Tying the straps tighter for generalized linear models

Ya'acov Ritov

Weining Wang

Wolfgang K. Härdle

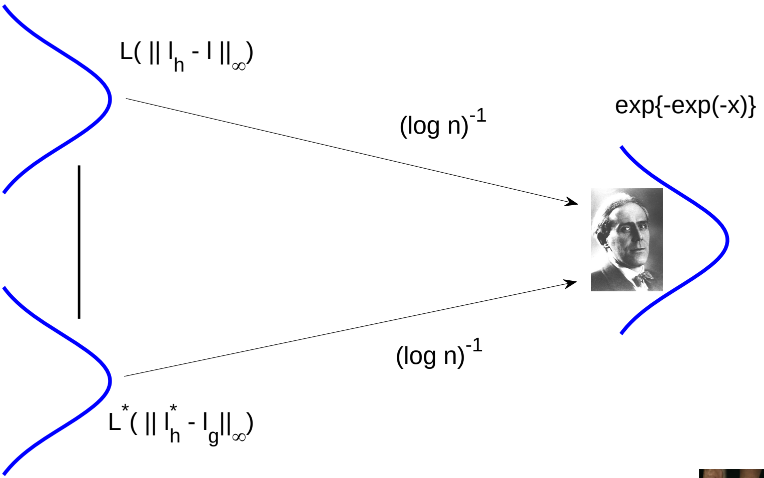
Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. - Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin

Department of Statistics
The Hebrew University of Jerusalem
http://www.case.hu-berlin.de
http://www.stat.huji.ac.il

Motivation

- Bootstrap improves finite sample performance
- Regression Problem: resample from residuals
- For nonlinear functionals (e.g. sup), little improvement
- $\|l_n(x) - l(x)\|_\infty \stackrel{\text{def}}{=} \sup_x |l_n(x) - l(x)|$, where $l(x)$ is a mean or quantile curve, and $l_n(x)$ is a quantile or mean smoother

Challenges



Song et. al. (2012)

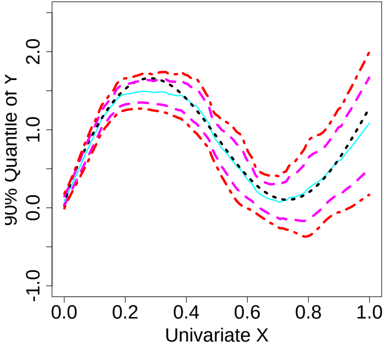


Figure 1: The real 0.9 quantile curve, 0.9 quantile estimate with corresponding 95% uniform confidence band from asymptotic theory and confidence band from bootstrapping.

Theorem (Song et. al. (2010))

An approximate $(1-\alpha)\times 100\%$ *confidence band over* $[0,1]$ *is*

$$\begin{aligned} l_h(t) \quad \pm \quad & (nh)^{-1/2}\{p(1-p)/\hat{f}_X(t)\}^{1/2}\hat{f}^{-1}\{l_h(t)|t\} \\ & \times\{d_n + c(\alpha)(2\delta\log n)^{-1/2}\}\cdot\{\lambda(K)\}^{1/2}, \qquad (1) \end{aligned}$$

where $c(\alpha) = \log 2 - \log|\log(1-\alpha)|$ *and* $\hat{f}_X(t)$, $\hat{f}\{l_h(t)|t\}$ *are consistent estimates for* $f_X(t)$, $f\{l(t)|t\}$.

n	Cov. Prob.
50	0.144 (0.642)
100	0.178 (0.742)
200	0.244 (0.862)

Table 1: Simulated coverage probabilities based on asymptotics (bootstrap methods).

Bootstrap Improvement

- Bootstrap improvements for the quantile smoother have been shown
- Quantile estimator is of Bounded Influence
- M- smoothers or a general QMLE may be expected to have same performance?
- How to handle high dimensional objects?

Opportunities

Extend this to $x \in \mathbb{R}^d$ and improve band precision?

- Hall (1991): bootstrap improvement
- Hahn (1995): consistency of bootstraping cdf
- Horowitz (1998): bootstrap (pointwise) for median
- Additive Models: Horowitz (2001), Horowitz and Lee (2005) and Stone(1985)

How to tie the straps tighter?

