# Shape restricted nonparametric regression with overall noisy measurements

Georg Pflug *
Department of Statistics, University of Vienna
Universitaetsstrasse, A-1010 Wien, Austria
and Roger J-B Wets
Department of Mathematics, University of California, Davis

February 2012

**Abstract**

For a nonparametric regression problem with errors in variables, we consider a shape-restricted regression function estimate, which does not require the choice of bandwidth parameters. We demonstrate that this estimate is consistent for classes of regression function candidates, which are closed under the graph topology.

**Keywords.** nonparametric regression, error-in-variables, graph topology, constrained maximum likelihood, shape restrictions
**AMS classification.** 62G08, 62G20

## 1 Introduction

We consider a nonparametric error-in-variables regression model on the unit hypercube $[0,1]^m$, which is of the following form: There are (unobservable) i.i.d. random variables $U_i, i = 1, \ldots, \nu$ with unknown distribution $\mu$ on $[0,1]^m$. The functional relation $c$ between the independent variable $U$ and the dependent variable $c(U)$ can only be observed together with observation error. The observable data are $(X_i, Y_i) = (U_i + V_i, c(U_i) + W_i)$, where $(V_i, W_i)$ are i.i.d., independent of $U_i$ and stem from a bivariate distribution with density $h(v,w)$. The problem is to recover $c$ from the observations $(X_i, Y_i)$, $i = 1, \ldots, \nu$.

*georg.pflug@univie.ac.at

Our setup is nonparametric, i.e. no special parametric form of the (generalized) regression function $c$ is assumed. However, we assume that the possible regression functions stem from a collection $\mathcal{C}$ of candidate regression functions on $[0,1]^m$, which is endowed with the graph topology, which we define below.

The graph of a function $c \in \mathcal{C}$ is given by

$$\text{gph } c = \overline{\{(u, c(u)) : u \in [0,1]^m\}}$$

where the overline means closure: We suppose that there is a given translation invariant distance $d$ in $[0,1]^m \times \mathbb{R}$, which generates the same topology as the Euclidean distance. Translation invariance means that $d(z_1, z_2)$ depends only on $z_1 - z_2$.

The graph topology on $\mathcal{C}$ is induced by the set-convergence of the graphs of the functions that belong to $\mathcal{C}$, cf. Chapter 5 in Rockafellar/Wets [5]. Since we will assume that our potential choice of regression functions are uniformly bounded, these graphs are then closed subsets of a compact space, in which case the graph-topology corresponds to that induced by the Pompeiu-Hausdorff distance between the graphs of these functions defined as follows:

$$\begin{aligned} d(c_1, c_2) &= d_H(\text{gph } c_1, \text{gph } c_2) \\ &= \max[\sup_{z \in \text{gph } c_1} \inf_{w \in \text{gph } c_2} d(z,w), \sup_{z \in \text{gph } c_2} \inf_{w \in \text{gph } c_1} d(z,w)]. \end{aligned}$$

**Assumptions.**

(i) All functions $c \in \mathcal{C}$ are uniformly bounded, i.e. $|c(u)| \leq K$ and of bounded variation.

(ii) $\mathcal{C}$ is closed in graph topology, implying that $\mathcal{C}$ is compact in graph topology, see Section 4.E in Rockafellar/Wets [5].

(iii) The density of $(V, W)$ belongs to the family of functions, which are of the form

$$h_s(z) = \exp\Big( - s\gamma[d(z,0)] + b(s)\Big)$$

for $z \in \mathbb{R}^{m+1}$ and $\gamma$ being a known increasing function on $\mathbb{R}^+$, with $\gamma(0) = 0$ which is Lipschitz on each finite subinterval, i.e.

$$|\gamma(u) - \gamma(v)| \leq C_t |u - v|$$

for $0 \leq u, v \leq t$. Moreover, we require that $\lim_{v \to \infty} \frac{\gamma(v + r(v))}{\gamma(v)} = 1$ for every bounded function $r(v)$. The scaling parameter $s$ is an unknown nuisance parameter, but it is known that it lies in the finite interval $S = (0, s_{max}]$ with $s_{max} < \infty$. $b(s)$ is the norming constant.

(iv) Let $m_K(z) = \sup\{d(z,(u,y)) : u \in [0,1]^m, y \in [-K,K]\}$. Then

$$\int \gamma(m_K(z))\, d\bar{P}(z) < \infty$$

where $\bar{P}$ is the true distribution of $(X,Y)$ under the true values $\bar{\mu}$, $\bar{s}$ and $\bar{c}$.

