Robust bandwidth selectors in semiparametric partly linear regression models

D. Rodriguez[1] and G. Boente[2]

[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, Bandwidth.

Abstract

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

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

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. Least square estimators have been studied by several authors (see for instance, Härdle, Liang and Gao (2000)).

All these estimators, as nonparametric estimators, depend on a smoothing parameter that should be choosen by the practitioner. As it is well known, large bandwidths produce estimators with small variance but high bias, while small values produce more wiggly curves. This trade–off between bias and variance lead to several proposals to select the smoothing parameter, such as cross-validation procedures and plug–in methods. Linton (1995), using local polynomial regression estimators, obtained an asymptotic expression for the optimal bandwidth in the sense that it minimizes a second order approximation of the mean square error of the least squares estimate, $\widehat{\beta}_{\mathrm{LS}}(h)$, of $\boldsymbol{\beta}$. This expression depends of the regression function we are estimating and on parameters which are unknown, as the standard deviation of the errors.

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. To avoid that problem, Bianco and Boente (2003) considered a three–step robust estimate for the regression parameter and the regression function. Besides, for the nonparametric regression setting, i.e., when $\boldsymbol{\beta} = 0$, the sensitivity of the classical bandwidth selectors to anomalous data was discussed by several authors, such as, Leung, Marrot and Wu (1993), Wang and Scott (1994), Boente, Fraiman and Meloche (1997) and Cantoni and Roncheti (2001).

In this talk, we will introduce a robust plug–in selector for the bandwidth, under a partly linear model (1) which converges to the optimal one and leads to robust data–driven estimates of the regression function g and the regression parameter $\boldsymbol{\beta}$. Through a Monte Carlo study, the behavior of the classical approach and of the resistant selectors is compared under normality and contamination.

References

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

G. Boente, R. Fraiman and J. Meloche (1997). Robust plug-in bandwidth estimators in nonparametric regression. *J. Statist. Planning and Inference*, 57, 109-142.

Cantoni, E. and Ronchetti, E. (2001). Resistant Selection of the Smoothing Paramater for Smoothing Splines, *Statistics and Computing*, **11**, 141-146.

H. Chen (1988). Convergence rates for parametric components in a partly linear model. *Ann. Statist.*, 16, 136-146

T. Gasser and H. Müller (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.*, 11, 171-185.

T. Gasser, T. and W. Härdle (1985). On robust kernel estimation of derivatives of regression functions. *Scand. J. Statist.*, 12, 233-240.

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

Leung, Marriott y Wu (1993). Bandwidth selection in robust smoothing. *J. Nonpar. Statist.*, **2**, 333-339

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.

Wang F. and Scott D. (1994). The L_1 method for robust nonparametric regression. *J. Amer. Stat. Assoc.*, 89, 65-76.