Outline

1. Motivation $\checkmark$
2. Bootstrap Confidence Bands
3. Bootstrap Confidence Bands for Additive Models
4. Monte Carlo Study
5. Applications

Regression with a general loss function

- $\{(X_i, Y_i)\}_{i=1}^n$ i.i.d. rv's, $x \in J^* = (a, b)$
- Suppose $Y_i = l(X_i) + \varepsilon_i$, $\varepsilon_i \sim F_{\varepsilon|X_i}(\cdot)$. Both l & F are smooth.
- Suppose

$$l(x) = \arg\min_\theta \mathsf{E}_{(Y|X=x)} \rho(Y - \theta), \tag{2}$$

where $\rho(.)$ is a loss function of Hampel/Huber type or more generally a negative (quasi) log likelihood.

Robust Statistics

For $\rho(.)$ is a trimmed mean (Tukey's biweight),

$$\rho(x) = \begin{cases} x^2, & |x| \leq k, \\ k^2, & |x| > k \end{cases} \tag{3}$$

or a form of Winsorized mean (Huber):

$$\rho(x) = \left\{ \begin{array}{rl} x^2/2, & |x| \leq k, \\ -k^2/2 + k|x|, & |x| > k. \end{array} \right\}. \tag{4}$$

The Bootstrap Couple

- $\{U_i\}_{i=1}^n$: i.i.d. uniform $U[0,1]$ rv's
- Bootstrap sample

$$Y_i^* = l_g(X_i) + \hat{F}^{-1}_{(\varepsilon_i|X_i)}(U_i), \quad i = 1, \ldots, n$$

- Couple with the true conditional distribution:

$$Y_i^{\#} = l(X_i) + F^{-1}_{(\varepsilon_i|X_i)}(U_i), \quad i = 1, \ldots, n.$$

Given $X_1, \ldots, X_n$: $Y_1, \ldots, Y_n$ and $Y_1^{\#}, \ldots, Y_n^{\#}$ are equally distributed.

Bootstrapping Approximation Rate

Theorem

If assumptions (A1)–(A3) hold, then

$$\sup_{x\in J^*} |l_h^*(x) - l_g(x) - l_h(x) + l(x)| = \mathcal{O}_p(\delta_n) = \mathcal{O}_p(h^2\Gamma_n),$$

with Γ_n a slowly varying sequence (a sequence a_n is slowly varying if $n^{-\alpha}a_n \to 0$ for any $\alpha > 0$).

Bootstrap improves the rate of convergence

Sketch of the Proof

- The basic elements: smoothness of $F_{Y|X}(\cdot)$ and bounded influence of $\rho(.)$
-
$$
\begin{aligned}
\max_i |\varepsilon_i^\sharp - \varepsilon_i^*| &= \mathcal{O}_p(h^2\Gamma_n)\\
\max_i |\psi(\varepsilon^*) - \psi(\varepsilon^\sharp)| &= \mathcal{O}_p(h^2\Gamma_n)\\
\max_{||X_i - X_j|| \le ch} |l_g(X_i) - l_g(X_j) - \{l(X_i) - l(X_j)\}| &= \mathcal{O}_p(h^2\Gamma_n)
\end{aligned}
$$
-
$$
|\hat{F}^{-1}(u|X_i) - F^{-1}(u|X_i)| \le h^2\Gamma_n, \forall u \in B.
$$
-
$$
\mathsf{E}_{\hat{F}_{\varepsilon|x=X_i}} \psi(\varepsilon_i^*) = 0 = \mathsf{E}_{F_{\varepsilon|x=X_i}} \psi(\varepsilon^\sharp) \tag{5}
$$

$$
\begin{aligned}
G_n^*(\theta, X_i) &\stackrel{\text{def}}{=} n^{-1} \sum_{j=1}^{n} W_{h,j}(X_i)[\psi\{\varepsilon_j^* - \theta + \hat{l}_g(X_j)\}] \\
&= n^{-1} \sum_{j=1}^{n} W_{h,j}(X_i)\{\psi(Y_j^* - \theta)\} \\
G_n^\sharp(\theta, X_i) &\stackrel{\text{def}}{=} n^{-1} \sum_{j=1}^{n} W_{h,j}(X_i)[\psi\{\varepsilon_j^\sharp - \theta + l(X_j)\}] \\
&= n^{-1} \sum_{j=1}^{n} W_{h,j}(X_i)\{\psi(Y_j^\sharp - \theta)\} \\
T_{n,1}(X_i) &\stackrel{\text{def}}{=} G_n^*\{\hat{l}_g(.), X_i\} - G_n^\sharp\{l(.), X_i\}
\end{aligned}
$$