By assumption, the true distribution $\bar{P}$ has density $f_{\bar{\mu},\bar{s},\bar{c}}(x,y)$ which is given by

$$f_{\bar{\mu},\bar{s},\bar{c}}(z) = \int_{[0,1]^m} \exp\Big( - \bar{s}\gamma[d(z,(u,\bar{c}(u))] + b(\bar{s})\Big)\, d\bar{\mu}(u). \tag{1}$$

**Identifiability.** The density $f$ has three parameters. We have to make sure that there is a one-to-one correspondence between a density from this family and the parameter set $\mu, s, c$. We show first, that

$$\lim_{z\to\infty} \log f_{\mu,s,c}(z)/\gamma(d(z,0)) = -s \tag{2}$$

for all measures $\mu$ and $c \in \mathcal{C}$. To prove (2), write $d(z,(u,c(u))) = d(z,0) + r(z,u)$. The function $r(z,u)$ is uniformly bounded by $\sup\{d(0,(u,y)) : u \in [0,1]^m, y \in [-K,K]\}$ and therefore by assumption (iii)

$$\gamma[d(z,(u,c(u)))]/\gamma[d(z,0)] \to 1$$

as $z \to \infty$, for all $u$ and all $c$, which entails (2). Thus the parameter $s$ and hence the density $h_s$ is fully determined by $f_{\mu,s,c}$.

It remains to show that also $\mu$ and $c$ are determined by $f_{\mu,s,c}$. The distribution of $(X,Y)$ is the convolution of the distributions of $(U,c(U))$ and $(V,W)$. Since the density $h_s$ of $(V,W)$ is known, the distribution of $(U,c(U))$ and hence that of $U$ and the function $c(u)$ is uniquely determined by the distribution of $(U+V, c(U)+W)$, i.e. the density $f_{\mu,s,c}$.

**The generalized regression estimate.** Given the sample $(X_i,Y_i)$ we estimate the log-likelihood at the regression function $c$ by

$$\bar{\Psi}^\nu(c) = \sup_{u_1,\ldots,u_\nu} \sup_{s\in S} \frac{1}{\nu}\sum_{i=1}^{\nu} \log \frac{1}{\nu}\left[\sum_{j=1}^{\nu} \exp\Big( - s\gamma[d((X_i,Y_i),(u_j,c(u_j)))]\Big)\right] + b(s).$$

Let $\mathcal{M}$ be the family of all probability measures on $[0,1]^m$ and let $\mathcal{M}^\nu$ be the subfamily of probability measures $\mu$, which are of the form $\mu = \frac{1}{\nu}\sum_{i=1}^{\nu} \delta_{u_i}$, where $\delta_u$ is the unit mass sitting at point $u$. With this definition we may write

$$\bar{\Psi}^\nu(c) = \sup_{\mu\in\mathcal{M}^\nu, s\in S} \int \log\left[\int_{[0,1]^m} \exp\Big( - s\gamma[d(z,(u,c(u))]\Big)\, d\mu(u)\right] dP^\nu(z) + b(s),$$

where $P^\nu$ is the empirical distribution of $(X_i,Y_i)$, $i = 1,\ldots,\nu$.

Our estimate for $c$ is then

$$\hat{c}^\nu = \text{argmax}\,\{\bar{\Psi}^\nu(c) : c \in \mathcal{C}\}. \tag{3}$$

Notice that we consider the whole argmax set as the estimate: Typically the argmax is not unique for finite $\nu$.

Our goal is to show the consistency of this estimate under the given assumptions. The closedness of $\mathcal{C}$ in graph topology is crucial. It prevents the sequence of estimates to have variation increasing to infinity. As illustration, consider the functions $u \mapsto \sin(\nu u)$ on $[0,1]$. These functions graphically converge to the set $[0,1] \times [-1,1]$, which is not a function. Thus the family of all continuous or all differentiable functions is not closed in graph topology and thus does not qualify for the candidate set $\mathcal{C}$.

Before stating the main result, we review the usual method for nonparametric regression. If $f(u,y)$ is the joint density of $(U_i, Y_i)$, then the mean regression, i.e. the conditional expectation of $Y$ given $U = u$ is

$$\bar{c}(u) = \mathbb{E}(Y|U=u) = \frac{\int y\, f(u,y)\, dy}{\int f(u,y)\, dy}.$$

The Nadaraya-Watson estimate uses a smoothing kernel $K$ for the estimation of the numerator and the denominator:

$$c^\nu(u) = \frac{(\nu h_\nu)^{-1} \sum_{i=1}^{\nu} Y_i K((u - U_i)/h_\nu))}{(\nu h_\nu)^{-1} \sum_{i=1}^{\nu} K((u - U_i)/h_\nu))}.$$

Here $K$ is a kernel function and $(h_\nu)$ is a bandwidth sequence. In the error-in-variables model, the values of $U_i$ are only observable together with noise, i.e. $X_i = U_i + V_i$. This implies that the pertaining characteristic functions (Fourier Transforms) satisfy $\phi_X = \phi_U \cdot \phi_V$ i.e. $\phi_U = \phi_X / \phi_V$. By applying the inverse Fourier Transform (IFT) one gets the *deconvolution kernel*

$$D_\nu(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx) \frac{\phi_K(t)}{\phi_V(t/h_\nu)}\, dt,$$

where $\phi_K$ is the Fourier Transform of the kernel $K$ and the error-in-variables regression kernel estimate is

$$\bar{c}^\nu(u) = \frac{(\nu h_\nu)^{-1} \sum_{i=1}^{\nu} Y_i D_\nu((u - X_i)/h_\nu))}{(\nu h_\nu)^{-1} \sum_{i=1}^{\nu} D_\nu((u - X_i)/h_\nu))}.$$

(see [3, 6, 4]). One of the drawbacks of kernel estimates is their dependency on the right choice of the bandwidth parameter $h_\nu$. This drawback is even more crucial for deconvolution kernels, since it is very time consuming to test several choices of $h_\nu$: Every time a new bandwidth is tried out, one has to recalculate the IFT anew.

