σ_u^2.
2. Same as step 2 from Procedure C
3. Under H_0 true, generate observations by generating values of a binomial distribution with sizes n_{dj} and probabilities

$$\mu_{dj}^* = \frac{exp\{T_{dj}^t \hat{\alpha} + X_{dj}^t \hat{\beta} + Z_{dj}^t u_d^*\}}{1 + exp\{T_{dj}^t \hat{\alpha} + X_{dj}^t \hat{\beta} + Z_{dj}^t u_d^*\}} \quad d = 1,\text{K}\ ,D;\ j = 1,\text{K}\ ,n_d\ .$$

4. Calculate the test statistic from the bootstrap sample (Y^*, X, T).

The computation of quantiles of the distributions of $R_l = R$ or R_w, can be done by Monte Carlo: generate B independent sets of bootstrap samples $(Y^{*(b)}, X, T)$, $b = 1,\text{L}\ ,B$. The $(1-\alpha)$ quantiles of the distributions R_l can be estimated then the $\{[(1-\alpha)B]+1\}$th order statistic of $R_l^{*(b)} = R_l^*(Y^{*(b)}, X, T)$ $(b = 1,\text{L}\ ,B)$. Theorem 1 shows that the bootstrap procedure works:

Theorem 1. *Under the assumptions of Corollary 3, it holds for* $R_l = R$, $R_l = R_w$ *respectively, that*

$$d_k\left(F_{R_l^*}^*, F_{R_l}\right) \longrightarrow 0,$$

where F_{R_l} *is the distribution of* R_l, $F_{R_l^*}^*$ *is the conditional distribution of* R_l^* *(given the sample), and* d_k *is the Kolgomorov distance, which is defined as*

$$d_k(\nu, \tau) = sup_{\{t \in \text{i}\ \}} |\nu(X \le t) - \tau(X \le t)|$$

for two probability measures ν *and* τ *on the real line.*

Proof. The consistency of bootstrap methods is proved by imitation. For a general discussion of the validation of bootstrap methods see Shao and Tu (1995, pp.76). Using the same process as in Härdle et al. (1998, see proof of their Theorem 2 in the Appendix), taking into account the asymptotics results of the previous section and that $|Y_{dj}^*|$ has a bounded conditional Laplace transform (in a neighbourhood of 0). For more details see Mammen and van de Geer (1997, Section 5); these authors studied the asymptotic distribution of the parametric component of a regression model using the wild bootstrap. Thus, it holds

$$d_k\left(F_{R_l^*}^*, N(b, v^2)\right) \longrightarrow 0$$

in probability, with b and v^2 introduced in Corollary 3.

4. Finite Sample Performance

As mentioned in the previous sections, there are several papers that discuss algorithms for the estimation of semiparametric mixed effects models for spline and wavelet methods, different implementations, and likelihood approaches. Therefore, we have concentrated here more on the testing side and have stuck to a model relatively easy to estimate. Here, we study in detail the size and power of both tests, R_w and R in samples of moderate size with different bandwidths, and in particular the effect of

different designs. Afterwards, we applied our procedures, estimation and testing, to the real data example discussed in the introduction.

4.1. Simulation Study

For the simulation study we consider on the nested-error regression model because, the random coefficient model is hardly used in practice, and the Fay-Herriot model is a particular case of the nested-error one. The data generating process in our study was

$$y_{dj} = 1 + (1-a)t_{dj} + a\sin(\pi t_{dj}) + \beta^t x_{dj} + u_d + \varepsilon_{dj}, \tag{18}$$

for $d = 1,\ldots,D,, j = 1,\ldots,n_d$, where $\beta^t = (2,1)$, $(t_{dj}, x_{dj}^t) \in ¡^3$ i.i.d., $u_d : N(0,1)$ i.i.d., and $\varepsilon_{dj} : N(0,0.25)$ i.i.d., where $N(\mu,\sigma^2)$ is the normal distribution with mean μ and variance σ^2. We simulated the case where a runs from 0 (giving the null hypothesis model) to 0.5 to study the error of the first and second type. Further, for the explanatory variables (t_{dj}, x_{dj}^t) we simulated three different (always random) designs; first $U[0,2]^3$, second normal with mean 1.0, variance 0.6 but uncorrelated, and finally normal with mean 1.0, variance 0.6 but covariance 0.15. We did this because it is well known that non- and semiparametric inference is unfortunately strongly affected by the experimental design; in this context it obviously is of special interest to see the (expected) loss in power when we first change from uniform ("optimal" for nonparametric estimation) to normal, and second from uncorrelated to correlated designs.

Note that they all have the same mean, but in case of normal distribution about 10 to 20% of the observations fall outside of the $[0,2]^3$ cube. We studied two sample sizes, $n = 100$ and $n = 200$. When $n = 100$ we set $D = 10$ with n_1 to n_D equal to 5, 7, 8, 9, 10, 10 ,11, 12, 13, and 15. For $n = 200$ we set $D = 20$ and each of the above n_d occurred twice. We used always $B = 500$ bootstrap replications to estimate the critical values of the test statistic.

