Economics Bulletin

Volume 29, Issue 3

Uniform in Bandwidth Consistency of Smooth Varying Coefficient Estimators

Juan carlos Escanciano
Indiana University, Bloomington

David Jacho-chavez
Indiana University

Abstract

We prove the strong consistency, uniformly in the bandwidth, of the smooth varying coefficient conditional least squares estimator. Our results justify data-driven choices of bandwidths, such as Silverman's rule-of thumb, or standard cross-validation, that are usually implemented by most practitioners.

Citation: Juan carlos Escanciano and David Jacho-chavez, (2009) "Uniform in Bandwidth Consistency of Smooth Varying Coefficient Estimators", *Economics Bulletin*, Vol. 29 no.3 pp. 1889-1895.
Submitted: Jul 09 2009. **Published:** August 07, 2009.

1 Introduction

Consider the following variation of the varying coefficient model introduced by Hastie and Tibshirani (1993)

$$Y_i = X_i^{\top}\beta\,(Z_i) + \varepsilon_i,\ i = 1,\dots,n \tag{1.1}$$

where $(Y_1, X_1^{\top}, Z_1^{\top}),\dots,(Y_n, X_n^{\top}, Z_n^{\top})$ are independent and identically distributed (i.i.d.) observations taking values in $\mathbb{R}^{1+p+d}(p \geq 1,\ q \geq 1)$, distributed as $(Y, X^{\top}, Z^{\top})$. Furthermore, we assume $\{E[XX^{\top}|Z = z]\}^{-1}$ exists almost surely (a.s.), and X and Z do not have any component in common. The functions $\beta\,(\cdot)$ are continuously differentiable but unknown, while ε is an unobserved error such that $E[X_i\varepsilon_i|Z_i] = 0$. Notice that the last moment condition, and the non-singularity assumption of $E[XX^{\top}|Z = z]$ identifies $\beta\,(z)$ as

$$\beta\,(z) = \left\{E[XX^{\top}|Z = z]\right\}^{-1} E[XY|Z = z],$$

provided the moment $E[XY|Z = z]$ exists a.s.

The model described in (1.1) is very general and flexible as it comprises the classical linear regression model, the partially linear model, the linear quantile regression model or the classical fully nonparametric model, as special cases. It has been used in various settings in Statistics and Economics; see Li et al. (2002) and references therein. For example, in a housing hedonic price model, Y could represent the price paid for a house by a buyer, X can include home specific characteristics such as number of rooms, area of living space, and number of bathrooms, while Z could represent variables that can indirectly determine house prices, such as average neighborhood income, and number of houses the home owner inspected prior to buying the house.

Based on the previous observation, Li et al. (2002) showed the pointwise consistency and asymptotic normality of the following least squares estimator

$$\widehat{\beta}(z) = \left\{\sum_{i=1}^{n} X_i X_i^{\top} K\left(\frac{z - Z_i}{h^{1/d}}\right)\right\}^{-1} \sum_{i=1}^{n} X_i Y_i K\left(\frac{z - Z_i}{h^{1/d}}\right),$$

where the kernel K is any measurable function that satisfies the following conditions

$$\int_{\mathbb{R}^d} K\,(s)\,ds = 1, \text{ and} \tag{1.2}$$

$$\|K\|_{\infty} := \sup_{z\in\mathbb{R}^d} |K\,(z)\,| < \infty, \tag{1.3}$$

and $h \in \mathbb{R}_{++}$ is a smoothing parameter, the so-called bandwidth.

In this paper, we establish for suitable sequences $0 < a_n < b_n$, with probability 1 (w.p.1), and uniformly in $z \in I$ (henceforth I denotes any generic compact subset of $\mathbb{R}^d$),

$$\sup_{a_n \leq h \leq b_n} \|\,\widehat{\beta} - \beta\,\|_I \to 0, \tag{1.4}$$

where $\|f\|_I := \sup_{z\in I} |f(z)|$. Result (1.4) has an important practical implication. It formally justifies the usage of data-dependent bandwidths in empirical applications of model (1.1), where practitioners usually choose bandwidths by Silverman's rule-of-thumb or standard cross-validation for example.

