Nonparametric Regression in Time Series with Errors-in-Variables

Dimitris Ioannides[1] and Philippos Alevizos[2]

[1] Department of Economics
University of Macedonia
54006 Thessaloniki, Greece
(e-mail: dimioan@uom.gr)

[2] Department of Mathematics
University of Patras
26500 Patras, Greece
(e-mail: philipos@math.upatras.gr)

Abstract. In this paper, we study the nonparametric estimation of the regression function for dependent data with measurement errors in responses and covariates. The usual assumption in the errors-in-variables problem of indepedent errors can be replaced by dependent errors when the data are time series. Both cases are examined, and it is considered for first time the effect of measurement errors in responses when we are estimating nonparametrically the regression function.
Keywords: Deconvolution, nonparametric estimation, $\alpha-$mixing, noisy observations, regression function, uniform convergence.

1 Intoduction

Let $\{(X_i, Y_i)\}$, $i \geq 1$, be a strictly stationary process, where (X_i, Y_i) takes values in $\mathbb{R}^d \times \mathbb{R}$, $d \geq 1$, and has probability density function (pdf) $f(x, y)$. Consider the deconvolution model

$$Z_i = Y_i + \eta_i, \quad \text{and} \quad S_i = X_i + \epsilon_i, \tag{1}$$

where the noise processes $\{\eta_i\}$, and $\{\epsilon_i\}$, $i \geq 1$, are independent of the processes $\{Y_i\}$ and $\{X_i\}$, $i \geq 1$, respectively. In addition, we assume that the marginal distributions of the noise processes $\{\eta_i\}$, $i \geq 1$, and $\{\epsilon_i\}$, $i \geq 1$, are known, and also the components $\epsilon_{i1}, ..., \epsilon_{id}$ of the random vector ϵ_i are indentically distributed according to a r.v. ϵ. Models of this type and the deconvolution problems to which they lead arise in a variety of contexts in economic statistics, biostatistics, and various other fields. For example, if $d = 1$ in (1), X_i may represent the true income of a household at time i measured with error ϵ_i, Y_i its expenditures for some good which is subject to the measurement error η_i, and S_i, Z_i its measured income and expenditures, respectively. The interested reader may find additional applications of this problem in [Carroll *et al.*, 1995].

On the basis of the observations $(Z_1, S_1), ..., (Z_n, S_n)$, the problem is that of providing nonparametric estimate of the kth conditional moment function

$m(k;x) = E(Y^k/X = x)$, where Y and X are distributed as the r.v.'s Y_i and X_i, respectively. For the special case, $k = 1$, this problem was extensively studied in the literature and also when the covariates X_i are measured with some noise (i.e. $\eta_i \equiv 0, \epsilon_i \neq 0$). See, for example, [Carroll *et al.*, 1995], [Fan and Masry, 1992] and [Ioannides and Alevizos, 1997].

Here, we investigate the more complicate deconvolution model defined as in (1).

If we were using a Nadaraya-Watson type estimator, this problem could not be solved since for $k = 1$ the noise could not be extracted from the responses. Instead to use a Nadaraya-Watson type estimator, we construct first an estimator $\hat{f}_n(y/x)$ for the conditional density of Y given X, $f_{Y/X}(y/x)$, on the basis of our osbervations $(Z_1, S_1), ..., (Z_n, S_n)$. Then one natural estimator of $m(k;x)$ is obtained if we integrating apropriate the quantity $y^k \hat{f}_n(y/x)$ with respect to y. Because the type of this estimator was first introduced for uncontaminated data by [Roussas, 1969], we call it Roussas's estimator. In order to construct an estimator for $f_{Y/X}(y/x)$, the introduction of some notation and related concepts is necessary. Let $\widetilde{\Phi}_{K_1}(t)$ and $\widetilde{\Phi}_{K_2}(\tau)$ be the Fourier transforms of the univariate kernel density functions $\widetilde{K}_1(x)$ and $\widetilde{K}_2(y)$, and let $\widetilde{\Phi}_{\epsilon}(t)$ and $\widetilde{\Phi}_{\eta}(\tau)$ be the characteristic functions of the noise variables ϵ and η, respectively. Then, as in [Fan, 1991], we define the corresponding deconvoluting kernel functions by the following relations,