	U[0.2]^3			N(1,0.6)					
				Cov = 0.0			Cov = .15		
h_0	1.0	1.5	2.0	1.0	1.5	2.0	1.0	1.5	2.0
p	.481	.501	.526	.491	.554	.600	.502	.588	.639
1%	.012	.002	.002	.014	.004	.002	.014	.002	.002
5%	.068	.052	.024	.066	.034	.014	.078	.028	.016
10%	.110	.098	.072	.120	.066	.052	.130	.066	.034

Table 1. The p-values (p) and first error type at 1, 5, and 10% level for $n = 100$, $D = 10$.

Even though our smoother suffers from boundary effects, we did neither boundary corrections nor any trimming, i.e. we set $\pi(t) = 1$ throughout; instead, we trusted in the ability of the bootstrap to capture these effects adequately. For estimation we used the Epanechnikov kernel. The literature on bandwidths selection for nonparametric (kernel)

estimation is abundant but it is also well known that the optimal bandwidth for testing has a faster rate, i.e. should be slightly under-smoothing in practice. Although the cross validation bandwidth is asymptotically optimal for estimation rather then for testing, in our experience it behaves fairly well for testing problems with finite samples, probably because of its tendency to slightly under-smooth. Alternatively, there exists an increasing amount of literature on adaptive testing, i.e. choosing a bandwidth that maximizes the power of the test. However, these methods are only available for some specific testing problems, are difficult to implement and take a long time to compute. An approach that could probably be extended to our testing problem has been recently proposed by Rodríguez-Póo et al. (2004) and is based on the Spokoiny (2001) idea. The results presented were calculated with bandwidths $h = h_0 / n^{2/9}$ where $h_0 = 1.0$, 1.5, and 2.0 respectively. Note that, as in (18) the canonical link function is the identity, the test statistics R_w and R coincide.

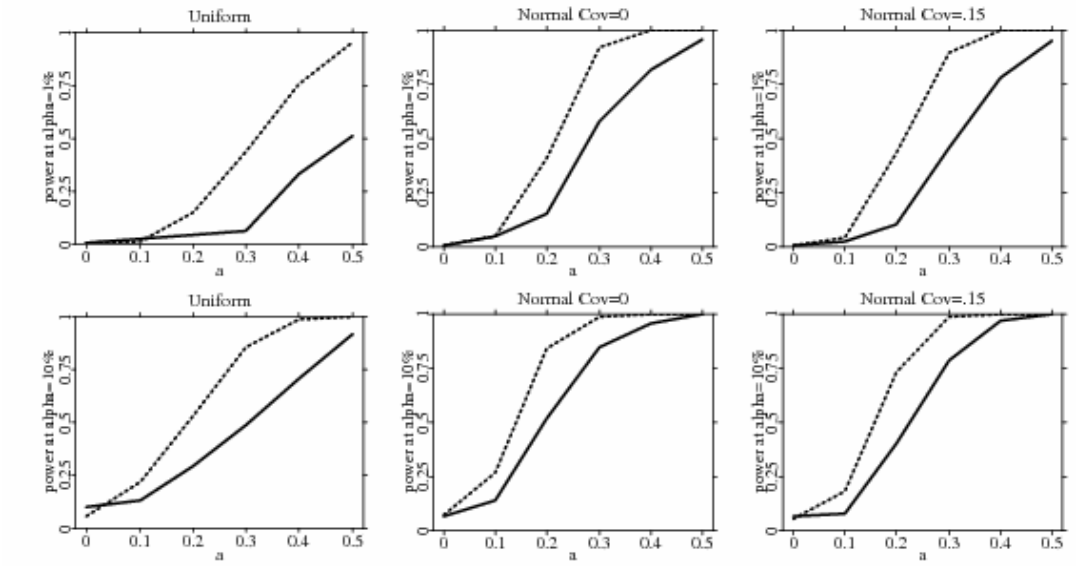

Figure1. The power of R for $n = 100$, $D = 10$ (solid), and $n = 200$, $D = 20$ (dashed) for different designs and rejection levels ($\alpha = 1\%$ upper line, $\alpha = 10\%$ lower line).

In Table 1 we provide the real $p-$value and rejection levels for different nominal levels under the null hypothesis of linearity of $m(\cdot)$, i.e. setting $a = 0$ in (18), calculated from 500 simulation runs. As expected, depending on the bandwidth, the rejection levels vary somewhat but due to the implemented bias reduction (compare discussion in Section 3, Procedure B) this test holds the level. It even tends to be conservative which at least is better than being too liberal in our opinion.