2 Theoretical Framework

Let V, and $V_1, \ldots, V_n$ be a sequence of i.i.d. $\mathbb{R}^q$ $(q \geq 1)$ valued random vectors from the probability space $(\mathcal{V}, \mathcal{A}, P)$ with common distribution function with lebesgue density function f_V. Let $\mathcal{G}$ denote a class of measurable functions on $\mathbb{R}^q$ with a finite valued measurable envelope function G, i.e.

$$G(v) \geq \sup_{g\in\mathcal{G}} |g(v)|\,,\; v \in \mathbb{R}^q.$$

Furthermore, assume that $\mathcal{G}$ is pointwise measurable (see van der Vaart and Wellner (1996, Definition 2.3.3, pp. 110)) such that for constants C, and ν,

$$N(\epsilon, \mathcal{G}) \leq C\epsilon^{-\nu},\; 0 < \epsilon < 1, \text{ where} \tag{2.1}$$

$$N(\epsilon, \mathcal{G}) := \sup_{Q} N(\epsilon\sqrt{Q(G^2)}, \mathcal{G}, d_Q), \tag{2.2}$$

and the suppremum is taken over all probability measures Q on $(\mathcal{V}, \mathcal{A})$ for which $0 < Q\left(G^2\right) < \infty$. If we were to replace expectations by outer expectations, the pointwise measurability is not longer needed. d_Q is the $L_2(Q)$-metric, and $N(\epsilon, \mathcal{G}, d)$ is the minimal number of balls $\{\gamma : d(\gamma, g) < \epsilon\}$ or d-radius ϵ needed to cover $\mathcal{G}$. Consider the following class of function

$$\mathcal{K} = \{K((z - \cdot)/h^{1/d}) : h > 0,\; z \in \mathbb{R}^d\},$$

and assume that $\mathcal{K}$ is pointwise measurable and satisfies (2.1) - (2.2) with $\mathcal{G}$ replaced by $\mathcal{K}$, $q = d$, $V \equiv Z$, and $v \equiv z$.

Now let W, W_1, …, W_n, be another sequence of i.i.d. $\mathbb{R}^r$ $(r \geq 1)$ valued random vectors, and let Φ denote another class of measurable functions on $\mathbb{R}^r$ with a finite valued measurable envelope function $(G \equiv)$ F. Furthermore assume that Φ is pointwise measurable and satisfies (2.1) - (2.2) with $\mathcal{G}$ replaced by Φ, $q \equiv r$, $V \equiv W$, and $v \equiv w$. Furthermore, let $\mathcal{C} := \{c_\varphi : \varphi \in \Phi\}$ represent the class of continuous function on a compact subset J of $\mathbb{R}^d$ indexed by Φ. This class is always assumed to be relatively compact with respect to the sup-norm topology. Finally, for any: i) $\varphi \in \Phi$, ii) continuous functions c_φ on a compact subset J of $\mathbb{R}^d$, iii) $x \in J$, define the process

$$\eta_{\varphi,\,n,\,h}(z) = \frac{1}{nh}\sum_{i=1}^{n} c_\varphi(z)\,\varphi(W_i)\,K\left(\frac{z - Z_i}{h^{1/d}}\right),$$

where K is a kernel with support contained in $[-1/2,\,1/2]^d$ such that (1.2) - (1.3) hold. Let $\mathbb{I}\{\cdot\}$ represent the indicator function that equals one if its argument is true and zero otherwise, then Einmahl and Mason (2005) proved

Theorem 1 *(Einmahl and Mason (2005, Proposition 2, pp. 1397)) Let $J = I^{\eta} = \{x \in \mathbb{R}^d : |x-y| \leq \eta,$ $y \in I\}$, for I a compact set of $\mathbb{R}^d$ for some $0 < \eta < 1$. Also assume that*

$$f_{WZ} \text{ is continuous and strictly positive on } J.$$

Further assume that the envelope function F of the class Φ satisfies

$$\exists M > 0 : \; F(W)\,\mathbb{I}\{Z \in J\} \leq M, \; a.s. \tag{2.3}$$

or for some $p > 2$

$$\alpha := \sup_{x \in J} E[F^p(W)\,|Z = x] < \infty. \tag{2.4}$$

Then we have for any $c > 0$ and $0 < h_0 < (2\eta)^d$, with probability 1,

$$\limsup_{n\to\infty} \sup_{c(\log n/n)^{\gamma} \leq h \leq h_0} \frac{\sup_{\varphi\in\Phi} \|\eta_{\varphi,\,n,\,h} - E\eta_{\varphi,\,n,\,h}\|_I}{\sqrt{nh\,(|\log h| \vee \log\log n)}} =: Q(c) < \infty,$$

where $\gamma = 1$ in the bounded case (2.3) and $\gamma = 1 - 2/p$ under assumption (2.4).

The next result is due to Dony et al. (2006) and it will be use in our subsequent theory.

