Chapter 2. Semi-parametric Models (I)
Part 3

February 14, 2007

1 The Single-index model

Recall that the linear regression model is

$$Y = \beta_0 + \beta_1 \mathbf{x}_1 + \cdots + \beta_p \mathbf{x}_p + \varepsilon.$$

where $\beta_0, \beta_1, ..., \beta_q$ are parameters.

Brillinger (1983) considered *the generalized linear model* by imposing a link function $\phi(.)$ and

$$Y = \phi(\beta_0 + \beta_1 \mathbf{x}_1 + \cdots + \beta_p \mathbf{x}_p) + \varepsilon. \tag{1.1}$$

where the link function is known. He called these kind of model the generalized linear regression model.

If $\phi(.)$ is unknown, then we call model (1.1) *the single-index model* (SIM) following Stoker (1986). For model identification, we can write it as

$$Y = \phi(\beta_1 \mathbf{x}_1 + \cdots + \beta_p \mathbf{x}_p) + \varepsilon. \tag{1.2}$$

or

$$Y = \phi({\alpha_0}^\top X) + \varepsilon. \tag{1.3}$$

where $\alpha_0 = (\beta_1, \cdots, \beta_p)^\top$, $E(\varepsilon|X) = 0$ and $||\alpha_0||^2 = \beta_1^2 + ... + \beta_p^2 = 1$.

As an illustration of the SIM, consider the following latent dependent variable model. In this model, we do not observe Y^* but Y, which is a transformation of Y^*. Formally,

$$Y^* = \beta_0 + \beta_1 \mathbf{x}_1 + \cdots + \beta_p \mathbf{x}_p + \epsilon, \quad Y = \tau(Y^*).$$

If the function $\tau(.)$ takes the form

$$\tau(s) = \begin{cases} s, & \text{if } s > c, \\ 0, & \text{if } s \le c, \end{cases}$$

for some constant c, then it is the *Tobit model*.

If the function $\tau(.)$ takes the form

$$\tau(s) = \begin{cases} 1, & \text{if } s > c, \\ 0, & \text{if } s \le c, \end{cases}$$

then it is the *binary choice model*.

Assume that X is independent of ϵ. In both case, we have

$$E(Y|X = x) = E(\tau(Y^*)|X = x) = E(\tau({\alpha_0}^\top x + \epsilon)) = \int \tau({\alpha_0}^\top x + v) f_\epsilon(v) dv$$

where $f_\epsilon(v)$ is the density function of ϵ. Therefore, we have the single-index model with $\phi({\alpha_0}^\top x) = \int \tau({\alpha_0}^\top x + v) f_\epsilon(v) dv$.

Example 1.1 (Swiss banknotes data) *The data contains 6 explanatory variables which are certain measures of Swiss banknotes including measures on the Length, Left, Right, Bottom, Top and Diagonal, denoted by* $\mathbf{x}_1, \cdots, \mathbf{x}_6$ *respectively. The response variable* Y *is coded as 0 and 1 describing whether a banknote is genuine or not. There are 200 banknotes. The first 100 banknotes are genuine, and the others are counterfeit.*

If we applied the linear regresion model to classify the banknotes, we will have at least one misclassification.

Now, we try a single-index model

$$Y = \phi({\alpha_0}^\top X) + \varepsilon$$

where $X = (\mathbf{x}_1, \cdots, \mathbf{x}_6)^\top$

Applied the single-index model, we have the estimator of α_0 *and the link function as shown in figure 1. If we classify the banknotes according to* $\hat{\phi}(\hat{\alpha_0}^\top x) > 0.5$ *or* ≤ 0.5*, we can classify the banknotes accurately.*