To study the power performance we determined the real rejection levels for a in model (18) running from 0 to 0.5, based on 100 simulation runs with $h_0 = 1.5$. In Figure 1 we plotted the power functions (solid line for $n = 100$, $D = 10$, dashed for $n = 200$, $D = 20$). From these plots we can draw several conclusions. Obviously, even for this rather small sample size of only 100 ($D = 10$) observations our tests work quite well detecting already moderate deviations from the null hypothesis. For $n = 200$, $D = 20$ we see that the size remains basically the same, the differences between the models change slightly, and the power improves (not surprisingly) considerably, especially for

the uniform design. Finally, the loss of power caused by introducing correlation in the design is moderate but visible. The case study with uniform design does, maybe surprisingly, no better than the one with normally uncorrelated distributed regressors. This is either due to the larger boundary effects the nonparametric estimator suffers from in the uniform case or due to the implicit weighting by the design density when in R we average over the data.

4.2. Real Data Application

For further illustration of the feasibility of our methods but also to show their usefulness in empirical studies, we now turn to the example mentioned in the Introduction, predicting the forested hectares in Galicia's 53 *comarcas* (administrative regions). Especially in the Mediterranean area, knowledge of the amount of forest (not used in forestry) is of great interest for environmental, tourism and health reasons. However, the measurement of forested hectares in a rocky and mountainous region like Galicia is quite expensive. All the data is from 1999, and was collected by Galician section of the Spanish National Statistical Institute (INE), and is accessible on the Internet. For the prediction we only wanted to use easily available indicators (quick and inexpensive). In this example we have the density of population (*popdens*) measured in persons by square kilometre, herbaceous cultivation (*herbac*) in hectares, (in Galician *cultivos herbáceos, barbeitos, hortas familiares*, including fields that are left fallow, and large family vegetable gardens) fruit farming (*fruit*) in hectares, and vineyards (*viney*) again in hectares. Note that there are no olive groves in Northern Spain (and therefore in Galicia) and so they are not included here. There are $D = 53$ comarcas (our small areas) with $2 \le n_d \le 12$, and in total $n = 314$ observations.

As in the simulation study, we concentrate mainly on the model specification test. The model under consideration is

$$y_{dj} = \beta^t x_{dj} + m(t_{dj}) + u_d + \varepsilon_{dj}, \tag{19}$$

for $d = 1,\ldots,53, \quad j = 1,\ldots,n_d$. In a first step we estimated (19) with t_{dj} being any one of the four covariates *popdens*, *herbac*, *fruit* or *viney* respectively, while the three remaining covariates entered the model linearly. Afterwards, in a second step we consider two dimensional t_{dj}.

For smoothing reasons, we used a simple bandwidth choice and a modified Epanechnikov kernel

$$K_h(t_0 - t_{dj}) = \frac{0.75}{h}\left\{1 - \left(\frac{\log(t_0+1) - \log(t_{dj}+1)}{h}\right)^2\right\}_+ .$$

This can be interpreted as estimating function $\omega(\cdot)$, where $\omega(\log t) = m(t)$, but with the original Epanechnikov kernel $K(z) = 0.75\left\{1 - z^2\right\}_+$. For bandwidth h we took for each element $t \in ¡$ of $t \in ¡^q$ the range of $\log(t+1)$ divided by a number h_d, see below.

In Figure 2 are given the nonparametric estimates of the four different models, i.e. in the upper left we see $\hat{m}(popdens)$ from model

$$y_{dj} = \beta_1 herbac + \beta_2\, fruit + \beta_3 viney + m(\,popdens) + u_d + \varepsilon_{dj}\,.$$