$$\widetilde{W_{1n}}(u) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iut} \frac{\widetilde{\Phi}_{K_1}(t)}{\widetilde{\Phi}_{\epsilon}(\frac{t}{h_n})} dt, \quad \widetilde{W_{2n}}(v) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iv\tau} \frac{\widetilde{\Phi}_{K_2}(\tau)}{\widetilde{\Phi}_{\eta}(\frac{\tau}{h_n})} d\tau, \quad (2)$$

where $0 < h_n \downarrow 0$. Thus the deconvoluting nonparametric estimator for the conditional density function is given by:

$$\hat{f}_n(y/x) = \frac{\hat{f}_{2n}(x,y)}{\hat{f}_{1n}(x)}, \quad (3)$$

where $\hat{f}_{1n}(x) = \frac{1}{nh_n^d} \sum_{i=1}^n W_{1n}(\frac{x-S_i}{h_n})$ and
$\hat{f}_{2n}(x,y) = \frac{1}{nh_n^{d+1}} \sum_{i=1}^n W_{1n}(\frac{x-S_i}{h_n}) \widetilde{W}_{2n}(\frac{y-Z_i}{h_n})$ with $W_{1n}(x) = \prod_{j=1}^d \widetilde{W}_{1n}(x_j)$.

Consequently the Roussas's estimator for the kth conditional moment is defined as follows:

$$\begin{aligned} m_n(k;x) &= \int_{-B_n}^{B_n} y^k \hat{f}_n(y/x) dy \\ &= \frac{1}{h_n} \sum_{i=1}^n \frac{W_{1n}(\frac{x-S_i}{h_n}) \int_{-B_n}^{B_n} y^k \widetilde{W}_{2n}(\frac{y-Z_i}{h_n}) dy}{\sum_{i=1}^n W_{1n}(\frac{x-S_i}{h_n})}, \end{aligned} \quad (4)$$

where B_n goes to infinity as $n \to \infty$. We integrate the quantity $y^k \widetilde{W}_{2n}(\frac{y-Z_i}{h_n})$ from $-B_n$ to B_n, since this is not in general integrable. The proposed estimator can be used in certain prediction problems.

[Ioannides, 1999] proved that the modal regression estimator can be used for extracting the noises from both variables Y and X. [Hannan, 1963] and [Robinson, 1986] treating this problem in the case by which $m(1;x) = E(Y/X = x)$ is the simple linear regression model. This paper attempts to study this problem in a more general setting using the Roussas's estimator (4).

In most publications on nonparametric deconvolution problems, a distinction is made between the case, where the noise characteristic functions $\widetilde{\Phi}_\epsilon(t)$ and $\widetilde{\Phi}_\eta(\tau)$ decay for large $|t|$ and $|\tau|$ either algebraically (*ordinary smooth case*) or exponentially (*supersmooth case*).

In the case by which the noise variables follow an ordinary smooth distribution, one of our main results is that the rates of the uniform strong convergence for the Roussas's estimator in (4) is of order $\max\{(\frac{logn}{nh_n^{[(d+2\beta)+1+2\beta']}})^{\frac{1}{2}}, h_n\}$ with β and β' positive numbers greater than 1 denoted the degree of smoothness of the noise variables ϵ and η, respectively. See, also, Assumption (A5) in the Appendix. This rates is better than the rate $\max\{(\frac{logn}{nh_n^{2[(d+2\beta)+1+2\beta']}})^{\frac{1}{4}}, h_n\}$ found for the modal estimator in [Ioannides, 1999]. In the noiseless case our rate becomes of order $\max\{(\frac{logn}{nh_n^{d+1}})^{\frac{1}{2}}, h_n\}$, which is essentially the optimal rate of $\hat{f}_{2n}(x,y)$ for estimating the pdf $f(x,y)$, and it is slight weaker than the optimal rate $\max\{(\frac{logn}{nh_n^d})^{\frac{1}{2}}, h_n\}$, obtained by the Nadaraya-Watson estimator.