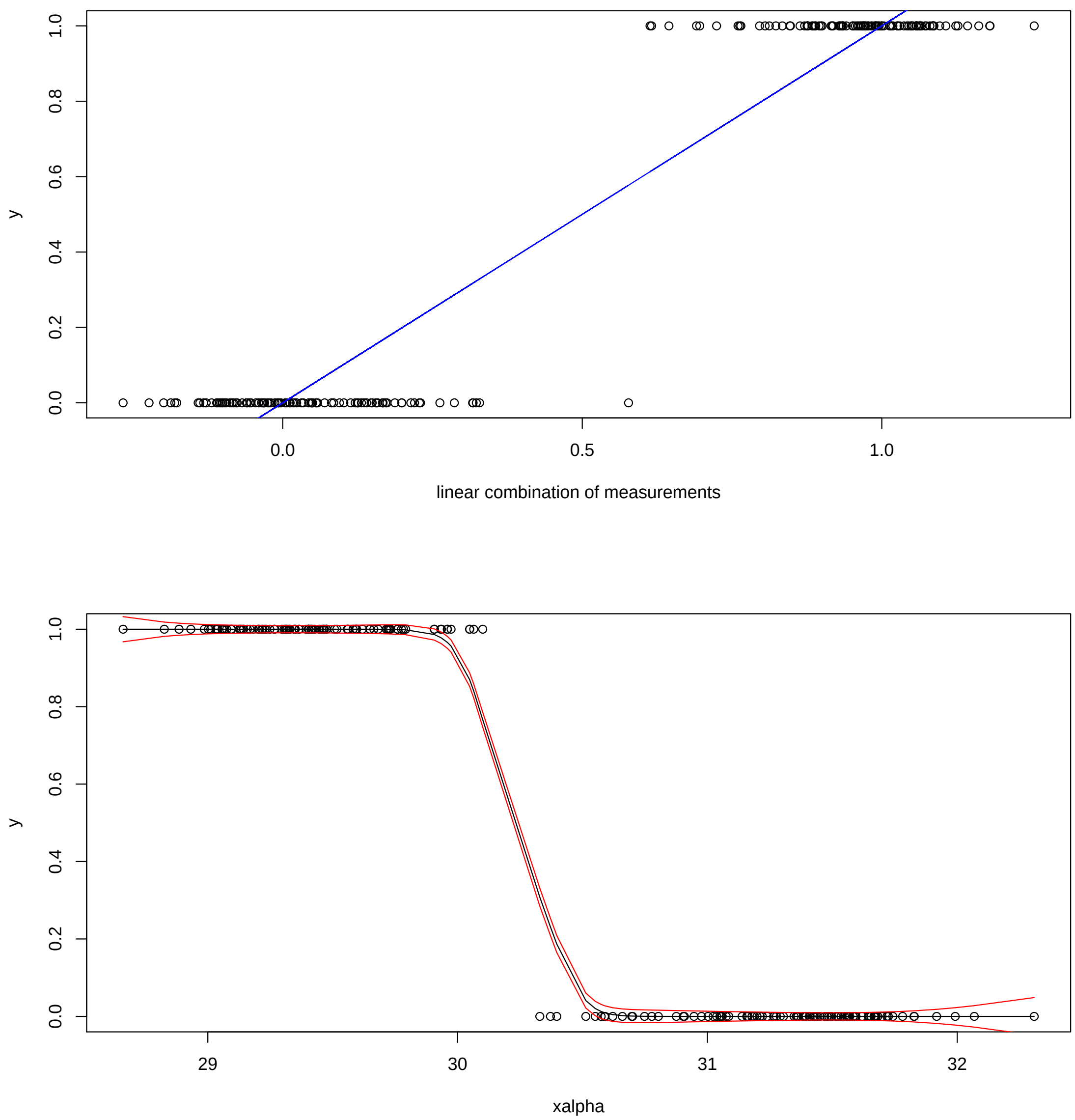

Figure 1: The upper panel, linear regression model is used for the classification. The lower panel, the single-index model is used. The linear regression model has one misclassification. The banknotes are correctly classified by the single-index model. **(sim.R) (c2c1.R)**

2 Estimation of SIM

Let $m(x) = E(Y|X = x)$. By the model, we have

$$m(x) = \phi({\alpha_0}^\top x).$$

Consider the partial derivative[1]. we have

$$\frac{\partial m(x)}{\partial x_k} = \phi'({\alpha_0}^\top x)\beta_k, \quad k = 1, 2, \cdots, p$$

in other words,

$$\begin{pmatrix} \frac{\partial m(x)}{\partial x_1} \\ \frac{\partial m(x)}{\partial x_2} \\ \cdots \\ \frac{\partial m(x)}{\partial x_p} \end{pmatrix} = \phi'({\alpha_0}^\top x) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdots \\ \beta_p \end{pmatrix} = \phi'({\alpha_0}^\top x)\alpha_0 \tag{2.4}$$

Therefore, we can estimate the direction α_0 by the following two methods. Suppose we have random sample $(X_i, Y_i), i = 1, ..., n$.