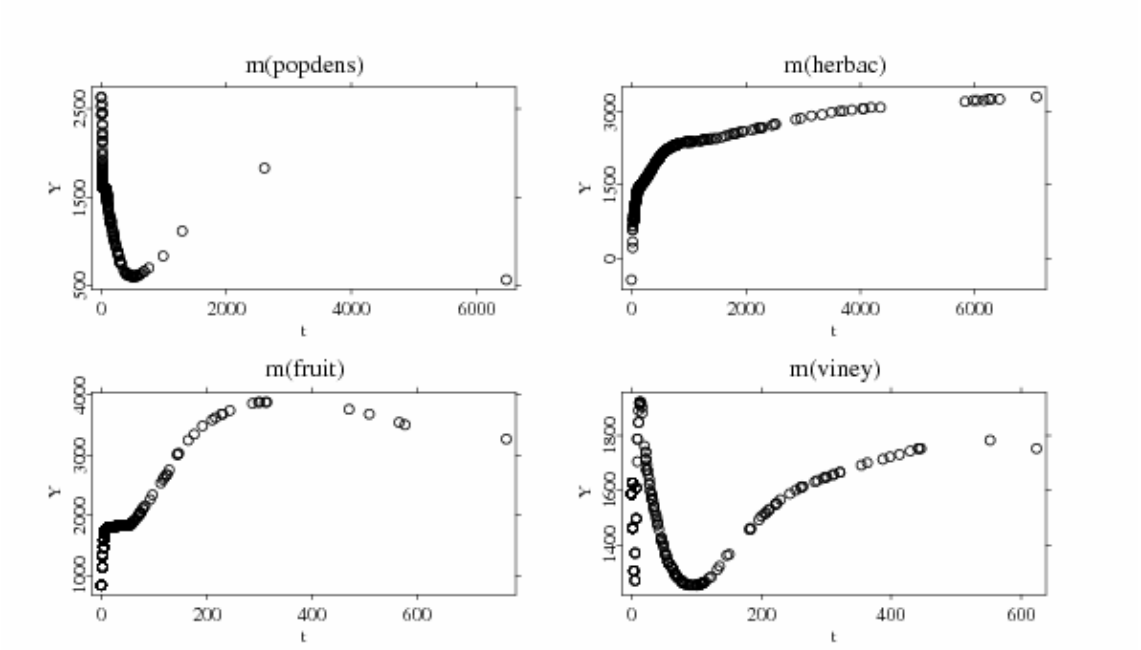

Figure 2. The nonparametric estimates of the four different models when only one variable enters non-parametrically, $h_d = 7$.

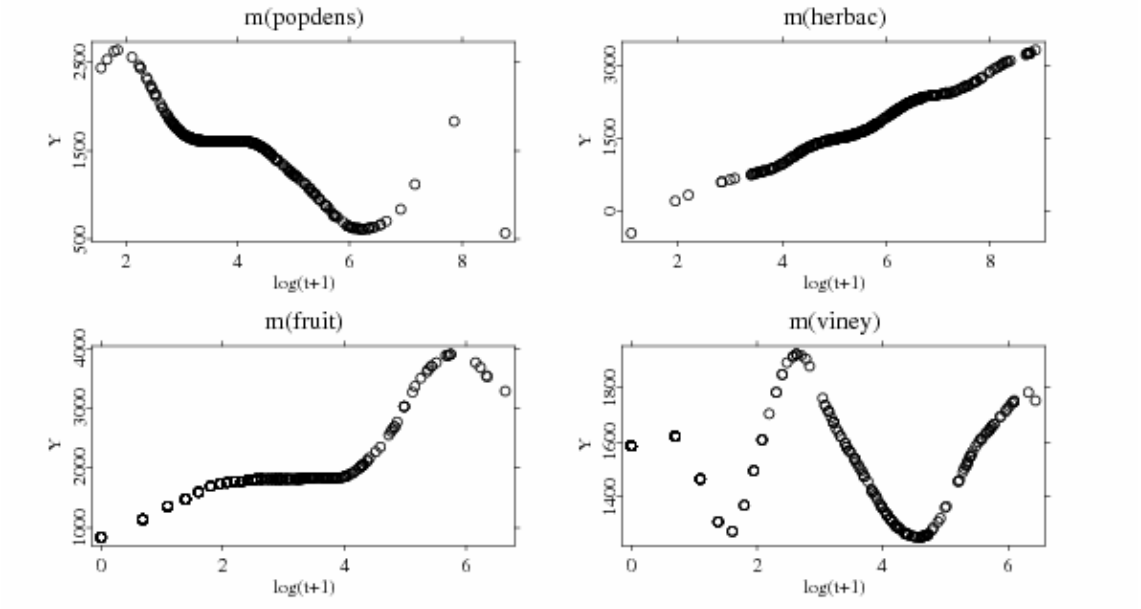

Figure 3. The nonparametric estimates of the four different models when only one variable enters non-parametrically versus $\log(t+1)$ for variable t, $h_d = 7$.

We have drawn the functions not as solid lines but in terms of circles indicating how the observations of covariate t are distributed over the range. As can be seen, none of them

looks really linear, but some of them might be log-linear. Therefore we plotted $\hat{m}(\cdot)$ also versus $\log(t+1)$, see Figure 3. These plots lead us to test $H_0^1 : m(t) = \gamma^t t$ and $H_0^2 : m(t) = \gamma^t \ln(t+1)$ against the nonparametric alternative for each covariate. We suspect that we will not reject H_0^2 for *herbac*. In addition it is likely that we will accept H_0^1 and H_0^2 for *viney* because the noise dominates there.

	$H_0^1 : m(t) = \gamma^t t$				$H_0^2 : m(t) = \gamma^t \ln(t+1)$			
$t = \backslash h_d =$	6	7	8	9	6	7	8	9
popdens	.000	.000	.000	.000	.104	.118	.130	.114
herbac	.000	.000	.000	.002	.716	.904	.918	.950
fruit	.000	.000	.000	.000	.002	.000	.000	.014
viney	.336	.282	.298	.332	.354	.302	.284	.346

Table 2. The p-values of H_0 when $t \in$ ¡ is either one of the four explanatory variables, assuming that the other covariates enter the model linearly. The critical values have been calculated by 500 bootstrap replications.