$$
\begin{aligned}
\mathsf{E}_{F(\varepsilon|X_i)} T_{n,1}(X_i) &= 0 \\
\mathrm{Var}_{(\varepsilon|X_i)} T_{n,1}(X_i) &= n^{-2} \sum_{j=1}^{n} W_{h,j}^2(X_i) \, \mathrm{Var}\{\psi(\varepsilon_j^*) - \psi(\varepsilon_i^\sharp)\} \\
&= \mathcal{O}_p(n^{-1} h^3)
\end{aligned}
$$

$$
\begin{aligned}
\hat{l}_{h,g}^*(X_i) - \hat{l}_g(X_i) &= -\frac{G_n^*\{\hat{l}_g(X_i), X_i\}}{G_n'^*\{\hat{l}_{h,g}^*(X_i), X_i\}} + o_p(h^2), \qquad (6) \\
\hat{l}_h(X_i) - l(X_i) &= -\frac{G_n^\sharp\{l(X_i), X_i\}}{G_n'^\sharp\{\hat{l}_h(X_i), X_i\}} + o_p(h^2). \qquad (7)
\end{aligned}
$$

Why Oversmoothing?

- Take care of bias with tuning parameter: g
- Härdle and Marron (1991), let

$$
\begin{aligned}
b_h(x) &\stackrel{\mathrm{def}}{=} \mathsf{E}\, l_h^{\#}(x) - l(x) \\
\hat{b}_{h,g}(x) &\stackrel{\mathrm{def}}{=} \mathsf{E}^*\, l_h^*(x) - l_g(x)
\end{aligned}
$$

- Investigate $\mathsf{E}\left[\left\{\hat{b}_{h,g}(x) - b_h(x)\right\}^2 | X_1, \ldots, X_n\right]$.
  How fast it converges to 0?

Oversmoothing

Theorem

Under some assumptions, for any $x \in J^*$

$$\mathrm{E}\left[\left\{\hat{b}_{h,g}(x) - b_h(x)\right\}^2 | X_1, \ldots, X_n\right] \sim h^4\{\mathcal{O}_p(g^4) + \mathcal{O}_p(n^{-1}g^{-5})\}$$

To minimize RHS, $g = \mathcal{O}(n^{-1/9})$, $g \gg h$, where $h = \mathcal{O}(n^{-1/5})$

The Multivariate Case

$$m(X_i) = \sum_{j=1}^{d} m_j(x_{i,j}), \tag{8}$$

Estimate the additive model via a basis function approach:

$$m_j(x_{i,j}) \approx \sum_{l=1}^{L_j+1} a_{l,j} g_l(x_{i,j}),$$

$\psi(x_{i,j})$s are B-splines, e.g. linear B-splines.

$$\psi_l(x) = \begin{cases} Hx - l + 1 & (l-1)H^{-1} \le x \le lH^{-1} \\ l + 1 - Hx & lH^{-1} \le x \le (l+1)H^{-1} \\ 0 & \text{otherwise} \end{cases}$$

Bootstrap Couple

Quantile

$$Z_i = \begin{cases} 1 & \text{with prob } \tau \\ -1 & \text{with prob } 1-\tau \end{cases},$$

where $\tau = 1/2$ is for symmetric error distribution. The bootstrap couple ε^* (the bootstrap residuals) and $\varepsilon^\sharp$ (the theoretical couple) are:

$$\varepsilon_i^* = Z_i |\hat{\varepsilon}_i| \tag{9}$$

$$\varepsilon_i^\sharp = Z_i \eta_i \tag{10}$$

Bootstrap

$$F_{i,s}(t) \stackrel{\text{def}}{=} \mathrm{P}(|\varepsilon_i| \le t | s\varepsilon_i > 0), \quad i = 1, \ldots, n, \ s \in \{1, -1\}, \tag{11}$$

$$\eta_i \stackrel{\text{def}}{=} F^{-1}_{i,Z_i}\{F_{i,sgn(\varepsilon_i)}(|\varepsilon_i|)\}, \quad i = 1, \ldots, n. \tag{12}$$

