Current status linear regression

Piet Groeneboom and Kim Hendrickx
Delft University and University of Hasselt

September 9, 2016

Figure: Kim Hendrickx

Current status regression

Consider the regression model

$$Y_i = \alpha + \beta' X_i + \epsilon_i,$$

where X_i is a k-dimensional covariate and ϵ_i an observation error with expectation zero.
We can not observe Y_i!
Instead, we observe $(X_1, T_1, \Delta_1), \ldots, (X_n, T_n, \Delta_n)$, where T_i is an observation time and

$$\Delta_i = 1_{\{Y_i \leq T_i\}}.$$

The ϵ_i are i.i.d. variables, independent of the T_i and X_i, with expectation zero; T_i and X_i are also taken to be independent.
We want to estimate α and β. How to do this?

Maximum likelihood

The log likelihood of the observations is

$$\sum_{i=1}^{n} \{\Delta_i \log F_{Y_i}(T_i) + (1-\Delta_i) \log\{1 - F_{Y_i}(T_i)\}\,,$$

where

$$F_{Y_i}(t) = \mathbb{P}\left\{\alpha + \beta' X_i + \epsilon_i \leq t\right\}.$$

If F is the distribution function of the ϵ_i, we can write the log likelihood in the form

$$\sum_{i=1}^{n} \left\{\Delta_i \log F(T_i - \alpha - \beta' X_i) + (1-\Delta_i) \log\{1 - F(T_i - \alpha - \beta' X_i)\right\}.$$

We also have the extra condition that the expectation of ϵ_i is zero, which amounts to:

$$\int x\, dF(x) = 0.$$

Reduction to simpler model

We drop the condition that the expectation of ϵ_i is zero, and reduce the model to:

$$Y_i = \beta' X_i + \epsilon_i,$$

where the ϵ_i are i.i.d. with unknown expectation α. Then the log likelihood becomes:

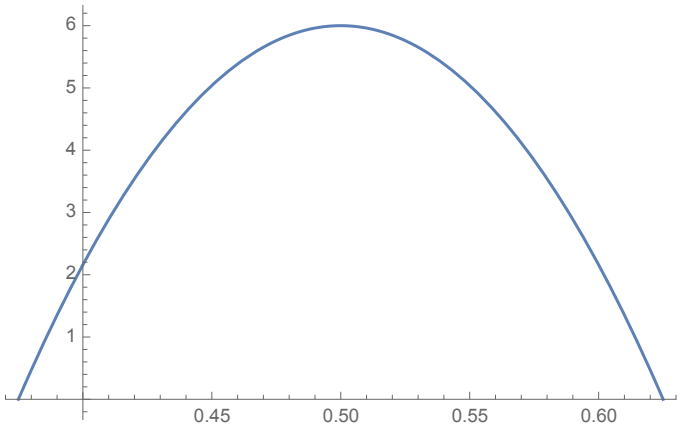
$$\sum_{i=1}^{n} \left\{ \Delta_i \log F(T_i - \beta' X_i) + (1 - \Delta_i) \log\{1 - F(T_i - \beta' X_i)\} \right.$$

where F is the distribution function of ϵ_i.

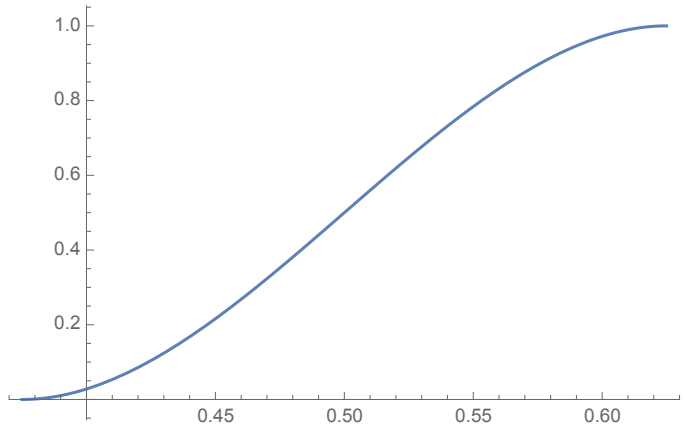
We now first estimate F and β by some estimates $\hat{F}$ and $\hat{\beta}$, and then estimate α by $\int x \, d\hat{F}(x)$ (or a variation on this).

Example

Example: X_i and T_i are uniform on $[0, 2]$, that $\beta = 1/2$, $\alpha = \mathbb{E}\epsilon_i = 1/2$, and suppose ϵ_i has as density a rescaled version of the density $6x(1-x)$ on $[0, 1]$. We rescale it to a density on $[3/8, 5/8]$:



(a) density of ϵ_i.