The results for these tests are given in Table 2. Since we don't have a data adaptive bandwidth choice searching the optimal bandwidth for each test, we simply repeated each test for various possible bandwidths. As for the estimation we took the range of $\ln(t+1)$ divided by (different) h_d. No matter what bandwidth we chose and whether we took a rejection level of 1, 5 or 10% our expectations were met except for *popdens* in H_0^2.

	$H_0^1 : m(t) = \gamma^t t$				$H_0^2 : m(t) = \gamma^t \ln(t+1)$			
$t = \backslash h_d =$	3	4	5	6	3	4	5	6
popdens , *fruit*	.004	.002	.000	.000	.148	.052	.016	.024
popdens , *viney*	.060	.014	.004	.004	.416	.350	.282	.250
fruit , *viney*	.024	.024	.006	.000	.024	.010	.004	.010

Table 3. The p-values of H_0 when $t \in$ ¡ ^{2} are either one of the combinations of explanatory variables shown in the left column. The other variables were assumed to enter the model linearly. The critical values have been calculated by 500 bootstrap replications.

As a final step we checked the linearity of two covariates jointly. For this we had to estimate a two dimensional functional $m(\cdot)$. This was quite problematic with only $n = 314$ observations, so we expected a serious loss of power. For the smoothing we simply took a product kernel from the above described modified Epanechnikov kernel. The bandwidth needs to be chosen larger for the two dimensional case, compare h_d from Table 2 and Table 3. In Table 3 we provide the results of selected combinations of the covariates which enter non-parametrically. That is, in the first line we tested whether

$m(popdens, fruit)$ is (log-)linear or not. As can be seen from the p-values, we rejected linearity for almost all combinations of covariates and bandwidths. In contrast, log-linearity cannot be rejected for $m(popdens, viney)$. For $m(popdens, fruit)$ the decision depends on the supposed smoothness, i.e. the used bandwidth and the rejection level.

5. Conclusions

We have introduced generalized partial linear mixed effect models which combine most of the known linear mixed effects models with flexible semiparametric modelling. For these kinds of models we first have derived estimation procedures and have afterwards used them to construct model specification tests. Finally, we have provided a bootstrap procedure to calculate the critical values of the test. However, this bootstrap procedure can certainly also be used for constructing confidence intervals or bands. Note that thanks to using the specific variance-covariance structure in mixed effects models, not only the parametric but also the nonparametric part is estimated efficiently.

Unlike most of the existing literature on semiparametric mixed effect models, we have been able to provide both feasible procedures and asymptotic theory for all our methods. The particular importance of these models and this kind of inference has been discussed for small area statistics although mixed effects models are also frequently used in many other fields. Our examples, especially the real data application, come from this field, as well. It is important to emphasize again that for any model-based inference the question of model specification is essential.

Our simulation study demonstrates the feasibility and the good behaviour of our quite complex methods. Even for small sample sizes the nonparametric smoothing is no problem, and the bootstrap-based test easily detects deviations from the null hypothesis. Correlations in the design turned out not to cause serious problems here. We concluded our findings with a real data application, also illustrating the usefulness and importance of our methods for empirical studies in practice.

Extensions to other estimation approaches (already studied in the parametric context), alternative test statistics, bootstrap procedures for heteroscedastic errors, and extensions to other models like generalized additive models, are deferred to Lombardía and Sperlich (2006a) for the sake of clarity and brevity of this paper.

Appendix

To provide our technical assumptions and definitions, we first introduce some more notations. We define

$$D^{\mu_x} a(x) = \frac{\partial^{|\mu_x|}}{\partial x_1^{\mu_1},\mathrm{L}\ ,\partial x_k^{\mu_k}} a(x),$$

being μ_x a k-vector of nonnegative integer constants, $|\mu_x| = \sum_{j=1}^{k} \mu_j$ and $a(x) \in ¡^{k}$ any function. We denote by $D_{m,\delta}^{r_m,s_\delta}(Y) = D^{r_m} D^{s_\delta} l(Y;m,\delta)$ and by $f_{m\delta}^{(r_m,s_\delta)}(y,u,x\,|\,t)$ the conditional density of $D_{m,\delta}^{r_m,s_\delta}(Y)$ given $T=t$. Let us define for each $\delta \in \Delta$ and $t \in \Im$

$$h(\delta,m,t) = E[l(Y;m,\delta)\,|\,T=t] \tag{20}$$

and $\bar{m}_\delta(t)$ the solution to

$$\frac{\partial}{\partial m} h(\delta, m, t) = 0, \tag{21}$$

with respect to m for each fixed δ and t.
Recall that $l(Y; m, \delta) = \log f(Y \mid u, T, X; m, \delta) + \log p(u; \sigma_u^2)$.