Lemma 2.1 *(Dony et al. (2006, Lemma A.2, pp. 119)) Let φ be a measurable function on $\mathbb{R}^d$, which for some $\gamma > 0$ is bounded and uniformly continuous on $D_{\gamma} = \{x \in \mathbb{R}^d : |x-y| \leq \gamma,\; y \in D\}$, where D is a closed subset of $\mathbb{R}^d$. Then for any $L_1(\mathbb{R}^d)$ function H, which is equal to zero for $x \notin I^d$,*

$$\sup_{z\in D} |\varphi * H_h(z) - I(H)\,\varphi(z)| \to 0, \; as\; h \to 0,$$

where $I(H) = \int_{\mathbb{R}^d} H(u)\,du$ and $\varphi * H_h(z) := h^{-1}\int_{\mathbb{R}^d} \varphi(x)\,H(h^{-1/d}(z-x))dx$.

We are now able to establish the desired result.

3 Main Consistency Result

For $W = [Y, X^{\top}]^{\top}$, define the class of functions

$$\Phi := \{\varphi(w) = x^{(l)}x^{(j)} \text{ or } \varphi(w) = yx^{(l)},\; l, j = 1, \ldots, p\},$$

where $x^{(l)}$ denotes the $l-th$ component of the vector x. Obviously, the class Φ, being of finite cardinality, is pointwise measurable, satisfying (2.1) - (2.2). Hence the assumptions of Theorem 1 hold and therefore, for each $1 \leq l,\, j \leq p$, and sequence a_n satisfying

$$a_n \searrow 0, \text{ and } na_n/\log n \to \infty, \tag{3.1}$$

we have

$$\sup_{z\in I}\sup_{a_n\leq h\leq h_0}\left|\eta^{(l,j)}_{\varphi,\,n,\,h}(z)-E\eta^{(l,j)}_{\varphi,\,n,\,h}(z)\right|\to 0, \text{ a.s., and}$$
$$\sup_{z\in I}\sup_{a_n\leq h\leq h_0}\left|\eta^{(l)}_{\varphi,\,n,\,h}(z)-E\eta^{(l)}_{\varphi,\,n,\,h}(z)\right|\to 0, \text{ a.s.,}$$

where

$$\eta^{(l,j)}_{\varphi,\,n,\,h}(z) := \frac{1}{nh}\sum_{i=1}^{n} X_i^{(l)}X_i^{(j)}K\left(\frac{z-Z_i}{h^{1/d}}\right), \tag{3.2}$$
$$\eta^{(l)}_{\varphi,\,n,\,h}(z) := \frac{1}{nh}\sum_{i=1}^{n} X_i^{(l)}Y_iK\left(\frac{z-Z_i}{h^{1/d}}\right); \tag{3.3}$$

with

$$E\eta^{(l,j)}_{\varphi,\,n,\,h}(z) = \frac{1}{h}\int_{\mathbb{R}^d} E[X^{(l)}X^{(j)}|Z=t]K\left(\frac{z-t}{h^{1/d}}\right)f_Z(t)\,dt := m^{(l,j)}f_Z * K(z),$$
$$E\eta^{(l)}_{\varphi,\,n,\,h}(z) = \frac{1}{h}\int_{\mathbb{R}^d} E[X^{(l)}Y|Z=t]K\left(\frac{z-t}{h^{1/d}}\right)f_Z(t)\,dt := m^{(l)}f_Z * K(z),$$

and since $m^{(l,j)}(z) \equiv E[X^{(l)}X^{(j)}|Z=z]$, $m^{(l)}(z) \equiv E[X^{(l)}Y|Z=z]$, and f_Z are assumed to be continuous on $J = I^{\eta}$, Lemma 2.1 implies that as $h \searrow 0$,

$$\sup_{z\in I}\left|E\eta^{(l,j)}_{\varphi,\,n,\,h}(z)-m^{(l,j)}f_Z(z)\right|\to 0, \tag{3.4}$$
$$\sup_{z\in I}\left|E\eta^{(l)}_{\varphi,\,n,\,h}(z)-m^{(l)}f_Z(z)\right|\to 0. \tag{3.5}$$

Therefore, it then follows from (3.2) - (3.5) that uniformly in $z \in I$ and for $a_n < b_n$, with a_n satisfying (3.1), and $b_n \searrow 0$,

$$\sup_{a_n\leq h\leq b_n}\left|\eta^{(l,j)}_{\varphi,\,n,\,h}(z)-m^{(l,j)}f_Z(z)\right|\to 0, \text{ a.s.}$$
$$\sup_{a_n\leq h\leq b_n}\left|\eta^{(l)}_{\varphi,\,n,\,h}(z)-m^{(l)}f_Z(z)\right|\to 0, \text{ a.s.}$$

Therefore, under the assumed (a.s.) invertibility of $E[XX^{\top}|Z=z]$, we have by the Continuous Mapping Theorem the uniform consistency of $\widehat{\beta}(z)$.