The case where the noise variable has a super smooth distribution can be treated similarly, but the a.s. convergence rate is expected to be of logarithmic order.

Another interesting aspect of this paper is that we are not dealing only with independent measurement errors, but in general we allow them to be dependent. Usually in nonparametric deconvolution problems it is assumed that the noise process cosists from i.i.d. r.v.'s, avoiding correlated noise as is considered by [Hannan, 1963] and [Robinson, 1986]. Assuming that the joint stochastic process $\{(X_i, \epsilon_i, Y_i, \eta_i)\}$, $i \geq 1$, is a strong mixing process and the noise process $\{(\epsilon_i, \eta_i)\}$, $i \geq 1$ either consists from i.i.d. r.v.'s or dependent indentical r.v.'s we are proving the strong consistency of Roussas's estimator with the above rates under some mixing conditions which are weaker for the i.i.d measurement errors case.

This paper is organized as follows. The main result, Theorem 3.1, is given in Section 3, while some preparatory lemmas are given in Section 2. All the asssumptions made in this paper are given at the end of the paper in Appendix.

2 Some preliminary results

Set

$$R_n^{B_n}(k;x) = \frac{1}{h_n^{d+1}} \sum_{i=1}^{n} W_{1n}(\frac{x-S_i}{h_n}) \int_{-B_n}^{B_n} y^k \widetilde{W}_{2n}(\frac{y-Z_i}{h_n}) dy, \qquad (5)$$

then the Roussas's estimator $m_n(k;x)$ can be written as

$$m_n(k;x) = \frac{R_n^B(k;x)}{\hat{f}_n(x)}.$$

Now, for $k > 0$, denote by C_k the Cube in IR^d which is the Cartesian product of d copies of $[-k,k]$. Then working similar as in [Roussas, 1990], dividing $[-k,k]$ into b_n subintervals each of length δ_n, and taking J_{nl}, $l = 1,...,N$ the sets into which C_k is divided. Let x_{nl} arbitrary points in J_{nl}. Pick k sufficiently large, so that $J \subset C_k$, J compact subinterval of IR^d. Then, clearly,

$$\begin{aligned} |m_n(k;x) - m(k;x)| \le |\hat{f}_n^{-1}|\{ & |ER_n(k;x) - ER_n^{B_n}(k;x)| \\ & + |R_n^{B_n}(k;x) - R_n^{B_n}(k;x_{nl})| \\ & + |ER_n^{B_n}(k;x) - ER_n^{B_n}(k;x_{nl})| \\ & + |ER_n^{B_n}(k;x) - R(k;x)| + |R(k;x)||\hat{f}_n(x) - f(x)| \\ & + |R_n^{B_n}(k;x_{nl}) - ER_n^{B_n}(k;x_{nl})|\}, \end{aligned} \qquad (6)$$

with $R(k;x) = \int_{\mathrm{IR}} y^k f(x,y) dy$, and
$R_n(k;x) = \frac{1}{h_n^{d+1}} \sum_{i=1}^n K_1(\frac{x-X_i}{h_n}) \int_{\mathrm{IR}} y^k \widetilde{K}_2(\frac{y-Y_i}{h_n}) dy$.

Lemma 2.1. Under Assumptions (A1)(ii)-(iv), and (A6)(ii), one has

$$|E\hat{R}_n(k;x) - E\hat{R}_n^{B_n}| \le cB_n^{k-s},$$

for all x in IR^d, and $s > k$.