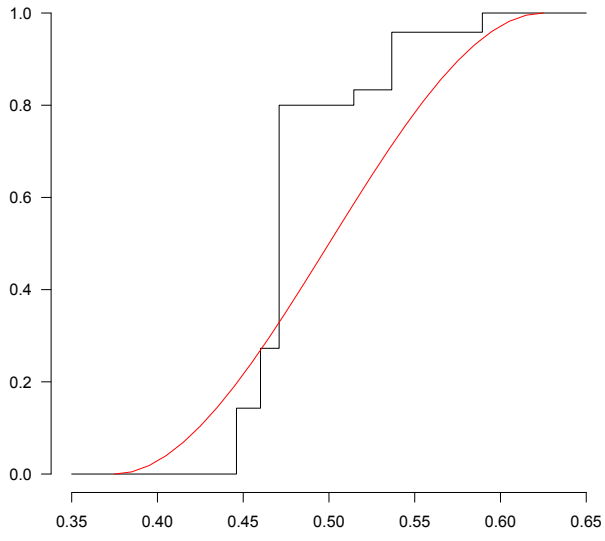
(b) df of ϵ_i.

We can try to straightforwardly maximize w.r.t. F and β:

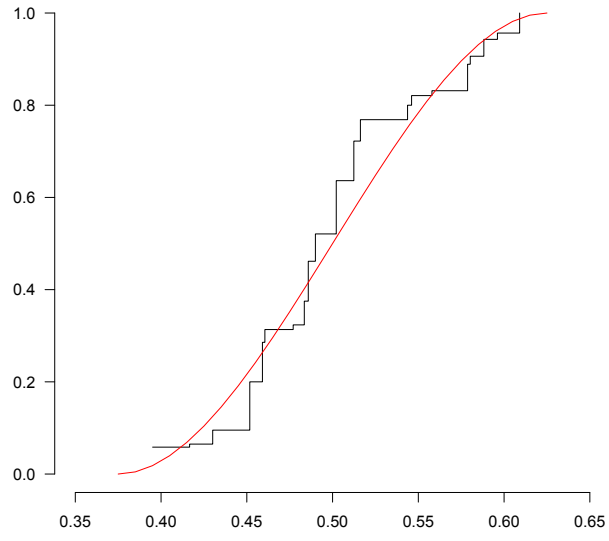
$$\sum_{i=1}^{n} \{\Delta_i \log F(T_i - \beta X_i) + (1 - \Delta_i) \log\{1 - F(T_i - \beta X_i)\}\}.$$

Example (continued)

Example for $n = 1000$: $\hat{\beta} = 0.514828$.
For $n = 10,000$: $\hat{\beta} = 0.499999$.



(a) MLE of F, $n = 1000$.



(b) MLE for F, $n = 10,000$.

Profile likelihood

Profile likelihood: Pick a β, maximize for this β the log likelihood

$$\sum_{i=1}^{n} \left\{ \Delta_i \log F(T_i - \beta' X_i) + (1 - \Delta_i) \log\{1 - F(T_i - \beta' X_i)\} \right\}.$$

over F, this gives $\hat{F}_\beta$ via one step (convex minorant) algorithm.
Now try to find an argmax for β over all $\hat{F}_\beta$ so obtained.
We use Brent's algorithm for maximization, with $\hat{F}_\beta$ as parameter.

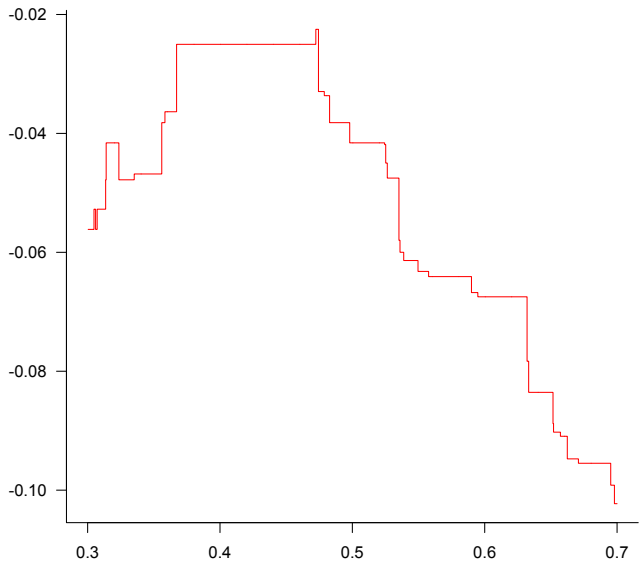


Figure: log likelihood for $\hat{F}_\beta$, as a function of β, $n = 100$.

Properties maximum likelihood estimator

What are the properties of the MLE of β so obtained?
Unknown!
Unknown whether the MLE of β is $\sqrt{n}$-consistent

Li and Zhang (1998) conjecture that the MLE will give a $\sqrt{n}$-consistent but inefficient estimate.

Murphy, van der Vaart, and Wellner (1999) prove that in a model in which one only has observations from a part of F_0 where F_0 stays away from 0 and 1 that the MLE gives a $n^{1/3}$-consistent estimate of β_0.