Next we provide the conditions in the case where we consider the family of density functions $\{ f(\cdot \mid u, t, x; m, \delta) : \delta \in \Delta, m \in M \}$.

A.1 For fixed but arbitrary $(m_1, \delta_1) \in M \times \Delta$, let

$$\rho(m, \delta) = \int l(y; m, \delta) f(y \mid u, t, x; m_1, \delta_1) dy,$$

with $(m, \delta) \in M \times \Delta$. If $\delta \neq \delta_1$ then $\rho(m, \delta) < \rho(m_1, \delta_1)$.

A.2 The matrix I_δ is positive definite for all $\delta \in \Delta$ and $m \in M$.

A.3 Assume that for vectors $| r_m | \leq 4$ and $| s_\delta | \leq 4$ such that $| r_m | + | s_\delta | \leq 4$ the function $D^{r_m} D^{s_\delta} l(Y; m, \delta)$ exists for almost all Y. Further, assume that

$$E\left\{ sup_\delta sup_m | D^{r_m} D^{s_\delta} l(Y; m, \delta) |^2 \right\} < \infty.$$

A.4 The Laplace transformation $E[exp\{t \mid Y_{dj} |\}]$ is finite for $t > 0$ small enough.

The condition [A.2] and [A.3] are usual in likelihood related problems. E.g. [A.3] allows differentiation and integration to be interchanged when differentiating

$$\rho(m, \delta) = \int l(y; m, \delta) f(y \mid u, t, x; m_1, \delta_1) dy.$$

The condition [A.4] is essential for the asymptotic expansions of Corollary 3, see Mammen and van de Geer (1997).

Next, we need to include some smoothness assumptions that are necessary because of the use of nonparametric smoothing methods:

B.1 For each $\delta \in \Delta$ and $t \in \Im$,

$$sup_{\{\delta, m, t\}} | D^{r_m} D^{s_\delta} D^{z_t} h(\delta, m, t) | < \infty$$

for $2 \leq | r_m | \leq 4$, $| s_\delta | \leq 2$, $| x_t | \leq 1$, and $| r_m | + | s_\delta | + | x_t | \leq 4$.

B.2 The solution to (21), $\bar{m}\delta(t)$, is unique and for any constant $\varepsilon > 0$ there exists another $\nu > 0$ such that

$$sup_\delta sup_t \left| \frac{\partial}{\partial m} h(\delta, \bar{m}_\delta(t), t) \right| \leq \nu$$

implies that

$$sup_{\delta} sup_t \left| \bar{m}_{\delta}(t) - m_{\delta}(t) \right| \leq \varepsilon.$$

B.3. Assume that

(a) $E[sup_m sup_{\delta} | D_{m,\delta}^{r_m, s_\delta}(Y) |] < \infty$ for $| r_m | \leq 5$ and $| s_\delta | \leq 3$,

(b) for some even integer $\xi \geq 10$ it holds that $sup_m sup_{\delta} E[| D_{m,\delta}^{r_m, s_\delta}(Y) |^q] < \infty$ for $| r_m | \leq 3$ and $| s_\delta | \leq 4$,

(c) $sup_m sup_{\delta} sup_{\{y,u,t,x\}} | f_{m\delta}^{(r_m, s_\delta)}(y,u,x \,|\, t) | < \infty$ for $| r_m | \leq 4$ and $| s_\delta | \leq 3$,

(d) $sup_t | D^{x_t} p(t) | < \infty$ and $sup_m sup_{\delta} sup_{\{y,u,t,x\}} | D^{x_t} f_{m\delta}(y,u,x \,|\, t) | < \infty$ for $| x_t | \leq a + 2$,

(e) and $0 < inf_{\{t \in \Im\}} p(t) < sup_{\{t \in \Im\}} p(t) < \infty$.

The assumptions [B.1] - [B.3] are sufficient to guarantee that the nonparametric estimator from step 1 of Procedures A and B fulfill $sup_{\{t_0 \in \Im\}} | \hat{m}(t_0) - m(t_0) | = o_p(n^{-1/4})$, and thus it is the estimator of least favourable curves.

Finally, we also need to impose some conditions on the kernel function $K(\cdot)$ and the bandwidth h:

N1. Function $K(\cdot)$ is a bounded kernel of order a with compact support, and

$$sup_z | D^{t_z} K(z) | < \infty \text{ for } | t_z | \leq a + 2 .$$

N2. The bandwidth vector h is of order $O(n^{-\alpha})$, $1/(4a) < \alpha < (\xi - 3)/4q(\xi + 6)$ such that $a/q > (\xi - 3)/(\xi + 6)$ with ξ from [B.3] b).