Recall $F_{Y|X=x_i}\{l(X_i)\} = \tau$ and $F_{\varepsilon|X=x_i}(0) = \tau$. $F_{i,\mathrm{sgn}(\varepsilon_i)}(|\varepsilon_i|)$ standard uniform

$$F_{i,+1}(t) = \frac{F_i(t) - 1 + \tau}{\tau}, F_{i,-1}(t) = \frac{1 - \tau - F_i(-t)}{1 - \tau}.$$

$$\mathcal{L}(\varepsilon_i^\sharp) = \mathcal{L}(\varepsilon_i). \tag{13}$$

Go to details

Uniform Consistency

Theorem

Let assumptions A.1- A.9 be fulfilled, then

$$\sup_x \left|(\hat{m}_j - m_j)(x) - \{(\hat{m}_j^* - \hat{m}_j)(x)\}\right| = o_p(h^2\Gamma_n),$$

Go to conditions

How to Bootstrap?

- Simulate $\{X_i, Y_i\}_{i=1}^n$,

$$f(x, y) = g\{y - \sin(\pi x)\}\mathbf{1}(x \in [0, 1]) \tag{14}$$

$$g(u) = 9\varphi(u)/10 + \varphi(u/9)/90 \tag{15}$$

- Compute the smoother (Tukey biweight) $\hat{l}_h(x)$, $\hat{\varepsilon}_i = Y_i - \hat{l}_h(X_i)$
- Compute the conditional edf:

$$\hat{F}_{\varepsilon|X_i=x}(t) = \frac{\sum_{i=1}^n K_h(x - X_i)\mathbf{1}(\hat{\varepsilon}_i \le t)}{\sum_{i=1}^n K_h(x - X_i)}$$

with Gaussian kernel

$$K_h(u) = (\sqrt{2\pi})^{-1} \exp\{-u^2/2h\}/h,$$

$h = 0.06$.

- For each $i = 1, \ldots, n$, generate random variable $\varepsilon^* \sim \hat{F}(t|x)$, $i^* = 1, \ldots, n^*$, where n^* is the bootstrap sample size:

$$Y_{i,i^*} = \hat{l}_g(X_i) + \varepsilon^*_{i,i^*},$$

with $g = 0.2$.

- For each sample X_i, Y_i^*, compute $\hat{l}_h^*(.)$ and the random variable

$$d_{i^*} \stackrel{\text{def}}{=} \sup_{x \in J^*} \{\sqrt{\hat{f}_X(x)}\, \mathsf{E}_{y|x}\{\psi'(\varepsilon_i^*)\} / \sqrt{\mathsf{E}_{y|x}\{\psi^2(\varepsilon_i^*)\}} |\hat{l}_h^*(x) - \hat{l}_g(x)|\}$$

- Calculate the $1-\alpha$ quantile d_α^* of $d_1, \ldots, d_{n^*}$.
- Construct the bootstrap uniform band centered around $m_h(x)$

$$\hat{l}_h(x) \pm [\sqrt{\hat{f}_X(x)}\, \mathsf{E}_{y|x}\{\psi'(\varepsilon_i^*)\} / \sqrt{\mathsf{E}_{y|x}\{\psi^2(\varepsilon_i^*)\}}]^{-1} d_\alpha^*$$

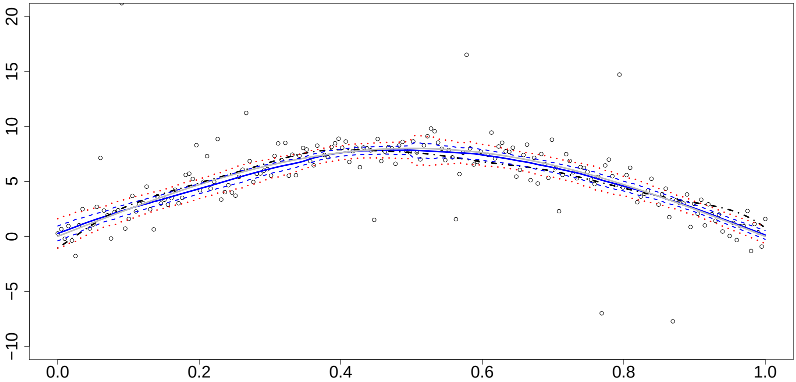


Figure 2: Plot of true curve (grey), BI estimation and bands (blue), local polynomial estimation (black), bootstrap band (red)

n	Cov. Prob.	Area
100	0.88(0.98)	1.23(2.51)
200	0.89(0.98)	0.89(1.95)
400	0.90(0.96)	0.78(1.32)

Table 2: Simulated coverage probabilities areas of nominal asymptotic (bootstrap) 95% confidence bands with 100 repetition.