Interlude: binary choice model

In econometrics, the binary choice model has been studied: special case of current status regression, where the T_i is degenerate at 0 and log likelihood of the form:

$$\sum_{i=1}^{n} \left\{ \Delta_i \log F(\beta' X_i) + (1 - \Delta_i) \log\{1 - F(\beta' X_i)\} \right\}.$$

The regression parameter β is only identifiable under an extra condition, for example $\|\beta\| = 1$.
Example: Whether a person i says yes ($\Delta_i = 1$) or no ($\Delta_i = 0$) on a question in a questionnaire is predicted by his (socio-economic) covariate X_i.
Accompanying theory is also still shrouded in mystery.
Literature: Cosslett (1983, 2007), Klein and Spady (1993), Dominitz and Sherman (2005), lecture notes Bruce Hansen (2009).

Difficulties

Why does estimating β seem so difficult?

β is a parameter which appears as argument of a function F which we cannot estimate nonparametrically at rate $\sqrt{n}$.

Nevertheless, we expect that it is possible to estimate β at rate $\sqrt{n}$.

So we have to "bypass" the non-$\sqrt{n}$ estimable parameter F.

Compare with proportional hazards model under current status: in this case the log likelihood is of the form

$$\sum_{i=1}^{n} \left\{ \Delta_i \log \left(1 - \exp \left\{ -\Lambda(T_i) e^{\beta X_i} \right\} \right) - (1 - \Delta_i) \Lambda(T_i) e^{\beta X_i} \right\},$$

where Λ is the baseline cumulative hazard function. Now β is not hiding inside a function F which is not-$\sqrt{n}$ estimable!

Ordinary maximum likelihood is efficient (Huang (1996))!

Smoothing

Consider restricting the estimates of F to a smooth plug-in estimator (Nadaraya Watson statistic):

$$F_{nh,\beta}(t-\beta x) = \frac{\int \delta K_h(t-\beta x-u+\beta y)\, d\mathbb{P}_n(u,y,\delta)}{\int K_h(t-\beta x-u+\beta y)\, d\mathbb{G}_n(u,y)}, \quad (1)$$

where, for a smooth symmetric kernel K:

$$K_h(u) = h^{-1}K(u/h).$$

Another way of writing $F_{n,\beta}$ is to use ordinary sums:

$$F_{nh,\beta}(t-\beta x) = \frac{\sum_{j=1}^n \Delta_j K_h(t-\beta x-T_j+\beta X_j)}{\sum_{j=1}^n K_h(t-\beta x-T_j+\beta X_j)}.$$

This type of estimate has also been considered in the econometrics literature. Note: $F_{n,\beta}$ is not necessarily monotone, so need not be a distribution function!

Example plugin estimator

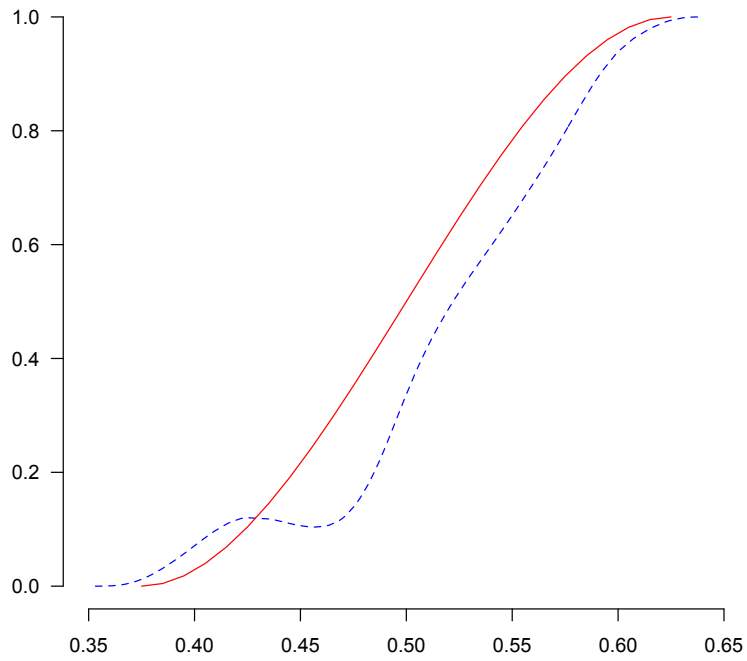
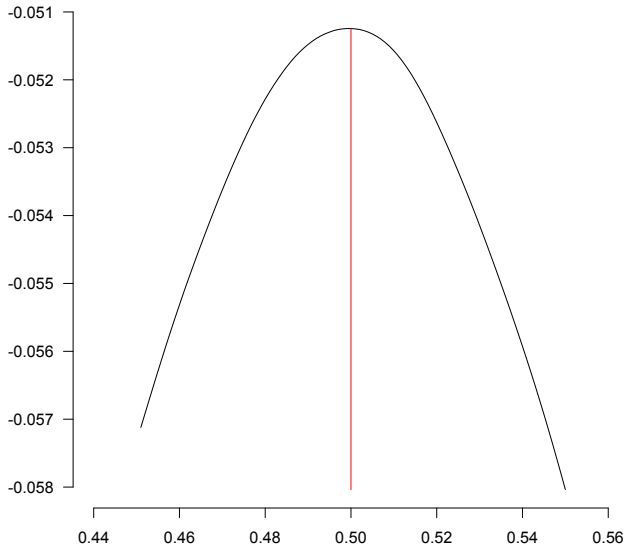
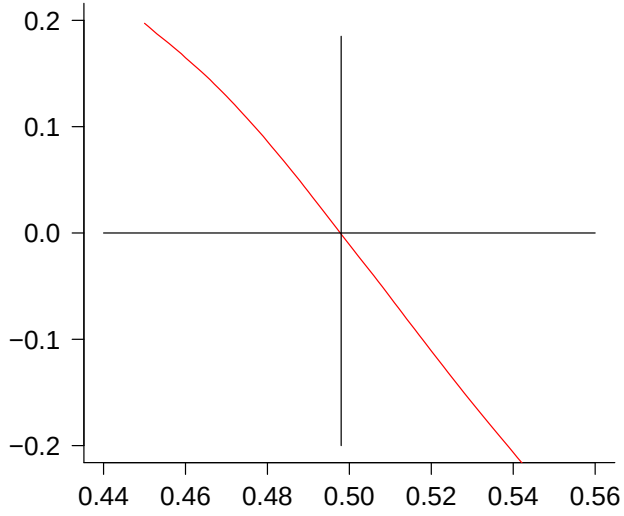