Let's now return to our main concern and derive the consistency of our estimates.

**Theorem.** Under the given assumptions, the shape restricted ML-estimate $\hat{c}^\nu$ converges a.s. to the true value $\bar{c} \in \mathcal{C}$.

**Proof.**

Let $\psi_{\mu,s,c}(z) = \log f_{\mu,s,c}(z) = \log\Big[\int_{[0,1]^m} \exp\Big(-s\gamma[d(z,(u,c(u))] + b(s)\Big)\,d\mu(u)\Big]$ for $z \in [0,1]^m \times \mathbb{R}$. Evidently, $z \mapsto \psi_{c,\mu,s}(z)$ is continuous and bounded from above by $\exp\Big(b(s_{max})\Big)$. Using the concavity of log we may estimate $\psi_{c,\mu,s}$ from below by

$$\begin{aligned} -\psi_{\mu,s,c}(z) &\leq -\int_{[0,1]^m} \log\exp\Big(-s\gamma[d(z,(u,c(u))] + b(s)\Big)\,d\mu(u) \\ &= \int_{[0,1]^m} s\gamma[d(z,(u,c(u))]\,d\mu(u) + b(s). \end{aligned}$$

The function

$$g(z) = s_{max}\cdot\gamma(m_K(z)) + b(s_{max})$$

satisfies $-\psi_{\mu,s,c}(z) \leq g(z)$ uniformly in $\mu, s, c$ and is $\bar{P}$ -integrable by assumption (iv). Thus there is for every $\varepsilon > 0$ a bounded set $K_\epsilon$ in $[0,1]^m \times \mathbb{R}$ such that $|\psi_{\mu,s,c}(z)|$ is bounded on $K_\epsilon$ and

$$\int_{\mathbb{R}^{m+1}\setminus K_\epsilon} |\psi_{\mu,s,c}(z)|\,d\bar{P}(z) \leq \epsilon \tag{4}$$

for all $\mu \in \mathcal{M}$, $s \in S$, $c \in \mathcal{C}$. Introduce

$$\begin{aligned} \Psi^\nu(\mu,s,c) &:= \int \psi_{\mu,s,c}(z)\,dP^\nu(z) \\ &= \frac{1}{\nu}\sum_{i=1}^{\nu} \log\left[\int_{[0,1]^m} \exp\Big(-s\gamma[d((X_i,Y_i),(u,c(u))]\Big)\,d\mu(u)\right] + b(s). \end{aligned}$$

By the the a.s. weak convergence of $P^\nu$ to $\bar{P}$, and the integrability and continuity of $\psi_{c,\mu,s}$, it follows that $\Psi^\nu(c,\mu,s)$ converges a.s. pointwise to

$$\begin{aligned} \Psi(c,\mu,s) &= \int \psi_{c,\mu,s}(z)\,d\bar{P}(z) \\ &= \int \log\left[\int \exp(-s\gamma[d(z,(u,c(u))])\right]\,d\mu(u)\,d\bar{P}(z) + b(s). \end{aligned}$$

We show that for fixed $c$, the family $\{\psi_{\mu,s,c}(\cdot)\}$ fulfills a bracketing condition in the sense of the Blum-de Hardt Theorem (see the Appendix), from which follows that

$$\sup_{\mu\in\mathcal{M},s\in S} |\Psi^\nu(\mu,s,c) - \Psi(\mu,s,c)| \to 0 \text{ a.s.} \tag{5}$$

Fix an $\varepsilon$. By (4), we may restrict ourselves to $z \in K_\varepsilon$. Since the function $c$ is bounded, we may find a finite collection of Borel sets $\{B_j^{(\varepsilon)}\}$ covering $[0,1]^m$ such that for all $j$

$$\sup_{u\in B_j^{(\varepsilon)}} c(u) - \inf_{u\in B_j^{(\varepsilon)}} c(u) \leq \varepsilon.$$

Likewise we may find a finite collection $s_1 = s_{min} < s_2, \ldots < s_p = s_{max} \in S$, such that $|s_{i+1} - s_i| \leq \eta$ and $|b(s_{i+1}) - b(s_i)| \leq \eta$, where $\eta$ will be chosen later.