Note that as in RSV03 we consider here the use of higher order kernels to allow for higher dimensions of t. Otherwise, one could substitute conditions [N.1], [N.2] for the ones of SW92 in Lemma 8 and 9.

Definition of least favourable curve and valid estimators

A curve $\delta \to m_{\delta}$ with $m_0 = m_{\delta_0}$ is called *least favourable* if

$$E_0 \left[\frac{d}{d\delta} l(\delta, m_{\delta}) \Big|_{\delta = \delta_0} \right]^2 \leq E_0 \left[\frac{d}{d\delta} l(\delta, m_{1\delta}) \Big|_{\delta = \delta_0} \right]^2$$

for any other smooth curve $\delta \to m_{1\delta}$ with $m_{1\delta_0} = m_0$. For more details and a geometrical interpretation see SW92.

We next give the so called *Conditions NP*, i.e. conditions for being a *valid estimator* to get an efficient parametric estimate for δ. Here, $\hat{m}'$ indicates a derivative w.r.t. δ :

(a) On $\aleph$, $\hat{m}_{\delta}(x)$ converges for each x to a constant $m_{\delta}(x)$ for all $\delta \in \Delta$, $m_{\delta} \in M$, and for all $r, s = 0,1,2$, $r + s \leq 2$, $\frac{\partial^{r+s}}{\partial x^r \partial \delta^s} m_{\delta}(x)$ and $\frac{\partial^{r+s}}{\partial x^r \partial \delta^s} \hat{m}_{\delta}(x)$ exist. Further, $\| \hat{m}_0 - m_0 \| = o_p(n^{-\alpha})$, $\| \hat{m}'_0 - m'_0 \| = o_p(n^{-\beta})$, where $\alpha + \beta \geq 0.5$ and

$\alpha \geq 0.25$. Furthermore, $\sup_{\delta \in \Delta} \left\| \hat{m}_\delta - m_\delta \right\|$, $\sup_{\delta \in \Delta} \left\| \hat{m}'_\delta - m'_\delta \right\|$, $\sup_{\delta \in \Delta} \left\| \hat{m}''_\delta - m''_\delta \right\|$ are all of order $o_p(1)$. Finally, for some $\delta > 0$, $\left\| \partial \hat{m}_0 / \partial x - \partial m_0 / \partial x \right\|$ and $\left\| \partial \hat{m}'_0 / \partial x - \partial m'_0 / \partial x \right\|$ are of order $o_p(n^{-\delta})$.

(b) The curve m_δ is a least favourable curve.

References

- BATTESE, G.E., R.M. HARTER AND W.A. FULLER. (1988) An Error Components Model for Prediction of County Crop Areas Using Survey and Satellite Data. *Journal of the American Statistical Association*, **83**: 28-36.
- BRESLOW, N.E. AND D.G. CLAYTON (1993) Approximate Inference in Generalized Linear Mixed Models. *Journal of American Statistical Association*, **88**: 9-25.
- DIGGLE, P.J., J.A. TAWN AND R.A. MOYEED (1998) Model-based Geostatistics (with discussion). *Applied Statistics*, **47**: 299-350.
- FAHRMEIR, L. AND G. TUTZ (2001) *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer Series in Statistics, 2nd edition, Springer-Verlag: New York.
- FAN, J., N. HECKMAN, AND M.P. WAND (1995) Local Polynomial Kernel Regression for Generalized Linear Models and Quasi-Likelihood Functions. *Journal of the American Statistical Association*, **90**: 141-150.
- GHOSH, M. AND J.N.K. RAO (1994) Small Area Estimation: an Appraisal. *Statistical Science*, **9**: 55-93 (with discussion).
- GHOSH, M., D. NATARAJAN, T.W.F. STROUD AND B.P. CARLIN (1998) Generalized Linear Models for Small-Area Estimation. *Journal of the American Statistical Association*, **93**: 273-282.
- GONZÁLEZ-MANTEIGA, W., M.J. LOMBARDÍA, I. MOLINA, D. MORALES AND L. SANTAMARÍA (2005) Estimation of the Mean Squared Error of Predictors of Small Area Linear Parameters Under a Logistic Mixed Model. *Computational Statistics & Data Analysis Journal*. In Press.
- 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 E. MAMMEN (1993) Testing Parametric Versus Nonparametric Regression. *The Annals of Statistics*, **21**: 1926-1947.
- HÄRDLE, W., E. MAMMEN, AND M. MÜLLER (1998) Testing Parametric Versus Semiparametric Modeling in Generalized Linear Models. *Journal of the American Statistical Association*, **93**: 1461-1473.
- HAMILTON, J.D. (2001) A Parametric Approach to Flexible Nonlinear Inference *Econometrica*, **69**: 537-573.
- HJELLVIK, V., Q. YAO, AND D. TJØSTHEIM (1998) Linearity testing using local polynomial approximation. *Journal of Statistical Planning and Inference*, **68**: 295-321.
- JIANG, J. AND P. LAHIRI (2006) Mixed Model Prediction and Small Area Estimation *Test*, **15**: 1-96.
- JIANG, J., P. LAHIRI AND C. WU (2001) A Generalization of the Pearson

χ^2 Goodnes-Of-Fit Test with Estimated Cell Frequencies *The Indian Journal of Statistics*, **63**: 260-276.