Figure: Plug-in estimator, $n = 1000$ and $\hat{\beta} = 0.49532$.

Maxima plug-in estimator



(a) log likelihood w.r.t. β



(b) scores of plugin w.r.t. β

$$\frac{\partial}{\partial \beta} \textit{log likelihood}(F_{nh,\beta}) = \textit{scores of plugin as function of } \beta$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{\{\Delta_i - F_{nh,\beta}(T_i - \beta X_i)\} \frac{\partial}{\partial \beta} F_{nh,\beta}(T_i - \beta X_i)}{F_{nh,\beta}(T_i - \beta X_i)\{1 - F_{nh,\beta}(T_i - \beta X_i)\}}$$

Questions

"Shrouded in mystery":
How should we choose the bandwidth h?
Can we choose $h \asymp n^{-1/5}$?
Can we use ordinary kernels?
Cosslett (2007): we should have $n^{-1/5} < h < n^{-1/8}$ (excluding $h \asymp n^{-1/5}$ and $h \asymp n^{-1/8}$).
Klein and Spady (1993): $n^{-1/6} < h < n^{-1/8}$ and one should take higher order (4th order) kernels.
Lecture notes Bruce Hansen (2009): "It is unclear to me if these are merely technical sufficient conditions, or if there is a substantive difference with the semiparametric regression case."

Note: The Klein-Spady estimates of the finite-dimensional regression parameter have been proved to be $\sqrt{n}$ consistent and to achieve the information lower bound.

Efficiency plug-in estimators

Theorem (Piet Groeneboom and Kim Hendrickx (2016))

Let $F_{nh,\hat{\beta}_n}$ *maximize the truncated log likelihood*

$$\sum_{F(T_i-\beta' X_i)\in[\epsilon,1-\epsilon]} \Big[\Delta_i \log F(T_i-\beta' X_i) + (1-\Delta_i)\log\{1-F(T_i-\beta' X_i)\}\Big]$$

over all plug-in estimators (ordinary kernels). As $n\to\infty$, *and* $h \asymp n^{-1/5}$, *then*

$$\sqrt{n}(\hat{\beta}_n-\beta_0) \xrightarrow{d} N(0, I_\epsilon(\beta_0)^{-1})$$

where $I_\epsilon(\beta_0)$ *is the matrix with elements*

$$\int_{F_0(u)\in[\epsilon,1-\epsilon]} \frac{covar(X_i,X_j|T-\beta_0' X=u)}{F_0(u)\{1-F_0(u)\}} f_0(u)^2 f_{T-\beta_0' X}(u)\,du.$$

Penalized estimates

Murphy, van der Vaart, and Wellner (1999) consider the penalized maximum likelihood estimator, obtained by maximizing

$$\sum_{i=1}^{n} \{\Delta_i \log F(T_i - \beta X_i) + (1 - \Delta_i) \log\{1 - F(T_i - \beta X_i)\} - \lambda_n \int F''(u)^2 \, du,$$

where

$$1/\lambda_n = O_p\left(n^{2/5}\right), \qquad \lambda_n^2 = o_p\left(n^{-1/2}\right),$$

and observations are only on region where $\epsilon < F_0(u) < 1 - \epsilon$ for some $\epsilon > 0$.

Translated into bandwidth choice ($h_n \asymp \sqrt{\lambda_n}$), the penalty condition correspond to:

$$n^{-1/5} \leq h < n^{-1/8}.$$

Is smoothing really necessary?