For $n > 1/\eta$, find a finite set of probability measures $\mathcal{M}_\eta = \{\mu_\ell\}$ on $[0,1]^m$ such that the measures of $B_j^{(\varepsilon)}$ take all possible values $k/n$, $k = 0, \ldots, n$. Then

$$\begin{aligned}\psi_{\mu,s,c}(z) &= \log \int \exp\Big( - s\gamma[d(z,(u,c(u)))]\Big)\, d\mu(u) + b(s) \\ &= \log \sum_j \int_{B_j^{(\varepsilon)}} \exp\Big( - \gamma[d(z,(u,c(u))]\Big)\, d\mu(u) + b(s).\end{aligned}$$

For this given $\psi_{\mu,s,c}$, find an element $\mu_\ell$ of $\mathcal{M}_\eta$ such that $|\mu(B_j^{(\varepsilon)}) - \mu_\ell(B_j^{(\varepsilon)})| \leq 2\eta$ and $s_i \leq s < s_{i+1}$. Define the upper function as

$$U_{\ell,i}(z) = \log \int \exp\Big( - s_i\gamma[d(z,(u,c(u))) - \eta)]\Big)\, d\mu_\ell(u) + K \cdot \eta + b(s_{i+1})$$

and the lower function

$$L_{\ell,i}(z) = \log \int \exp\Big( - s_{i+1}\gamma[d(z,(u,c(u))) + \eta]\Big)\, d\mu_\ell(u) - K \cdot \eta + b(s_i).$$

Obviously,

$$L_{\ell,i}(z) \leq \psi_{\mu,s,c}(z) \leq U_{\ell,i}(z).$$

By the Lemma in the Appendix,

$$\begin{aligned}U_{\ell,i}(z) - L_{\ell,i}(z) &\leq 2K\eta + \sup_{u,c(u)} |s_i\gamma[d(z,(u,c(u)) + \eta] - s_{i+1}\gamma[d(z,(u,c(u)) - \eta]| \\ &\quad + |b(s_{i+1}) - b(s_i)| \\ &\leq 2K\eta + 2C_t\eta + |s_{i+1} - s_i| \sup_{z \in K\epsilon} \gamma[m_K(z)] \\ &\leq \text{const.}\eta\end{aligned}$$

for $z \in K_\varepsilon$ with $t = \sup\{m_K(z) : z \in K_\varepsilon\}$. Thus, for an appropriate choice of $\eta$, $\int U_{\ell,i}(z) - L_{\ell,i}(z) dP(z) \leq \varepsilon$. There are only finitely many intervals of functions $[L_{\ell,i}, U_{\ell,i}]$ which cover all $\psi_{\mu,s,c}$. Therefore, (5) is shown.

We now argue pointwise for fixed $\omega$. Since $(\mu, s) \mapsto \Psi(\mu, s, c)$ is continuous, it follows that $\Psi^\nu(\cdot,\cdot,c)$ converges continuously to $\Psi(\cdot,\cdot,c)$ for fixed $c$ (see Basic Fact 1 in the Appendix). This implies that $\bar\Psi^\nu(c) := \sup_{\mu\in\mathcal{M}^\nu, s\in S} \Psi^\nu(\mu,s,c)$ converges pointwise to $\bar\Psi(c) := \sup_{\mu\in\mathcal{M}, s\in S} \Psi^\nu(\mu,s,c)$ (see Basic Fact 2 in the Appendix).

We show now that the functions $c \mapsto \bar\Psi^\nu(c)$ are equicontinuous. To this end, we show that for each pair $c_1$ and $c_2$, each $\mu_1 \in \mathcal{M}^\nu$ and each $s \in S$, there is a $\mu_2 \in \mathcal{M}^\nu$ such that for $z_i \in K_\varepsilon$

$$|\Psi^\nu_{c_1,\mu_1,s}(z_1,\ldots,z_\nu) - \Psi^\nu_{c_2,\mu_2,s}(z_1,\ldots,z_\nu)| \leq s_{max} \cdot C_t \cdot \varepsilon$$

with $\varepsilon = d(c_1, c_2)$: If $\mu_1 = \frac{1}{\nu}\sum_{i=1}^{\nu} \delta_{u_j^1}$ then choose $u_j^2$ such that

$$d((u_j^1, c_1(u_j^1)), (u_j^2, c_2(u_j^2))) \leq \varepsilon.$$

Then, using the Lemma in the Appendix,

$$\begin{aligned}
& |\frac{1}{\nu}\sum_{i=1}^{\nu} \log \frac{1}{\nu}\sum_{j=1}^{\nu} k \cdot \exp\Big( - s\gamma[d(z_i, (u_j^1, c_1(u_j^1)))]\Big) \\
& -\frac{1}{\nu}\sum_{i=1}^{\nu} \log \frac{1}{\nu}\sum_{j=1}^{\nu} k \cdot \exp\Big( - s\gamma[d(z_i, (u_j^2, c_2(u_j^2)))]\Big)| \\
\leq\ & \sup_i |\log \sum_{j=1}^{\nu} \exp\Big( - s\gamma[d(z_i, (u_j^1, c_1(u_j^1)))]\Big) - \log \sum_{j=1}^{\nu} \exp\Big( - s\gamma[d(z_i, (u_j^2, c_2(u_j^1)))]\Big)| \\
\leq\ & \sup_i \sup_j |s\gamma[d(z_i, (u_j^1, c_1(u_j^1)))] - s\gamma[d(z_i, (u_j^1, c_1(u_j^1)))]| \\
\leq\ & s_{max} \cdot C_t \cdot \varepsilon.
\end{aligned}$$