Bootstrap for Additive Models

- Simulate $\{X_i, Y_i\}_{i=1}^{n}$. The variable $X_i = (x_{1i}, x_{2i}, x_{3i}, x_{4i}) \sim U(-2.5, 2.5)$,

  $m_1(x_1) = \sin(\pi x_1), m_2(x_2) = \Phi(3x_2), m_3(x_3) = x_3^3, m_4(x_4) = x_4^4,$

  and ε_i is as in (15).
- Compute $\hat{m}_1(x_1), \hat{m}_2(x_2), \hat{m}_3(x_3), \hat{m}_4(x_4)$ and $\hat{\varepsilon}_i = Y_i - \sum_{j=1}^{4} \hat{m}_l(x_{i,j})$.
- For each $i = 1, \ldots, n$, generate random variable ε_i^*, $i^* = 1, \ldots, n^*$ as in (9), where n^* is the bootstrap sample size:

$$Y_{i,i^*} = \sum_{j=1}^{4} \hat{m}_l(x_{i,j}) + \varepsilon_{i,i^*}^*.$$

- For each sample $(x_{1i}, x_{2i}, x_{3i}, x_{4i}, Y_i^*)$, compute $m_j^*(.)$ and the random variable

$$d_{i^*} \stackrel{\mathrm{def}}{=} \sup_{x \in J^*} \{\sqrt{\hat{f}_{x_j}(x_j)}\, \mathsf{E}_{\,y|x}\{\psi'(\varepsilon_i^*)\}/\sqrt{\mathsf{E}_{\,y|x}\{\psi^2(\varepsilon_i^*)\}}|\hat{m}_j^*(x) - \hat{m}_j(x)|\}$$

- Calculate the $1-\alpha$ quantile d_α^* of $d_1, \ldots, d_{n^*}$.
- Construct the bootstrap uniform band centered around $\hat{m}_j(x_j)$

$$\hat{m}_j(x_j) \pm [\sqrt{\hat{f}_{x_j}(x_j)}\, \mathsf{E}_{\,y|x_j}\{\psi'(\varepsilon_i^*)\}/\sqrt{\mathsf{E}_{\,y|x_j}\{\psi^2(\varepsilon_i^*)\}}]^{-1} d_\alpha^*$$

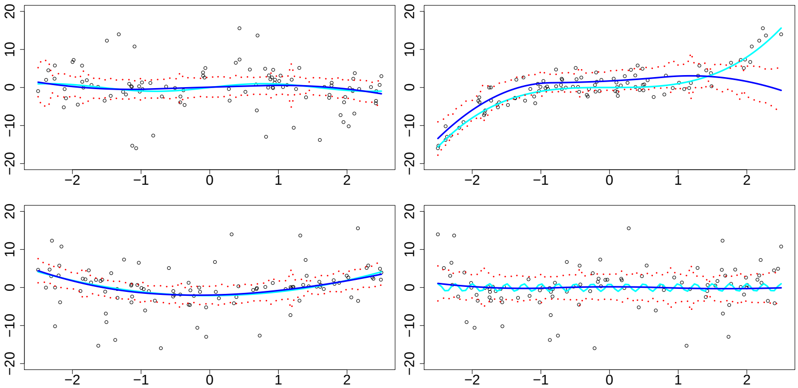


Figure 3: Plot of true curve (dark blue), robust estimation and bands (cyran), bootstrap band (red dotted)

n	Cov. Prob.	Area
100	0.95, 0.98, 0.83, 0.95	6.06, 5.37, 5.44, 5.21
200	0.88, 0.95, 0.93, 0.88	5.50, 4.74, 4.54, 4.65
400	0.84, 0.95, 0.96, 0.84	4.83, 3.63, 3.76, 3.70

Table 3: Simulated coverage probabilities areas of nominal asymptotic (bootstrap) 95% confidence bands with 100 repetition, 100 bootstrap sample.

n	Cov. Prob.	Area
100	0.89, 0.94, 0.85, 0.92	5.88, 5.07, 5.04, 5.30
200	0.90, 0.95, 0.86, 0.88	4.84, 3.84, 3.85, 4.00
400	0.85, 0.90, 0.92, 0.84	4.02, 3.25, 3.11, 3.03

Table 4: Simulated coverage probabilities areas of nominal asymptotic (bootstrap) 90% confidence bands with 100 repetition, 100 bootstrap sample.