Consider the following estimator: $\hat{\beta}_n$ is a value of β such that

$$\sum_{i=1}^{n} X_i \{\Delta_i - \hat{F}_{n,\beta}(T_i - \beta X_i)\} = 0,$$

where $\hat{F}_{n,\beta}$ is the MLE of F_0 based on the values $T_i - \beta X_i$.

Theorem (Piet Groeneboom and Kim Hendrickx (2016))

$$\sqrt{n}\{\hat{\beta}_n - \beta_0\} \xrightarrow{\mathcal{D}} N(0, \sigma^2),$$

where

$$\sigma^2 = \frac{\int \mathrm{var}(X|T - \beta_0 X = u) F_0(u)\{1 - F_0(u)\} f_{T-\beta_0 X}(u)\, du}{\left\{\int \mathrm{var}(X|T - \beta_0 X = u) f_0(u)\, f_{T-\beta_0 X}(u)\, du\right\}^2}.$$

Consequence: One can construct $\sqrt{n}$-consistent estimates of β_0 on the basis of the non-smoothed MLE's $\hat{F}_{n,\beta}$'s.

Simple score function for MLE

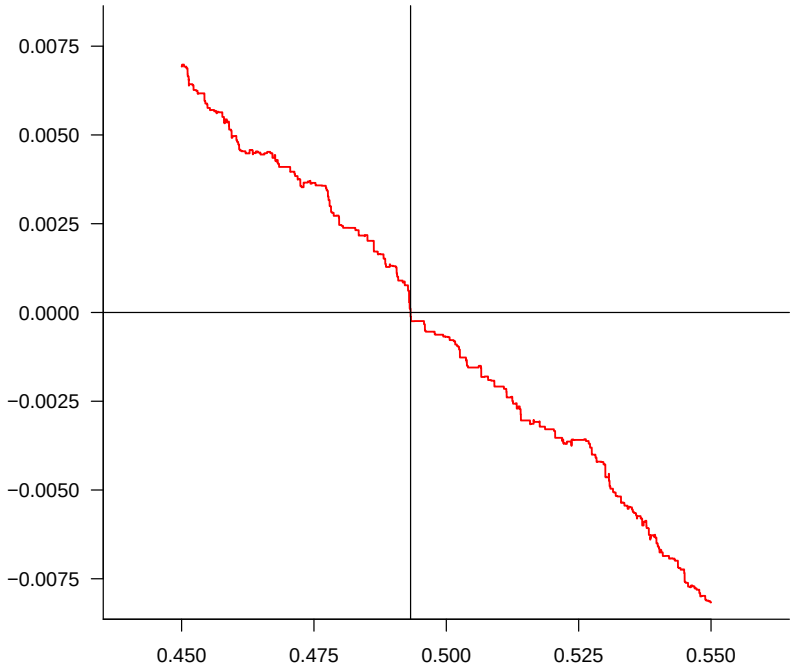


Figure: Simple score function for MLE, n=1000.

Maximum rank estimator of Aragón and Quiroz (1995)

The regression parameter β is estimated by the maximizer $\hat{\beta}_n$ of

$$\Gamma_n(\beta) = \sum_{i=1}^{n} \Delta_i R_i(\beta),$$

where $R_i(\beta)$ is the rank of $T_i - \beta X_i$ in $T_1 - \beta X_1, \ldots, T_n - \beta X_n$.
Similar to Han's maximum rank correlation estimator (Han (1987)) in the econometric literature.
Abrevaya (1999) proves that $\hat{\beta}_n$ is $\sqrt{n}$-consistent and asymptotically normal, using methods of Sherman (1993) in treating Han's maximum rank correlation estimator.

Proof for maximum rank estimator

Use Theorem 1 in Sherman (1993):

$$\|\hat{\beta}_n - \beta_0\| = O_p(n^{-1/2}),$$

if $\hat{\beta}_n$ is the maximizer of $\Gamma_n(\beta)$, with population equivalent $\Gamma(\beta)$ and

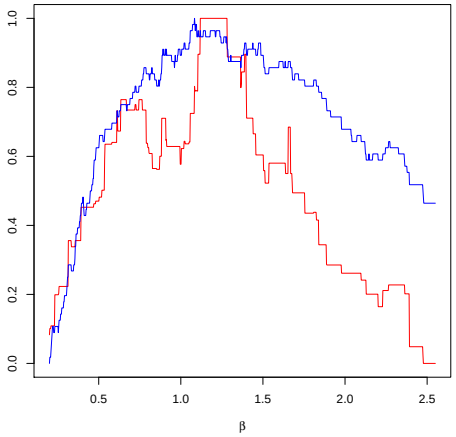
(a) there exists a neighborhood N of β_0 and a constant $k > 0$ such that