Lemma 2.2. Under Assumptions (A2), and (A5), it holds: $|R_n^{B_n}(k;x) - R_n^{B_n}(k;x')|$ and $|E\hat{R}_n^{B_n}(k;x) - E\hat{R}_n^{B_n}(k;x')|$ are bounded by $c_1 B_n^k h_n^{-(d(\beta+1)+\beta'+2)} \sum_{i=1}^d |x_i - x_i'|$, for any x, $x' \in \mathrm{IR}^d$, and $c_1 > 0$.

Lemma 2.3. Under Assumptions (A1)(ii)-(iii), it holds

$$|E\hat{R}_n(k;x) - R(k;x)| \le c_2 h_n,$$

for all $x \in \mathrm{IR}^d$, and some $c_2 > 0$.

Lemma 2.4. (i) Under Assumptions (A1)(i)-(iv), (A4), (A5), and if the noise processes $\{\epsilon_i\}$ and $\{\eta_i\}$, $\{i \geq 1\}$, consists from i.i.d. r.v.'s, then

$$\lim_{n\to\infty} \sup_{\mathrm{IR}^d} nh_n^{d(1+2\beta)+1+2\beta'} Var(R_n^{B_n}(k;x)) < c',$$

for some $c' > 0$.

(ii) Under the additional Assumption (A1)(v), one has

$$\lim_{n\to\infty} \sup_{\mathrm{IR}^d} nh_n^{d(1+2\beta)+1+2\beta'} Var(R_n^{B_n}(k;x)) < c',$$

for some $c' > 0$.

Lemma 2.5. Under Assumptions (A1), (A4), (A5) and (A6) one has

$$|R_n^{B_n}(x_{nl}) - ER_n^{B_n}(x_{nl})| = O\left[(\frac{logn}{nh_n^{d(1+2\beta)+1+2\beta'}})^{\frac{1}{2}} \right], \quad \text{a.s.}$$

Lemma 2.6. Under Assumptions (A1), (A4), (A5) and (A6) one has

$$|\hat{f}_n(x_{nl}) - E\hat{f}_n(x_{nl})| = O\left[(\frac{logn}{nh_n^{d(1+2\beta)+1+2\beta']}})^{\frac{1}{2}} \right], \quad \text{a.s.}$$

3 Main Result

The main result of this paper is the following theorem whose proof is a consequence of the preliminary results established. More precisely, one has:

3.1 Theorem

Under Assumptions (A1)-(A6), then

$$sup_{x\in J}|m_n(k;x) - m(k;x)| \leq O(h_n) + O\left[(\frac{logn}{nh_n^{d(1+2\beta)+1+2\beta'}})^{\frac{1}{2}} \right], \quad \text{a.s.}$$

Proof: The proof follows from Lemmas 2.1-2.6, in conjunction with the relation (6) using the same technique as in [Roussas, 1990] and [Ioannides, 1999]. ■

4 Appendix

The basic assumptions under which the a.s. uniform convergence of $m_n(k;x)$ is established.

Assumption (A1)

(i) The process $\{(X_i, Y_i, \epsilon_i, \eta_i)\}$, $i \geq 1$, is strictly stationary.
(ii) The processes $\{(X_i, Y_i)\}$, $i \geq 1$, and $\{(\epsilon_i, \eta_i)\}$, $i \geq 1$, are independent.
(iii) The processes $\{\epsilon_i\}$, $i \geq 1$, and $\{\eta_i\}$, $i \geq 1$, are independent.
(iv) The process $\{(X_i, Y_i, \epsilon_i, \eta_i)\}$, $i \geq 1$, is $\alpha-$mixing with mixing coefficient $\alpha(i) = O(i^{-k})$, $k > 1 + \frac{2}{\delta}, \delta > 0$.
(v) The process $\{(X_i, Y_i, \epsilon_i, \eta_i)\}$, $i \geq 1$, is $\alpha-$mixing with mixing coefficient $\alpha(i)$ satisfying the requirement $\frac{1}{h_n^{(2d\beta+2\beta')}} \sum_{j=h_n^{-d-1}}^{\infty} \alpha(j)^{\frac{2}{2+\delta}} < \infty$, for $h_n \to 0$, and $c_n \to \infty$, as $n \to \infty$.