Therefore also

$$|\bar{\Psi}^{\nu}(c_1) - \bar{\Psi}^{\nu}(c_2)| = | \sup_{\mu \in \mathcal{M}^{\nu}, s \in S} \Psi(\mu, s, c_1) - \sup_{\mu \in \mathcal{M}^{\nu}, s \in S} \Psi(\mu, s, c_2)| \leq s_{max} \cdot C_t \cdot \varepsilon$$

uniformly for all $\nu$. Thus we have shown that $\bar{\Psi}^{\nu}(c)$ are equi-continuous and converge pointwise to $\Psi(c)$. By Basic Fact 3, $\bar{\Psi}$ converges continuously to $\Psi$. By the Basic Fact 4, $\text{limsup}_{\nu} \text{argmax}\ _{\mathcal{C}} \bar{\Psi}^{\nu} \subseteq \text{argmax}\ _{\mathcal{C}} \bar{\Psi}$.

Finally, notice that $\text{argmax}\ _{c \in \mathcal{C}} \bar{\Psi}(c)$ is a singleton and consists only of $\bar{c}$. By the proven identifiability, $f_{\bar{\mu}, \bar{s}, \bar{c}}$ is different from any other $f_{\mu, s, c}$ and the argmax of $\bar{\Psi}(c)$, i.e. the last component of argmax of $\Psi(\mu, s, c)$ is $\bar{c}$. This follows from the Gibb's inequality, stating that for any two densities $f_1$, $f_2$ which are not a.s. identical

$$\int f_1(x) \log(f_2(x))\, dx < \int f_1(x) \log(f_1(x))\, dx,$$

applied for

$$\begin{aligned}
f_1(z) &= \int \log \Big[ \int \exp\Big( - \bar{s}\gamma[d(z, (u, \bar{c}(u))] + b(\bar{s})\Big)\Big]\, du \\
\text{and } f_2(z) &= \int \log \Big[ \int \exp\Big( - s\gamma[d(z, (u, c(u))] + b(s)\Big)\Big]\, du.
\end{aligned}$$

□

# 2 Graphically closed sets of candidate regression functions

In this section, we identify a few classes of regression functions that are closed with respect to the graph topology. Of course, these are just examples and it's easy to see how such classes could easily be enriched. For example, if it is known, or at least one suspects, that the regression function one is estimating is convex on $[0, \alpha)$ and concave on $[\alpha, 1]$, one can rely on the first one of the lemmas below to conclude that the collection of such bounded functions is closed with respect to the graph topology. Of course, many other such combinations are possible.

**Lemma 1.**

The family $\mathcal{C}_c$ of uniformly bounded convex functions on $[0,1]^m$ is graphically closed.

**Proof.** Let $c^\nu$ be a sequence from $\mathcal{C}_c$ converging graphically to a bounded closed set $C$. We have to show that $C$ is the graph of a convex function. We show first that the $c^\nu$ converge pointwise. Suppose to the contrary. Then there is a point $x$, such that $\liminf c^\nu(x) = a < b = \limsup c^\nu(x)$. We may find a sequence $\nu_k$ such that $c^{\nu_k}(x) \to a$. Since $(x, b)$ is in the graph limit of $c^{\nu_k}$, there is a sequence of points $x_{\nu_k} \to x$ such that $c^{\nu_k}(x_{\nu_k}) \to b$. Without loss of generality we may assume that right from the beginning we found a sequence $x_\nu \to x$ such that $c^\nu(x) \to a$ and $c^\nu(x_\nu) \to b$. Now find a sequence $\beta_\nu > 1, \beta_\nu \to \infty$ slowly enough such that $y_\nu := x(1 - \beta_\nu) + \beta_\nu x_\nu \to x$. Notice that $x_\nu = y_\nu \frac{1}{\beta_\nu} + (1 - \frac{1}{\beta_\nu})x$. Hence by convexity

$$c^\nu(x_\nu) \leq \frac{1}{\beta_\nu} c^\nu(y_\nu) + (1 - \frac{1}{\beta_\nu}) c^\nu(x)$$

or equivalently

$$c^\nu(y_\nu) \geq \beta_\nu[c^\nu(x_\nu) - c^\nu(x)] + c^\nu(x).$$

By construction $c^\nu(x_\nu) - c^\nu(x)$ tends to $b - a > 0$ and $c^\nu(x)$ tends to $a$. Thus $c^\nu(y_\nu) \to \infty$, which leads to a contradiction. The proof is complete, since the pointwise limit of uniformly bounded convex functions is a bounded convex function. □

Define the graph of a monotonic function in $[0,1]^m \times \mathbb{R}$ as a closed set $C$, whose projection to $[0,1]^m$ is $[0,1]^m$ and which has the following property: There do not exist pairs $(x, a)$ and $(y, b)$ in $C \subset [0,1]^m \times \mathbb{R}$ such that $x < y$ and $a > b$. Here $x < y$ means that all components satisfy $x_i < y_i$. Define the family $\mathcal{C}_m$ of monotonic functions by this property.