$$\Gamma(\beta) - \Gamma(\beta_0) \leq -k\|\beta - \beta_0\|^2,$$

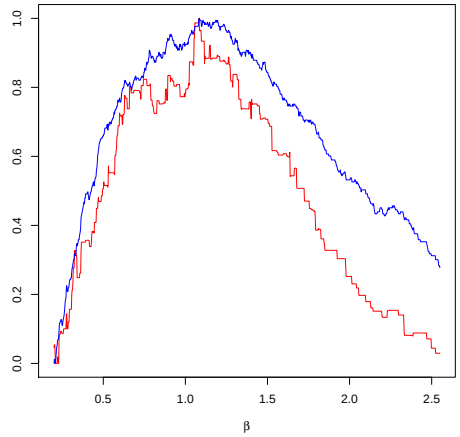
for $\beta \in N$, and

(b) uniformly over $o_p(1)$ neighborhoods of β_0,

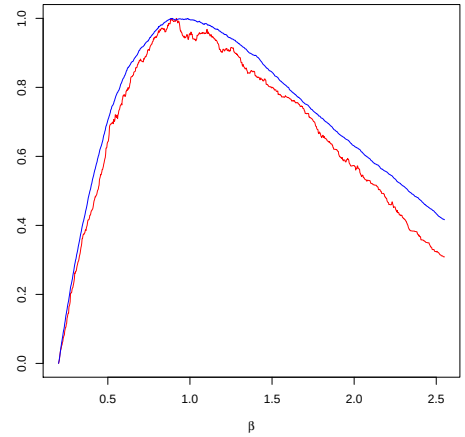
$$\begin{aligned}
&\Gamma_n(\beta) - \Gamma_n(\beta_0) \\
&= \Gamma(\beta) - \Gamma(\beta_0) + O_p\left(\|\beta - \beta_0\|/\sqrt{n}\right) + o_p\left(\|\beta - \beta_0\|^2\right) \\
&\quad + O_p\left(n^{-1}\right).
\end{aligned}$$



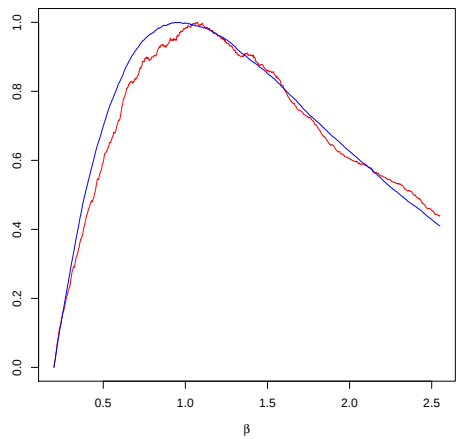
(a) $n = 50$



(b) $n = 100$



(c) $n = 500$

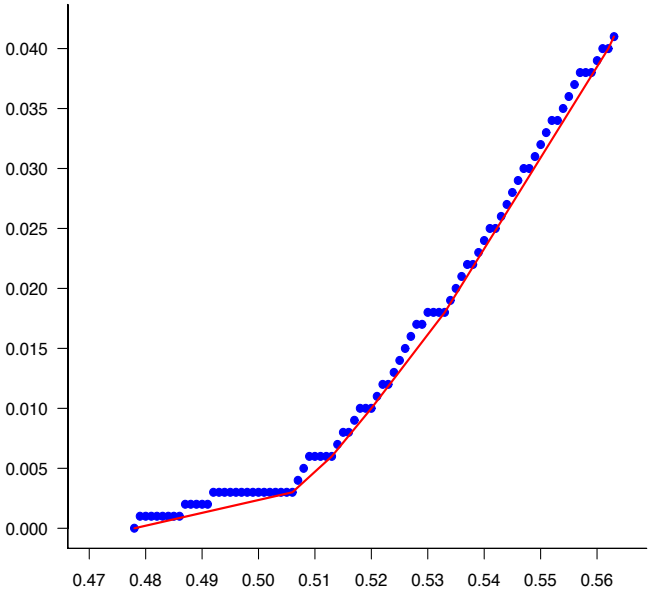


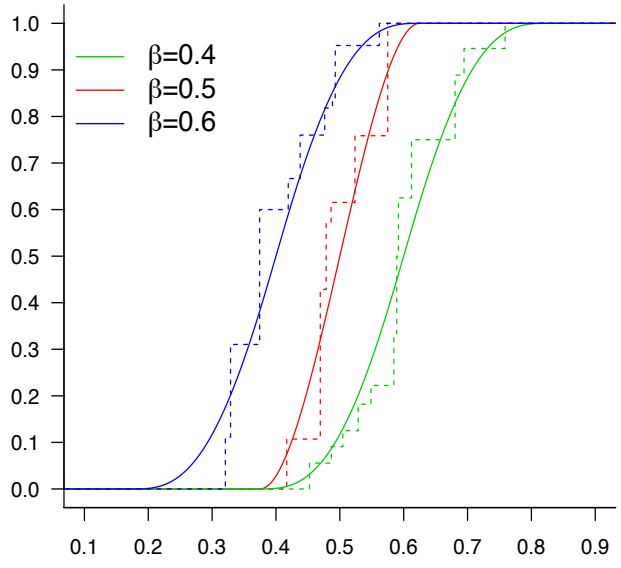
(d) $n = 1000$

Figure: The log likelihood (red) and the criterion function Γ_n (blue) for sample sizes $n = 50, 100, 500$ and 1000. $\beta_0 = 1$ is the true regression parameter. T_i, X_i and ϵ_i are independent and standard normal.