We summarize our findings in the following theorem, preceded by a set of assumptions. First, define the envelope $F(W) = \max_{1\leq l,j\leq p}\{\left|X^{(l)}X^{(j)}\right|, \left|X^{(l)}Y\right|\}$.

Assumption (K) The kernel function K has support contained in $[-1/2,\ 1/2]^d$ and is such that (1.2) - (1.3) hold, and the corresponding class $\mathcal{K}$ is pointwise measurable and satisfies (2.1) - (2.2);

Assumption (D) The joint density of $(Y, X^\top, Z^\top)$ is continuous and strictly positive on $J \equiv I^\eta$. Also, the components of $E[XX^\top|Z = z]$ and $E[XY|Z = z]$ are continuous on J. The matrix $E[XX^\top|Z = z]$ is positive definite a.s. on J.

Theorem 2 *Let Assumptions (K) and (D) hold, if F satisfies*

$$\exists M > 0 :\ F(W)\,\mathbb{I}\{Z \in J\} \leq M,\ a.s., \tag{3.6}$$

or for some $p > 2$

$$\alpha := \sup_{z\in J} E[F^p(W)\,|Z = z] < \infty. \tag{3.7}$$

Then for any $c > 0$ and $b_n \searrow 0$,

$$\sup_{c(\log n/n)^\gamma \leq h \leq b_n} \parallel \widehat{\beta} - \beta \parallel_I \to 0\ a.s. \tag{3.8}$$

where $\gamma = 1$ in the bounded case (3.6) and $\gamma = 1 - 2/p$ under assumption (3.7).

An important implication of (3.8) is that it justifies the usage of constant data-driven bandwidth sequences $\widehat{h}_n$ such that, for large enough n,

$$a_n \leq \widehat{h}_n \leq b_n, \text{ w.p.1.} \tag{3.9}$$

Andrews (1995) pointed out that (3.9) holds in many cases for common data-dependent methods of choosing bandwidths, including cross-validation, generalized cross-validation, and plug-in procedures. For example, if $d = 1$ and $\beta(\cdot)$ has at least two continuous derivatives, (3.8) implies that

$$\parallel \widehat{\beta}_{\widehat{h}_n} - \beta \parallel_I \to 0, \text{ w.p.1,}$$

for a plug-in bandwidth of the form $\widehat{h}_n = \widehat{c}n^{-1/5}$, and deterministic sequences $a_n = an^{-1/5}$, $b_n = bn^{-1/5}$; where $\widehat{c} \to c$ w.p.1, $c \in [a, b]$, $0 < a < b < \infty$, and $n^{-1/5}$ is the pointwise optimal rate of convergence, see Li et al. (2002). A similar argument holds for local data-driven bandwidth sequences, $\widehat{h}_n \equiv \widehat{h}_n(z)$, satisfying

$$\Pr\{a_n \leq \widehat{h}_n(z) \leq b_n : z \in I\} \to 1, \text{ as } n \to \infty.$$

References

Andrews, D. W. K. 1995, Nonparametric kernel estimation for semiparametric models, *Econometric Theory*, **11**, pp. 560–596.

Dony, J., Einmahl, U., and Mason, D. M. 2006, Uniform in bandwidth consistency of local polynomial regression function estimators, *Austrian Journal of Statistics*, **35** (2 & 3), pp. 105–120.

Einmahl, U. and Mason, D. M. 2005, Uniform in bandwidth consistency of kernel-type function estimators, *Annals of Statistics*, **33** (3), pp. 1380–1403.

Hastie, T. and Tibshirani, R. 1993, Varying-coefficient models, *Journal of the Royal Statistical Society. Series B (Methodological)*, **55** (4), pp. 757–796.

Li, Q., Huang, C. J., Li, D., and Fu, T.-T. 2002, Semiparametric smooth coefficient models, *Journal of Business & Economic Statistics*, **20** (3), pp. 412–422.

van der Vaart, A. W. and Wellner, J. A. 1996, *Weak convergence and empirical processes With applications to statistics*, Springer Series in Statistics, New York: Springer-Verlag, 1st edition, pp.xvi+508.