Robust tests for the nonparametric component in semiparametric partly linear regression models

A. Bianco[1], G. Boente[2] and E. Martínez[1]

[1] Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Ciudad Universitaria, Pabellón 2. C1428EHA, Buenos Aires, Argentina.
[2] CONICET and Departamento de Matemática and Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Ciudad Universitaria, Pabellón 1. C1428EHA, Buenos Aires, Argentina.

Keywords: Partly Linear Models, Robust Estimation, Smoothing Techniques, Hypothesis testing, Asymptotic Properties.

1 Introduction

Let $(y_i, \mathbf{x}_i', t_i)'$ be independent observations such that $y_i \in I\!R$, $t_i \in I\!R$, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})' \in I\!R^p$ and

$$y_i = \boldsymbol{\beta}'\mathbf{x}_i + g(t_i) + \epsilon_i \qquad 1 \le i \le n\,, \tag{1}$$

where the errors ϵ_i are independent and independent of $(\mathbf{x}_i', t_i)'$. As in Speckman (1988), Linton (1995), He, Zhu and Fung (2001) and González Manteiga and Aneiro Pérez (2003), we will assume that for $1 \le j \le p$

$$x_{ij} = \phi_j(t_i) + z_{ij} \qquad 1 \le i \le n\,, \tag{2}$$

where the errors z_{ij} are independent and independent of t_i. Denote $\mathbf{z}_i = (z_{i1}, \dots, z_{ip})'$.

This model, which has been studied by several authors (see, for instance, Härdle, Liang and Gao (2000)), generalizes the linear model and is more flexible since it includes a nonparametric component. Model (1) can be a suitable choice when one suspects that the response y linearly depends on $\mathbf{x}$, but that it is nonlinearly related to t. In the context of hypothesis testing, Gao (1997) established a large sample theory for testing $H_{0\boldsymbol{\beta}} : \boldsymbol{\beta} = \mathbf{0}$ under model (1), while Härdle, Liang and Gao (2000) tested $H_{0g} : g = g_0$ too. Recently, González Manteiga and Aneiro Pérez (2002) studied the case of dependent errors.

2 Robust estimators

It is well known that, both in linear regression and in nonparametric regression, least squares estimators can be seriously affected by anomalous data. The same statement holds for partly linear models.

Let $(y, \mathbf{x}', t, \epsilon, \mathbf{z}')$ be a random vector with the same distribution as $(y_i, \mathbf{x}_i', t_i, \epsilon_i, \mathbf{z}_i')$ and assume that ϵ and z_j have symmetric distributions $F\left(\cdot/\sigma_o\right)$ and $G_j\left(\cdot/\sigma_{z_j}\right)$ respectively. Denote $\phi_o(t) = E(y_i|t_i = t)$ and $\boldsymbol{\phi}(t) = (\phi_1(t), \dots, \phi_p(t))'$. Then, we have $g(t) = \phi_o(t) - \boldsymbol{\beta}'\boldsymbol{\phi}(t)$ and hence $y - \phi_o(t) = \boldsymbol{\beta}'(\mathbf{x} - \boldsymbol{\phi}(t)) + \epsilon$. In a robust setting, $\phi_o(t)$ and $\boldsymbol{\phi}(t)$ will also be the robust conditional location functionals if the errors ϵ and z_j have a symmetric distribution.

Bianco and Boente (2003) introduced a three step robust procedure which can be described as follows:

- **Step 1:** Estimate $\phi_o(t)$ and $\phi_j(t)$ through a robust smoothing, as the local medians or local M–type estimates. Denote $\widehat{\phi}_o(t)$ and $\widehat{\phi}_j(t)$ these estimates and $\widehat{\boldsymbol{\phi}}(t) = \left(\widehat{\phi}_1(t), \dots, \widehat{\phi}_p(t)\right)'$.
- **Step 2:** Estimate the regression parameter by applying a robust regression estimate to the residuals $y_i - \widehat{\phi}_o(t_i)$ and $\mathbf{x}_i - \widehat{\boldsymbol{\phi}}(t_i)$. Let $\widehat{\boldsymbol{\beta}}$ denote the obtained estimator.

- **Step 3:** Define the estimate of the regression function g as $\widehat{g}(t,\widehat{\boldsymbol\beta}) = \widehat{\phi}_o(t) - \widehat{\boldsymbol\beta}'\widehat{\boldsymbol\phi}(t)$.

In Step 3, an alternative estimator of the regression function g can be obtained by robustly smoothing the residuals $y_i - \widehat{\boldsymbol\beta}'\mathbf{x}_i$. However, it should be expected that a different smoothing parameter than the one in Step 1 may be preferable, as the residuals $y_i - \widehat{\boldsymbol\beta}'\mathbf{x}_i$ have a smaller variability than the original variables y_i. This gives another possible estimate $\widehat{\widehat{g}}(t,\widehat{\boldsymbol\beta})$.

Bianco and Boente (2003) studied the asymptotic behavior of the estimate $\widehat{\boldsymbol\beta}$ defined, in Step 2, as any solution of

$$\sum_{i=1}^{n} \psi_1\left(\frac{\widehat{r}_i - \widehat{\boldsymbol\beta}'\widehat{\mathbf{z}}_i}{s_n}\right) w_2\left(\|\widehat{\mathbf{z}}_i\|\right)\widehat{\mathbf{z}}_i = 0, \tag{3}$$

with ψ_1 and w_2 a score and a weight function, respectively, $\widehat{r}_i = y_i - \widehat{\phi}_o(t_i)$, $\widehat{\mathbf{z}}_i = \mathbf{x}_i - \widehat{\boldsymbol\phi}(t_i)$ and s_n an estimate of the residuals scale. These authors showed that $\sqrt{n}(\widehat{\boldsymbol\beta}-\boldsymbol\beta)$ is asymptotically normally distributed under model (1).

3 Test Statistics

Beyond the importance of developping robust estimators in more general settings, the work on testing also deserves attention. An up–to–date review of robust hypothesis testing results can be found in He (2002).

In some situations, we are interested in finding out the impact of a covariable on the response variable y. That is, we would like to make inference on the regression function g. In this talk, we will focus on the problem of testing on the model (1) the following nonparametric hypothesis $H_{0\boldsymbol\beta} : g = g_o$, and we will propose a class of tests based on the three step robust procedure described above.

The asymptotic behavior of the test statistics, under the null hypothesis and contiguous alternatives, is derived heuristically. Through a Monte Carlo study we illustrate the performance of the proposed tests.

References

A. Bianco and G. Boente (2003). Robust estimators in semiparametric partly linear regression. To appear in *Journal of Statistical Planning and Inference*.

J. Gao (1997). Adaptive parametric test in a semiparametric regression model. *Communication in Statistics, Theory and Methods*, 26, 787-800.

W. González Manteiga and G. Aneiro Pérez (2003). Testing in partial linear regression models with dependent errors. *J. Nonparam. Statist.***15**, 93-111.

W. Härdle, H. Liang and J. Gao (2000). *Partially Linear Models.* Springer-Verlag.

X. He (2002). Robust Tests in Statistics – a Review of the Past Decade. To appear in *Estadística.*

X. He, Z. Zhu and W. Fung (2001). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Working Paper.*

O. Linton (1995). Second order approximation in the partially linear regression model.*Econometrica*, 63 , 1079-1112.

P. Speckman (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society*, Ser. B, 50, 413-436.