Assumption (A2)

(i) The probability density $f(x)$ of X satisfies the Lipschitz condition of order 1 on $\mathbb{R}^d$.
(ii) $\inf_{x \in J} |f(x)| > 0$, where J is a compact subset of $\mathbb{R}^d$.
(iii) The quantity $R(k;x)$ satisfies the Lipschitz condition of order 1 on $\mathbb{R}^d$.

Assumption (A3)

The kernel functions $\widetilde{K}_i(.)$, $i = 1, 2$ are bounded probability density functions on $\mathbb{R}$ with $\int_{\mathbb{R}} |u| \widetilde{K}_1(u) du < \infty$ and $\int_{\mathbb{R}} |v| \widetilde{K}_2(v) dv < \infty$.

Assumption (A4)

(i) The characteristic functions $\widetilde{\Phi}_\epsilon(t)$ and $\widetilde{\Phi}_\eta(\tau)$ satisfy $\widetilde{\Phi}_\epsilon(t) \neq 0, \quad \widetilde{\Phi}_\eta(\tau) \neq 0$ for all t and τ.
(ii) $|t|^\beta |\widetilde{\Phi}_\epsilon(t)| > d, \quad d > 0, \quad |t|^{\beta'} |\widetilde{\Phi}_\eta(\tau)| > d', \quad d' > 0$, for large t and τ, and for some positive β , β'.
(iii) The joint characteristic function of ϵ_1 and ϵ_j is ordinary smooth of order 2β.
(iv) The joint characteristic function of η_1 and η_j is ordinary smooth of order $2\beta'$.

Assumption (A5)

The characteristic functions $\widetilde{\Phi}_{K_1}(t)$ and $\widetilde{\Phi}_{K_2}(\tau)$ satisfy the requirements: $\int |t|^{1+\beta} \widetilde{\Phi}_{K_1}(t) dt < \infty, \quad \int |\tau|^{1+\beta'} \widetilde{\Phi}_{K_2}(\tau) d\tau < \infty$.

Assumption (A6)

(i) $E|Y|^s \leq \infty$, for $s > 1$, and $x \in E \subset \mathbb{R}^d$.
(ii) $sup_{x \in \mathbb{R}^d} [\int |y|^s f(x,y) dy] < \infty$, for some $s > 1$.

References

[Carroll *et al.*, 1995]R.J. Carroll, D. Ruppert, and L.A. Stefanski. *Measurement Error in Nonlinear Models.* Chapman Hall, Britain, 1995.

[Fan and Masry, 1992]J. Fan and E. Masry. Multivariate regression estimation with error-in-variables. *J. Multivariate Anal.*, pages 237–271, 1992.

[Fan, 1991]J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, pages 1257–1272, 1991.

[Hannan, 1963]E.J. Hannan. Regression for time series with errors of measurement. *Biometrica*, pages 293–302, 1963.

[Ioannides and Alevizos, 1997]D. Ioannides and P. Alevizos. Nonparametric regression with errors in variables and applications. *Statist. Probab. Letters*, pages 35–43, 1997.

[Ioannides, 1999]D. Ioannides. Estimating the conditional mode of a stationary process from noise observations. *Metrika*, pages 19–35, 1999.

[Robinson, 1986]P.M. Robinson. On the errors-in-variables problem for time series. *J. Multivariate Anal.*, pages 240–250, 1986.

[Roussas, 1969]G.G. Roussas. Nonparametric estimation of the transition distribution of a markov process. *Ann. of Math. Stat.*, pages 1386–1400, 1969.

[Roussas, 1990]G.G. Roussas. Nonparametric regression estimation under mixing conditions. *Stochastic Processes and their Applic.*, pages 107–116, 1990.