**Lemma 2.**

The family $\mathcal{C}_m$ of monotonic functions on $[0,1]^m$ is graphically closed.

**Proof.** Let $c^\nu$ be a sequence from $\mathcal{C}_m$ converging graphically to a bounded closed set $C$. If $C$ is not the graph of a monotonic function, then there exist pairs $(x, a)$ and $(y, b)$ in $C$ such that $x < y$ and $a > b$. But then there must exist sequences $x_\nu \to x$ and $y_\nu \to y$ such that $c^\nu(x_\nu) \to a$ and $c^\nu(y_\nu) \to b$. There are neighborhoods $N_x$ of $x$ and $N_y$ of $y$ such that $x' < y'$ for all $x' \in N_x, y' \in N_y$. We may choose $x_\nu \in N_x, y_\nu \in N_y$ for $\nu$ large enough and hence for these $\nu$ one has $c^\nu(x_\nu) \leq c^\nu(y_\nu)$, which contradicts $a > b$. □

**Lemma 3.**

The family $\mathcal{C}_u$ of unimodal functions on [0,1] is graphically closed.

**Proof.** A function $c$ is unimodal, if there is an argument $x$ such that $c$ is monotonically increasing in $[0, x]$ and monotonically decreasing in $[x, 1]$. Any value $x$, for which this condition holds, is called a mode and the set of modes is denoted by $M(c)$.

We prove first the following result: If $c^\nu$ is a sequence of unimodal functions converging in graph topology to a set $C$, then if

$$M^* = \text{limsup} M(c^\nu) = \{x : \exists x_{\nu_k} \in M(c^{\nu_k})\},$$

then $C$ cannot contain two points $(x_1, y_1)$ and $(x_2, y_2)$ such that $x_1, x_2 \in M^*$ and $y_1 \neq y_2$.

Suppose the contrary. Then there are points $x_1 < x_2$ in $M^*$, such that $(x_1, y_1)$ and $(x_2, y_2)$ are in $C$ and $y_1 < y_2$ (w.l.o.g). There is a sequence $\nu_k$ for which there exist $x_{\nu_k} \in M(c^{\nu_k})$ such that $(x_{\nu_k}, c(x_{\nu_k})$ converges to $(x_1, y_1)$. By unimodality, $\sup_{x \geq x_{\nu_k}} c^{\nu_k}(x) \leq c(x_{\nu_k})$ and we cannot find a sequence $u_{\nu_k} \geq x_{\nu_k}$ such that $c(u^{\nu_k}) \to y_2 > \lim_\nu c(x_{\nu_k})$. This however contradicts the assumption. Since we have proved that $C \cap M^* \times \mathbb{R}$ contains only one value, the rest of the proof follows from the proof for monotonic functions. □

**Lemma 4.**

The family of Lipschitz functions $\mathcal{C}_L$ on $[0, 1]^m$ is graphically closed.

**Proof.** Suppose that $c^\nu$ are $L$-Lipschitz and that $c^\nu$ graphically converge to $c$. We have to show that $c$ is also $L$-Lipschitz. Let $x$ and $y$ be two points. If the graphical distance between $c^\nu$ and $c$ is smaller than $\varepsilon$, then there exist points $x^\nu$ and $y^\nu$ such that $\|x^\nu - x\| \leq \varepsilon$, $\|y^\nu - y\| \leq \varepsilon$, $|c^\nu(x^\nu) - c(x)| \leq \varepsilon$, $|c^\nu(y^\nu) - c(y)| \leq \varepsilon$. Then

$$|c(x) - c(y)| \leq |c^\nu(x^\nu) - c^\nu(y^\nu)| + 2\varepsilon \leq L\|x^\nu - y^\nu\| + 2\varepsilon \leq L\|x - y\| + 2(L + 1)\varepsilon.$$

Since $\varepsilon$ is arbitrary, the result follows. □

# 3 Examples

We have implemented the estimate (3) using the MATLAB function `fmincon` for the optimization. In our figures, we show the true regression function as dashed, while the estimated curve is the solid line

We compared the errors-in-variables model with the "classical" monotonic regression estimate, which does not assume errors in the regressor. Using the same data as in Fig. 2, this estimate turns out to be quite unsatisfactory if not lousy.

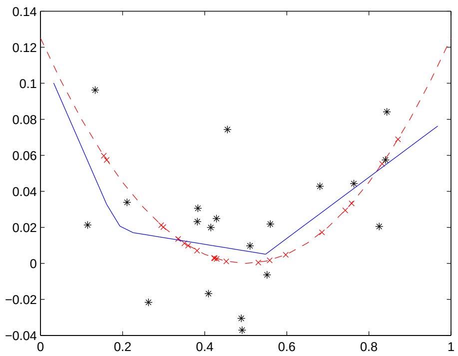


Figure 1: Convex regression: Here $\mathcal{C}$ is the family of all convex functions.