Firm expenses analysis

- Joel L. Horowitz and Sokbae Lee (2005)
- Whether a concentrated shareholding is associated with lower expenditure on activities
- The dependent variable Y: general sales and administrative expenses deflated by sales (denoted by MH5)
- The covariates: ownership concentration (denoted by TOPTEN, cumulative shareholding by the largest ten shareholders), firm characteristics: the log of assets, firm age, and leverage (the ratio of debt to debt plus equity)
- $n = 185$

$$
\begin{aligned}
MH5 &= \beta_0 + m_1(TOPTEN) + m_2\{log(Assets)\} \\
&\quad + m_3(Age) + m_4(Leverage) + error
\end{aligned}
$$

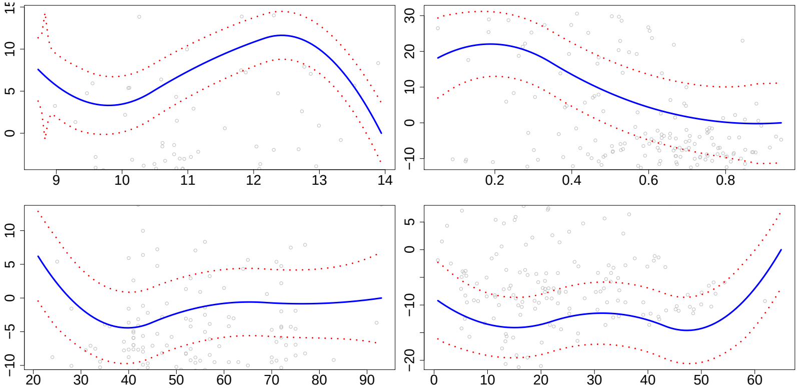


Figure 4: Robust estimation (blue), bootstrap band (red dotted), left up: Log(Asset), right up: Leverage, left below: Age, right below: Leverage.

The impact on stock market

- (Oil, currency, bond, real estate) affect the stock market.
- (http://www.lnstatistical.com /Main.jsp;jsessionid=009E36E74DFA15C80B74EE0BDAEB5746)
- The X variables are: the crude oil price, EUR- USD exchange rate, the 10 year treasury constant maturity inflation index %, and the y variable is S&P 500 index returns.
- $20080903 - 20111128$, $n = 170$

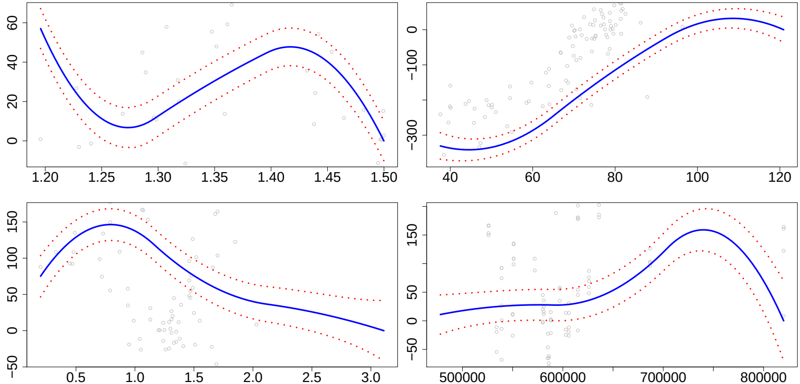


Figure 5: Robust estimation (blue), bootstrap band (red dotted), y S&P index, left up: exchange rates EUR-USD, right up: crude oil price, left below: inflation index, right below: real estate price.

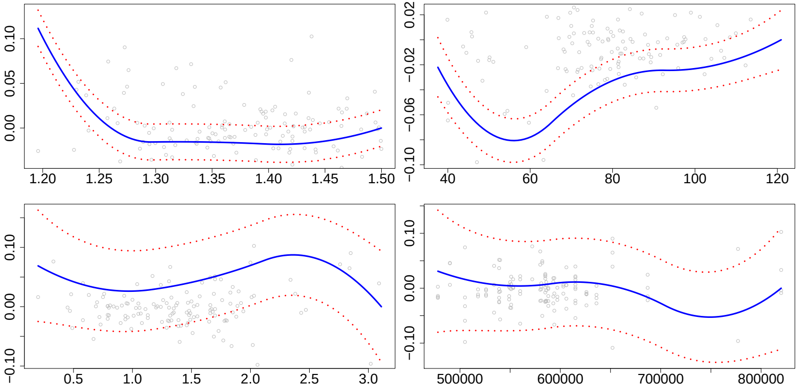


Figure 6: Robust estimation (blue), bootstrap band (red dotted), y S&P index return, left up: exchange rates EUR-USD, right up: crude oil price, left below: inflation index, right below: real estate price.