Cumulative sum diagram for different β

MLE $\hat{F}_{n,\beta}$ has jumps at a subset of $\{T_i - \beta' X_i$, i=1,. . . ,n$\}$.





$$\hat{F}_{n,\beta}(u) \xrightarrow{P} P\left\{\Delta_i = 1 | T_i - \beta' X_i = u\right\}$$

$$= \int F_0(u + (\beta - \beta_0)' x) f_{X|T-\beta' X}(x | T - \beta' X = u)\, dx$$

Efficiency

The efficient variance for estimating β is:

$$\int \frac{\mathrm{var}(X|T-\beta_0 X=u)f_0(u)^2}{F_0(u)\{1-F_0(u)\}}\, f_{T-\beta_0 X}(u)\, du.$$

We can change $\hat{\beta}_n$ to a value of β such that

$$\sum_{i=1}^{n} X_i \phi_n(T_i-\beta X_i)\{\Delta_i - \hat{F}_{n,\beta}(T_i-\beta X_i)\} = 0,$$

where $X_i\phi_n(T_i-\beta X_i)$ estimates the efficient score, for example

$$X_i\phi_n(T_i-\beta X_i) = \frac{X_i \hat{f}_{n,h}(T_i-\beta X_i)}{\hat{F}_{n,\beta}(T_i-\beta X_i)\{1-\hat{F}_{n,\beta}(T_i-\beta X_i)\}}.$$

where $\hat{f}_{n,h}(t) = \int K_h(t-u)\, d\hat{F}_{n,\beta}(u)$.

Theorem (Piet Groeneboom and Kim Hendrickx (2016))

Let $\hat{\beta}_n$ *be a value of* β *such that*

$$\int_{\hat{F}_{n,\beta} \in [\epsilon, 1-\epsilon]} x \phi_n(t - \beta' x)) \{\delta - \hat{F}_{n,\beta}(t - \beta' x)\} \, d\mathbb{P}_n = 0,$$

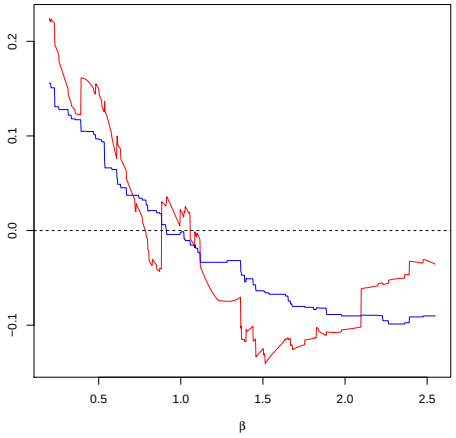
where:

$$\phi_n(t - \beta x) = \frac{\hat{f}_{n,h}(t - \beta' x)}{\hat{F}_{n,\beta}(t - \beta' x)\{1 - \hat{F}_{n,\beta}(t - \beta' x)\}},$$

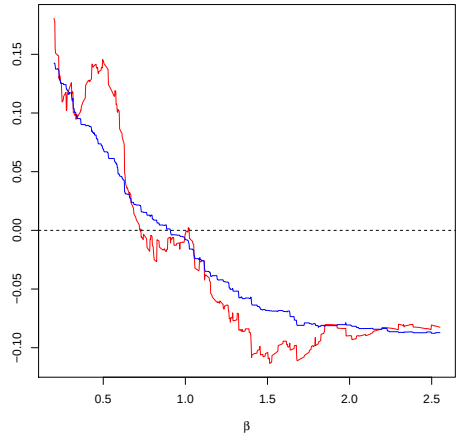
and $\hat{f}_{n,h}(t) = \int K_h(t - u) \, d\hat{F}_{n,\beta}(u)$. *Then*

$$\sqrt{n}\{\hat{\beta}_n - \beta_0\} \xrightarrow{\mathcal{D}} N(0, I_\epsilon(\beta_0)^{-1}),$$

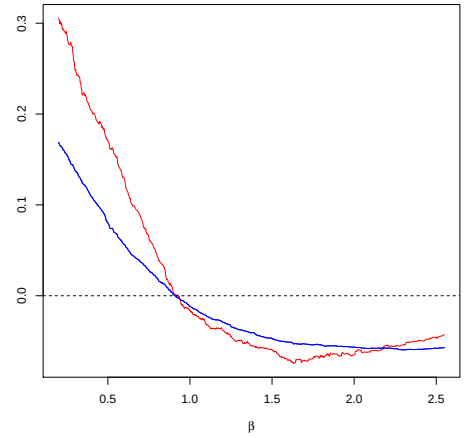
$$I_\epsilon(\beta_0) = \int_{F_0 \in [\epsilon, 1-\epsilon]} \frac{\mathrm{var}(X | T - \beta_0' X = u) f_0(u)^2}{F_0(u)\{1 - F_0(u)\}} \, f_{T - \beta_0' X}(u) \, du \, .$$



