Robust estimators in partly linear models

A. Bianco [1] and G. Boente[2]

[1] Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Ciudad Universitaria, Pabellón 2. 1428, Buenos Aires, Argentina.
[2] Departamento de Matemática and Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Ciudad Universitaria, Pabellón 1. 1428, Buenos Aires, Argentina.

Keywords: Partly Linear Models, ARMA models, Robust Estimation, Smoothing Techniques.

1 Abstract

Statistical inference for a multidimensional random variables commonly focuses on functionals of its distribution that are either purely parametric or purely nonparametric. A reasonable parametric model produces precise inferences, while a badly misspecified one possibly leads to seriously misleading models. On the other hand, nonparametric modeling is associated both with greater stability and lesser precision.

Recently, nonparametric autoregression models have gained a lot of attention in order to explore the nature of complex nonlinear phenomena. Moreover, nonlinear autoregressive models manage to handle situations in which linear models, as the autoregressive (AR) model, fail to predict due to the nonlinear features that appear in many practical problems. For instance, an AR(p) model, is a process $\{y_t : t \geq 1\}$ satisfying $y_t = \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \epsilon_t$, with ϵ_t i.i.d. random variables, independent of $\{y_{t-j}, j \geq 1\}$, $E(\epsilon_t) = 0$ and $E|\epsilon_t|^2 < \infty$. As mentioned in Gao and Yee (2000), in econometrical problems one way to solve nonlinearity is to consider non–gaussian AR processes, for instance through ARCH models. An alternative could be to use a fully nonparametric autoregressive model, $y_t = m(y_{t-1}, \ldots, y_{t-p}) + \epsilon_t$. However, this approach and either the additive approach can loose too much information about the relationship between y_t and $\{y_{t-i},\ i \geq 1\}$ and neglect a possible linear relation between y_t and any lag y_{t-k}. Thus, in this case, it may be sensible to fit a partly linear autoregressive model.

As in the i.i.d. setting, partly linear models were introduced for autoregressive models in order to combine the advantages of both parametric and nonparametric methods. To illustrate, we introduce the simplest partly linear model which considers a stationary process $\{y_t : t \geq 1\}$ such that

$$y_t = \beta y_{t-1} + g(y_{t-2}) + \epsilon_t \ , \tag{1}$$

with ϵ_t independent and identically distributed and ϵ_t independent of $\{y_{t-j},\ j \geq 1\}$, $E(\epsilon_t) = 0$ and $E|\epsilon_t|^2 < \infty$. The partly linear autoregressive model (1) is particularly important since it involves not only a linear autoregressive component, but a univariate smoothing which avoids the "curse of dimensionality".

Partly linear autoregression models (1) are more flexible than standard linear models since they have a parametric and a nonparametric component. They can be a suitable choice when one suspects that the dependence to the past cannot be adequately explained only through a linear autoregression of order p.

When dealing with independent observations, the partly linear regression model $y_t = \boldsymbol{\beta}'\mathbf{x}_t + g(v_t) + \epsilon_t$, has been studied by several authors. Gao and Anh (1999) considered semiparametric regression models under long–range dependent errors and studied convergence rates for the consistent estimates proposed.

Recently, for the autoregressive model $y_t = \beta y_{t-1} + g(y_{t-2}) + \epsilon_t$, Gao and Liang (1995) established the asymptotic normality of the least squares estimator of β. See also, Gao (1995, 1998), Liang (1996) and Gao and Yee (2000) for some other results. An extensive description of the different results obtained in partly linear regression and autoregression models can be found in Härdle, Liang and Gao (2000).

The sensitivity of the least squares estimates to a small fraction of outliers has been extensively described both in the purely parametric and in the nonparametric setting. For parlty linear regression and autoregression models robust methods, less sensitive to a single wild spike outlier, would be desirable.

The basic ideas from robust smoothing and from robust regression estimation were adapted to deal with the case of independent observations, using kernel weights or smoothing splines (see for instance, He and Shi (1996), Gao and Shi (1997) and Bianco and Boente (2001)). On the other hand, under an unspecified dependence structure, He, Zhu and Fung (2001) considered M–type estimates for repeated measurements after approximating the regression function by a regression spline.

Under the autoregression model (1), the effect of a single outlier is even worse than in the independent setting. Thus, it seems quite natural to estimate the conditional location functionals through a robust smoothing and the autoregression parameter by a robust regression estimator.

In this talk, we will discuss the approaches given in the litterature for partly linear models and remind the classical approach to partly linear autoregressive models. We will introduce a family of estimates with a more resistant behavior than those based on least squares. Finally, through a Monte Carlo study, we will compare the given proposal with the classical procedures, under the partly linear autoregressive model (1).

References

Bianco, A. and Boente, G. (2001). *Robust estimators in semiparametric partly linear regression.* Submitted.

Gao, J. (1995). Asymptotic properties of some estimators for partly linear stationary autoregressive models. *Commun. Statist.–Theory and Methods*, **24**, 2011-2026.

Gao, J. (1998). Semiparametric regression smoothing of nonlinear time series. *Scandinavian J. Statist.*, **25**, 521-539.

Gao, J. and Anh, V. (1999). Semiparametric regression with long–range dependent error processes. *J. Statist. Planning and Inference*, **80**, 37–57.

Gao, J. and Liang, H. (1995). Asymptotic normality of pseudo-LS estimator for partly linear autoregression models. *Statist. and Prob. Letters*, **23**, 27-34.

Gao, J. and Shi, P. (1997). M–type smoothing splines in nonparametric and semiparametric regression models. *Statistica Sinica*, **7**, 1155-1169.

Gao, J. and Yee, T. (2000). Adaptive estimation in partly linear autoregressive models. *Canad. J. Statist.*, **28**, 571–586.

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

He, X. and Shi, P. (1996). Bivariate tensor–product B–spline in a partly linear model. *J. Multivariate Anal.*, **58**, 162-181.

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

Please fill in this form and mail it together with your abstract.

My abstract fits best to topic number 8

List of Topics:

1. Algorithms
2. Applications
3. Biostatistics
4. Computing and graphics
5. Data analysis
6. Data mining
7. Economics, finance
8. Efficiency and robustness
9. Functionals and bias
10. Fuzzy statistics
11. Geostatistics
12. Inference for robust methods, model testing
13. Location depth and regression depth
14. Multivariate methods
15. Neural networks
16. Rank-based methods
17. Regression quantiles, trimming
18. Robust covariance
19. Robust designs
20. Robust regression
21. Time series analysis
22. Wavelets