- Exchange rate (EUR-USD): (< 1.27) negatively correlated with the stock indices, (> 1.43) a positive correlation follows
- Oil prices: negative impact
- Inflation index: the inflation rate high, interest rates typically high; A negative correlation >0.7
- The stock indices raise when the real estate prices gets higher
- ...

Tying the straps tighter for generalized linear models

Ya'acov Ritov
Weining Wang
Wolfgang K. Härdle

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. - Center for Applied Statistics and Economics
Humboldt-Universität zu Berlin

Department of Statistics
The Hebrew University of Jerusalem
http://www.case.hu-berlin.de
http://www.stat.huji.ac.il

References

A. Belloni and V. Chernozhukov
Intersection bounds: estimation and inference
CEMMAP Working Paper, 10/09, 2009.

M. Buchinsky
Quantile regression, Box-Cox transformation model, and the U.S. wage structure, 1963-1987
Journal of Econometrics, 65:109-154, 1995.

V. Chernozhukov, S. Lee and A. M. Rosen
L_1-Penalized quantile regression in high-dimensional sparse models
CEMMAP Working Paper, 19/09, 2009.

References

J. Franke and P. Mwita
Nonparametric Estimates for conditional quantiles of time series
Report in Wirtschaftsmathematik 87, University of Kaiserslautern, 2003.

P. Hall
On convergence rates of suprema
Probab. Th. Rel. Fields, 89:447-455, 1991.

Jinyong Hahn
Bootstrapping Quantile Regression Estimators
Econometric Theory, 11(1):105-121, 1995.

References

W. Härdle and J. S. Marron
Bootstrap simultaneous error bars for nonparametric regression
Ann. Statist., 19:778-796 1991.

W. Härdle and S. Song
Confidence bands in quantile regression
Econometric Theory, forthcoming, 2009.

J. Horowitz
Bootstrap methods for median regression models
Econometrica, 66:1327-1351, 1998.

References

K. Jeong, W. Härdle and S. Song
A Consistent Nonparametric Test for Causality in Quantile
Revise and resubmit to Econometric Theory, 2009.

C. Juhn, K. M. Murphy and B. Pierce
Wage Inequality and the Rise in Returns to Skill
Journal of Political Economy, 101(3): 410-442, 1993.

R. Koenker and K. F. Hallock
Quantile regression
Journal of Econometric Perspectives, 15(4):143-156, 2001.

References

J. Komlós, P. Major and G. Tusnády
An Approximation of Partial Sums of Independent RV'-s, and the Sample DF. I
Z. Wahrscheinlichkeitstheorie verw., 32:111-131, 1975.

K. Yu, Z. Lu and J. Stander
Quantile regression: applications and current research areas
J. Roy. Statistical Society: Series D (The Statistician), 52:331-350, 2003.

Wolfgang Härdle and Joel Horowitz and Jens-Peter Kreiss
Bootstrap Method for Time Series
International Statistical Review / Revue Internationale de Statistique, 71(2):435-459, 2003.

References

Joel L. Horowitz and Sokbae Lee
Nonparametric Estimation of an Additive Quantile Regression Model
Journal of the American Statistical Association, 100:1238-1249, 2005.

Joel L. Horowitz
Nonparametric Estimation of a Generalized Additive Model With an Unknown Link Function
Econometrica, 69:499-513, 2001.

Joel Horowitz and Jussi Klemelä and Enno Mammen
Optimal estimation in additive regression models
Bernoulli, 12(2):271-298, 2006.

References

Charles J. Stone
Additive Regression and Other Nonparametric Models
The Annals of Statistics, 13:689-705, 1985.