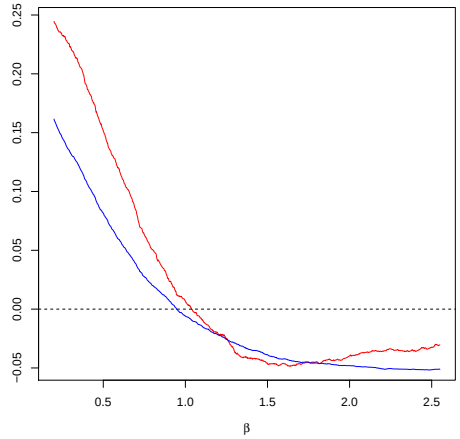
(a) $n = 50$



(b) $n = 100$



(c) $n = 500$



(d) $n = 1000$

Figure: Efficient score (red) and the simple score (blue) for sample sizes $n = 50, 100, 500$ and 1000. The true regression parameter $\beta_0 = 1$.

Performance of estimates of β

	Plug-in, β via argmax		MLE, β via score=0		MLE, β via argmax	
n	$mean(\hat{\beta}_n)$	$nvar(\hat{\beta}_n)$	$mean(\hat{\beta}_n)$	$nvar(\hat{\beta}_n)$	$mean(\hat{\beta}_n)$	$nvar(\hat{\beta}_n)$
100	0.499562	0.245172	0.502964	0.362099	0.487057	0.405507
500	0.498857	0.191857	0.500167	0.212293	0.498111	0.306398
1000	0.499502	0.192223	0.500178	0.194511	0.499528	0.296097
5000	0.500314	0.181421	0.500058	0.176652	0.500162	0.227073
10000	0.500120	0.172043	0.499989	0.174698	0.500159	0.233758
20000	0.500096	0.174197	0.500001	0.169884	0.500050	0.236070

Table: $h = 0.5n^{-1/5}$ and $N = 1000$. $(I_\epsilon(\beta_0)^{-1} = 0.158699,\ \epsilon = 0.001)$

Efficiency

Conclusion: We can construct an asymptoticallly efficient estimate with the non-smoothed MLE, using an estimate of the density based on the MLE.
This uses isotonic estimators and reordering properties.

Bibliography I

Jason Abrevaya. Rank regression for current-status data: asymptotic normality. *Statist. Probab. Lett.*, 43(3):275–287, 1999. ISSN 0167-7152. URL http://dx.doi.org/10.1016/S0167-7152(98)00267-3.

Jorge Aragón and Adolfo J. Quiroz. Rank regression for current status data. *Statist. Probab. Lett.*, 24(3):251–256, 1995. ISSN 0167-7152. URL http://dx.doi.org/10.1016/0167-7152(94)00180-G.

Stephen R. Cosslett. Distribution-free maximum likelihood estimator of the binary choice model. *Econometrica*, 51(3): 765–782, 1983. ISSN 0012-9682. URL http://dx.doi.org/10.2307/1912157.

Stephen R. Cosslett. Efficient estimation of semiparametric models by smoothed maximum likelihood. *Internat. Econom. Rev.*, 48 (4):1245–1272, 2007. ISSN 0020-6598. URL http://dx.doi.org/10.1111/j.1468-2354.2007.00461.x.

Bibliography II

Jeff Dominitz and Robert P. Sherman. Some convergence theory for iterative estimation procedures with an application to semiparametric estimation. *Econometric Theory*, 21(4):838–863, 2005. ISSN 0266-4666. URL `http://dx.doi.org/10.1017/S0266466605050425`.

Aaron K Han. Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator. *Journal of Econometrics*, 35(2-3):303–316, 1987.

Jian Huang. Efficient estimation for the proportional hazards model with interval censoring. *Ann. Statist.*, 24(2):540–568, 1996. ISSN 0090-5364. URL `http://dx.doi.org/10.1214/aos/1032894452`.

Roger W. Klein and Richard H. Spady. An efficient semiparametric estimator for binary response models. *Econometrica*, 61(2): 387–421, 1993. ISSN 0012-9682. URL `http://dx.doi.org/10.2307/2951556`.

Bibliography III

Gang Li and Cun-Hui Zhang. Linear regression with interval censored data. *Ann. Statist.*, 26(4):1306–1327, 1998. ISSN 0090-5364. URL `http://dx.doi.org/10.1214/aos/1024691244`.

S. A. Murphy, A. W. van der Vaart, and J. A. Wellner. Current status regression. *Math. Methods Statist.*, 8(3):407–425, 1999. ISSN 1066-5307.

Robert P. Sherman. The limiting distribution of the maximum rank correlation estimator. *Econometrica*, 61(1):123–137, 1993. ISSN 0012-9682. URL `http://dx.doi.org/10.2307/2951780`.