- KNEIP, A., R.C. SICKLES AND W. SONG (2005) On Estimation the Mixed Effects Models. *Technical Report, University of Bonn*.
- KUHN, E. AND M. LAVIELLE (2005) Maximum Likelihood Estimation in Nonlinear Mixed Effects Models. *Computational Statistics and Data Analysis*, In Press.
- LEE, Y. AND J.A. NELDER (1996) Hierarchical generalized linear models (with discussion) *Journal of the Royal Statistical Society. Series B*, : 619-678.
- LIU, R.Y. (1988) Bootstrap Procedures under some Non-I.I.D. Models. *The Annals of Statistics*, **58**: 1696-1708.
- LOMBARDÍA, M.J. AND S. SPERLICH (2006a) Notes on Semiparametric Inference in Small Areas. *Working Paper, Georg-August-Universität Göttingen*
- LOMBARDÍA, M.J. AND S. SPERLICH (2006b) Smooth Transition from Mixed Effects to Partial Linear and Fixed Effects Models. *Working Paper, Georg-August-Universität Göttingen*
- MALEC, D., J. SEDRANSK, C.L. MORIARITY AND F.B. LECLERE (1997) Small Area Inference for Binary Variables in the National Health Interview Survey. *Journal of the American Statistical Association*, **92**: 815-826.
- MAMMEN, E., S. VAN DE GEER (1997) Penalized Quasi-Likelihood Estimation in Partial Linear Models. *The Annals of Statistics*, **25**: 1014-1035.
- MCCULLOCH, C.E. AND S.R. SEARLE (2001) *Generalized, Linear, and Mixed Models*, John Wiley & Sons, Inc.
- MÜLLER, M. (2001) Estimation and Testing in Generalized Partial Linear Models - A Comparative Study. *Statistics and Computing*, **11**: 299-309.
- NEWEY, W.K. (1994) The Asymptotic Variance of Semiparametric Estimators. *Econometrica*, **62**: 1349-1382.
- PFEFFERMANN, D. AND C.H. BARNARD (1991) Some New Estimators for Small Area Means with Application to the Assessment of Farmland Values. *Journal of Business & Economic Statistics*, **9**: 73-83.
- 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.
- RODRÍGUEZ-PÓO, J.M., S. SPERLICH, AND P. VIEU (2003) Semiparametric Estimation of Weak and Strong Separable Models. *Econometric Theory*, **19**: 1008-1039
- RODRÍGUEZ-PÓO, J.M., S. SPERLICH, AND P. VIEU (2004) An Adaptive Specification Test for Semiparametric Models. *Working Paper, University Carlos III de Madrid, Spain.*
- RUPPERT, D., M.P. WAND, AND R.J. CARROLL (2003) *Semiparametric Regression*. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press.
- SEARLE, S.R., G. CASELLA, AND C.E. MCCULLOCH (1982) *Variance Components*. New-York. Wiley.
- SEVERINI, T.A. AND J.G. STANISWALIS (1994) Quasi-likelihood Estimation in Semiparametric Models. *Journal of the American Statistical*

Association, **89**: 501-511.

- SEVERINI, T.A. AND W.W. WONG (1992) Profile Likelihood and Conditionally Parametric Models. *The Annals of Statistics*, **4**: 1768-1802.
- SHAO, J. AND D. TU (1995) *The Jackknife and Bootstrap*. New-York. Springer.
- SKRONDAL, A. AND S. RABE-HESKETH (2005) Generalized Linear Latent and Mixed Models with Composite Links and Exploded Likelihoods. *http://www.gllamm.org/composite_conf.pdf*
- SPOKOINY, V. (2001) Data-driven Testing the Fit of Linear Models. *Mathematical methods of statistics*, **10**, 465-497.
- VERBEKE, G. AND G. MOLENBERGHS (2000) *Linear Mixed Models for Longitudinal Data*. New-York. Springer-Verlag.
- VERBYLA, A.P., B.R.CULLIS, M.G.KENWARD, AND S.J.WELHAM (1999) The Analysis of Designed Experiments and Longitudinal Data by using Smoothing Splines. *Journal of the Royal Statisticsl Society, Series C*, **48**, 269-311.
- WAND, M.P. (2003) Smoothing and Mixed Models. *Computational Statistics*, **18**: 223-249.
- ZHU, Z. AND W.K. FUNG (2004) Variance Component Testing in Semiparametric Mixed Models. *Journal of Multivariate Analysis*, **91**: 107-118.

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