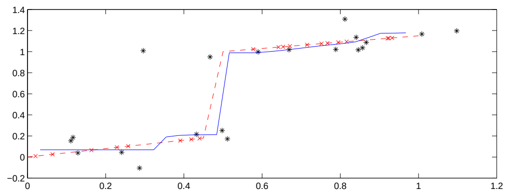

Figure 2: Monotonic regression: Here $\mathcal{C}$ is the family of monotonic functions.

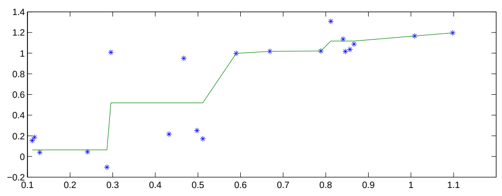

Figure 3: The same data as in Fig.2, but the regression estimate is here the classical monotonic regression estimate.

# References


[1] R.M. Dudley. A course on empirical processes (1984). Ecole d'Été de Probabilité de St. Flour XII *Lecture Notes in Math.* 1097, 1 – 142.

[2] Jianqing Fan and Young K. Truong (1993). Nonparametric Regression with Errors in Variables *Ann. Statist.* 21 (4), 1900–1925.

[3] D. A. Ioannides and P. D. Alevizos (1997). Nonparametric regression with errors in variables and applications *Statistics & Probability Letters.* 32 (1), 35–43.

[4] D. A. Ioannides and Eric Matzner-Loeber (2009). Regression quantiles with errors-in-variables *Journal of Nonparametric Statistics 21* (8), 1003 – 1015.

[5] R.T. Rockafellar and R. Wets (2009). *Variational Analysis.* Grundlehren der Mathematischen Wissenschaften 317, 2nd edition, Springer.

[6] S.M.Schennach (2004). Nonparametric regression in the presence of measurement errors. *Econometric Theory* 20, 1046–1093 (2004)

[7] S. Zwanzig. How to adjust nonparametric regression estimators to models with errors in the variables (1999). *Theory of Stochastic Processes* 5 (21) 272-277.


# Appendix

**Basic Facts about convergence of functions on metric spaces.**

Let $(\mathcal{M}, d)$ be some metric space.

1. Suppose that $\sup_{\mu\in\mathcal{M}} |\Psi^\nu(\mu) - \Psi(\mu)| \to 0$ and that $\Psi$ is continuous. Then $(\Psi^\nu)$ converges continuously to $\Psi$.

   Proof: Let $\mu^\nu \to \mu$. Then

$$\begin{aligned} |\Psi^\nu(\mu^\nu) - \Psi(\mu)| &\leq |\Psi^\nu(\mu^\nu) - \Psi(\mu^\nu)| + |\Psi(\mu^\nu) - \Psi(\mu)| \\ &\leq \sup_{\mu\in\mathcal{M}} |\Psi^\nu(\mu) - \Psi(\mu)| + |\Psi(\mu^\nu) - \Psi(\mu)| \to 0. \end{aligned}$$

2. Suppose that the upper semi-continuous (u.s.c.) functions $\Psi^\nu$ are defined on $\mathcal{M}^\nu$, which are increasing compact sets such that $\bigcup_\nu \mathcal{M}^\nu$ is dense in $\mathcal{M}$. Suppose that the u.s.c. function $\Psi$ is defined on $\mathcal{M}$ such that $(\Psi^\nu)$ converges continuously to $\Psi$ in the following sense: If $\mu^\nu \in \mathcal{M}^\nu$ converge to $\mu \in \mathcal{M}$, then $\Psi^\nu(\mu^\nu) \to \Psi(\mu)$. Then it follows that

$$\text{limsup}_\nu \text{argmax}\ _{\mathcal{M}^\nu}\Psi^\nu \subseteq \text{argmax}\ _{\mathcal{M}}\Psi$$

and

$$\lim_\nu \sup_{\mathcal{M}^\nu} \Psi^\nu = \sup_{\mathcal{M}} \Psi$$

as $\nu \to \infty$.

Proof: We show first that $\text{liminf}_\nu \sup_{\mathcal{M}^\nu} \Psi^\nu \geq \sup_{\mathcal{M}} \Psi$. Let $\Psi(\mu) \geq \sup_{\mathcal{M}} \Psi - \varepsilon$. Then there are $\mu^\nu \in \mathcal{M}^\nu$ such that $\mu^\nu \to \mu$. By continuous convergence

$$\text{liminf}_\nu \sup_{\mathcal{M}^\nu} \Psi^\nu \geq \lim_\nu \Psi^\nu(\mu^\nu) = \Psi(\mu) \geq \sup_{\mathcal{M}} \Psi - \varepsilon$$

and hence

$$\text{liminf}_\nu \sup_{\mathcal{M}^\nu} \Psi^\nu \geq \sup_{\mathcal{M}} \Psi.$$

Let now $\bar{\mu}^\nu \in \text{argmax}_{\mathcal{M}^\nu} \Psi^\nu$ and let $\bar{\mu}$ be a cluster point of $\bar{\mu}^\nu$, i.e. $\bar{\mu} = \lim_k \bar{\mu}^{\nu_k}$. By continuous convergence $\lim_k \Psi^{\nu_k}(\bar{\mu}^{\nu_k}) = \Psi(\bar{\mu})$ and by the previous result $\text{liminf}_k \Psi^{\nu_k}(\bar{\mu}^{\nu_k}) \geq \sup_{\mathcal{M}} \Psi$. Hence $\Psi(\bar{\mu}) \geq \sup_{\mathcal{M}} \Psi$, i.e. $\bar{\mu} \in \text{argmax}_{\mathcal{M}} \Psi$. In addition, $\lim_\nu \sup_{\mathcal{M}^\nu} \Psi^\nu = \Psi(\bar{\mu}) = \sup_{\mathcal{M}} \Psi$.