Appendix - Assumptions

A.1 $\psi(.) = \rho'(.)$ being a.s. differentiable and Lipschitz continuous: $\forall \mu_1, \mu_2 \in B$ $|\psi(\mu_1) - \psi(\mu_2)| < C|\mu_1 - \mu_2|$, and we assume that $\exists M > 0$ s.t.$\psi(\mu) \leq M$.
A.2 The support of X is $[0,1]^d$. The conditional density $f_{(y|X=x)(.)}$ is bounded from below $\infty > C_1 > \inf_t f_{(y|X=x)}(t) = c_1 > 0$.
A.3 The kernel function $K(.)$ is a product kernel composed from one dimension kernel with bandwidth $h = h_n$:

$$K_h(s) = \Pi_{j=1}^d K(s_j/h)/h, s = (s_1, \ldots, s_d)^\top \in \mathbb{R}^d. \qquad (16)$$

Appendix - Assumptions

A.4 The bandwidth satisfies $h \sim n^{-1/(4+d)}$. Let g be another bandwidth sequence $g >> h$. (Work our later the speed). Let Γ_n be a slowly increasing sequence in the sense that $n^{-\alpha} a_n \to 0$ for any $\alpha > 0$.

A.5 Assume $\sup_{x \in B} |\hat{l}''_g(x) - l''(x)| = \mathcal{O}_p(1)$, and $\sup_{x \in B} |\hat{l}'(x) - l'(x)| = \mathcal{O}_p(h^2 \Gamma_n)$.

A.6 There is an $\alpha > 0$ such that

$$\begin{aligned} \mathrm{E}_X \{\sum_{i=1}^{d} m_j(X_j)\}^2 &\geq \alpha \max_j \mathrm{E}_{X_j}\{m_j^2(X_j)\} \\ \mathrm{E}_{X_j}\{m_j(X_j)\} &= 0, m_j(\cdots) \in L_2(X_j). \end{aligned}$$

A.7 $\mathsf{E}_{X_j}\{\psi_l^2(x_{i,j})\} = 1$ for any $i \in 1, \ldots, n$ and $j \in 1, \ldots, d$.
$||\Phi_l(X_j)||_\infty \leq C_3/L, a.s.$, where $\Phi_l(X_j)), \ldots, \phi_l^2(x_{n,j})\}^\top$.

A.8 The inverse link function b' satisfies the following: $b' \in C(\mathbb{R})$, $b''(\theta) > 0, \theta \in \mathbb{R}$ while for a compact interval Θ whose interior contains $m([0,1]^d)$, $C_b > \max_{\theta \in \Theta} b''(\theta) > \min_{\theta \in \Theta} b''(\theta) > c_b$ for $C_b > c_b > 0$

A.9 The number of regressors $p = dL + 1$ (more precisely $L = L_j$) with $L \sim n^{1/5}$.

Go back

$$
\begin{aligned}
\mathrm{P}(\varepsilon_i^\sharp < t) &= \tau\,\mathrm{P}[F_{i,+1}^{-1}\{F_{i,\sim\eth\ltimes(\varepsilon_i)}(|\varepsilon_i|)\} < t] + 1 - \tau \\
&= \tau\,\mathrm{P}\{F_{i,\sim\eth\ltimes(\varepsilon_i)}(|\varepsilon_i|) < F_{i,+1}(t)\} + 1 - \tau \\
&= \tau\,\mathrm{P}\{\varepsilon_i < 0, F_{i,-1}(-\varepsilon_i) < F_{i,+1}(t)\} \\
&\qquad + \tau\,\mathrm{P}\{\varepsilon_i > 0, F_{i,+1}(\varepsilon_i) < F_{i,+1}(t)\} + 1 - \tau \\
&= \tau\,\mathrm{P}\{\varepsilon_i < 0, \frac{1 - \tau - F_i(\varepsilon_i)}{1 - \tau} < \frac{F_i(t) - 1 + \tau}{\tau}\} \\
&\qquad + \tau\,\mathrm{P}(0 < \varepsilon_i < t) + 1 - \tau \\
&= \tau\,\mathrm{P}[1 - \tau > F_i(\varepsilon_i) > \frac{1-\tau}{\tau}\{1 - F_i(t)\}] \\
&\qquad + \tau\,\mathrm{P}(0 < \varepsilon_i < t) + 1 - \tau \\
&= \tau[1 - \frac{1-\tau}{\tau}\{1 - F_i(t)\} - \tau] + \tau\{F_i(t) - 1 + \tau\} + 1 - \tau \\
&= F_i(t).
\end{aligned}
$$