3. If $\bar{\Psi}^\nu$ are equi-continuous and $\bar{\Psi}^\nu$ converge pointwise to $\bar{\Psi}$, then continuous convergence holds.

   Proof: If $c^\nu \to c$, then

$$|\bar{\Psi}^\nu(c^\nu) - \bar{\Psi}(c)| \leq |\bar{\Psi}^\nu(c^\nu) - \bar{\Psi}^\nu(c)| + |\bar{\Psi}^\nu(c) - \bar{\Psi}(c)| \to 0.$$

4. Let $\bar{\Psi}^\nu(c)$ be an hypo-convergent sequence of u.s.c. functions on a compact topological space $\mathcal{C}$. Let $c^\nu \in \text{argmax}_{\mathcal{C}} \bar{\Psi}^\nu$. Then all cluster points of the sequence $\{c^\nu\}$ lie in $\text{argmax}_{\mathcal{C}} \Psi(c)$. If $\text{argmax}_{\mathcal{C}} \Psi(c)$ is a singleton, then convergence holds.

**Remark.**

Basic Fact 2 is wrong if continuous convergence is replaced by hypo-convergence (for the definition of hypo convergence see section 7.B in [5]). As a counterexample, let $\mathcal{C} = [0, 1]$,

$$\Psi^\nu(x) = \begin{cases} 2 & \text{for } x = \pi/\nu \\ 1 & \text{for } x = 1 \\ 0 & \text{elsewhere} \end{cases}$$

and

$$\Psi(x) = \begin{cases} 2 & \text{for } x = 0 \\ 1 & \text{for } x = 1 \\ 0 & \text{elsewhere} \end{cases}$$

Then $\Psi^\nu$ hypo-converges to $\Psi$ but does not converge continuously. If $\mathcal{C}^\nu$ is the set of dyadic rationals in [0,1], which are multiples of $2^{-\nu}$, then $\cup_\nu \mathcal{C}^\nu$ is dense in [0,1], but $\text{limsup}_\nu \text{argmax}_{\mathcal{C}^\nu} \Psi^\nu = 1$, while $\text{argmax}_{\mathcal{C}} \Psi = 0$.

**Lemma.** The following inequalities are valid

$$|\log \int \exp[h_1(u)]\, d\mu(u) - \log \int \exp[h_2(u)]\, d\mu(u)| \leq \sup_u |h_1(u) - h_2(u)| \tag{6}$$

or in discrete form

$$|\log \left( \sum_i \exp a_i \right) - \log \left( \sum_i \exp b_i \right) | \leq \sup_i |a_i - b_i|. \tag{7}$$

**Proof.** This follows from

$$\begin{aligned} \int \exp[h_1(u)]\, d\mu(u) &\leq \int \exp[h_2(u)]\, d\mu(u) \cdot \sup_u \exp[|h_1(u) - h_2(u)|] \\ &= \int \exp[h_2(u)]\, d\mu(u) \cdot \exp[\sup_u |h_1(u) - h_2(u)|] \end{aligned}$$

Specialising to a discrete measure sitting on finitely many points, we get (7). □

**Theorem (Blum-De Hardt's Theorem on uniformity of the SLLN)** Let $(Z_i)$ be sequence of independent, identically distributed random variables on some probability space $(\Omega, \mathcal{A}, P)$ and $P^\nu$ the pertaining empirical measure. Let $\mathcal{H}$ be a family of measurable functions. Suppose that for every $\varepsilon > 0$ we can find a finite set of integrable functions on $\Omega$: $L_1(\omega), \ldots, L_{N_\varepsilon}(\omega)$ and $U_1(\omega), \ldots, U_{N_\varepsilon}(\omega)$ such that

(i) $L_i(\omega) \leq U_i(\omega)$ for $1 \leq i \leq N_\varepsilon$,

(ii) $\int U_i(\omega) - L_i(\omega)\; d\mu(\omega) \leq \varepsilon$,

(iii) For every $H \in \mathcal{H}$, there is an $i$ such that $L_i(\omega) \leq H(Z(\omega)) \leq U_i(\omega)$.

Then

$$\sup_{H \in \mathcal{H}} | \int H(z)\; dP^\nu(z) - \int H(z)\; dP(z)| \to 0 \qquad \text{a.s.}$$

as $\nu \to \infty$.

(See Dudley (1984), 6-1-5.)