1. The average derivative estimation method (Härdle and Stoker, 1989). Recall the local linear kernel estimator of $m(x) = E(Y|X = x)$ is as follows. The local linear expansion of $m(X_i)$ at any point x is

$$m(X_i) \approx m(x) + \frac{\partial m(x)}{\partial x_1}(\mathbf{x}_{i1} - x_1) + ... + \frac{\partial m(x)}{\partial x_p}(\mathbf{x}_{ip} - x_p).$$

To estimate $m(x)$ and $m_k(x) \overset{def}{=} \partial m(x)/\partial x_k, k = 1, ..., p$, we use the weighted least squares and minimize

$$\sum_{i=1}^{n} \{Y_i - a - b_1(\mathbf{x}_{i1} - x_1) - ... - b_p(\mathbf{x}_{i1} - x_1)\}^2 K_h(X_i - x).$$

Let

$$\mathbf{X} = \begin{pmatrix} 1 & \mathbf{x}_{11} - x_1 & ... & \mathbf{x}_{1p} - x_p \\ 1 & \mathbf{x}_{21} - x_1 & ... & \mathbf{x}_{2p} - x_p \\ ... & & & \\ 1 & \mathbf{x}_{n1} - x_1 & ... & \mathbf{x}_{np} - x_p \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ ... \\ Y_n \end{pmatrix}$$

[1] Partial derivatives are defined as derivatives of a function of multiple variables when all but the variable of interest are held fixed during the differentiation. For example, $m(x_1, x_2, x_3)$, then

$$\frac{\partial m(x_1, x_2, x_3)}{\partial x_1} = \lim_{h \to 0} \frac{m(x_1 + h, x_2, x_3) - m(x_1, x_2, x_3)}{h}$$

and $\mathbf{W} = diag(K_h(X_1 - x), K_h(X_2 - x), ..., K_h(X_n - x))$. The minimizer to the above minimization problem is the estimator of $m(x)$ and its derivatives

$$\begin{pmatrix} \hat{m}(x) \\ \hat{m}_1(x) \\ ... \\ \hat{m}_p(x) \end{pmatrix} = \{\mathbf{X}^\top \mathbf{W} \mathbf{X}\}^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{Y}.$$

By (2.4), we have

$$\begin{pmatrix} \hat{m}_1(x) \\ ... \\ \hat{m}_p(x) \end{pmatrix} \approx \phi({\alpha_0}^\top x)\alpha_0.$$

Thus

$$n^{-1} \sum_{i=1}^n \begin{pmatrix} \hat{m}_1(X_i) \\ ... \\ \hat{m}_p(X_i) \end{pmatrix} \approx n^{-1} \sum_{i=1}^n \phi({\alpha_0}^\top X_i)\alpha_0$$

We can estimate α_0 by the direction of the above vector. i.e.

$$\hat{\alpha_0} = n^{-1} \sum_{i=1}^n \begin{pmatrix} \hat{m}_1(X_i) \\ ... \\ \hat{m}_p(X_i) \end{pmatrix} /||n^{-1} \sum_{i=1}^n \begin{pmatrix} \hat{m}_1(X_i) \\ ... \\ \hat{m}_p(X_i) \end{pmatrix} ||.$$

2. the Out product of gradient (OPG) method. Note that

$$O \stackrel{def}{=} E\{ \begin{pmatrix} m_1(X) \\ ... \\ m_p(X) \end{pmatrix} \begin{pmatrix} m_1(X) \\ ... \\ m_p(X) \end{pmatrix}^\top \} = E[\phi'({\alpha_0}^\top X)]^2 \alpha_0 {\alpha_0}^\top.$$

Therefore, α_0 is the vector corresponding to the largest eigenvalue of the Out product of gradient. We can estimate the Out product of gradient by

$$\hat{O} = n^{-1} \sum_{i=1}^n \begin{pmatrix} m_1(X_i) \\ ... \\ m_p(X_i) \end{pmatrix} \begin{pmatrix} m_1(X_i) \\ ... \\ m_p(X_i) \end{pmatrix}^\top$$

The estimator $\hat{\alpha_0}$ is the eigenvector of $\hat{O}$ corresponding to the largest eigenvalue.

3. Minimizing the fitted errors. Given any α, we can estimate the link function $\phi(v)$ by kernel smoothing

$$\hat{\phi}_\alpha(v) = \frac{\sum_{i=1}^n K_h(\alpha^\top X_i - v)Y_i}{\sum_{i=1}^n K_h(\alpha^\top X_i - v)}$$

Then, α_0 is estimated by the minimiser of

$$\min_\alpha \sum_{i=1}^n \{Y_i - \hat{\phi}_\alpha(\alpha^\top X_i)\}^2.$$

3 Simulations and Examples for real data analysis

Example 3.1 (simulation) *We consider the following model*

$$Y = (\mathbf{x}_1 + 2\mathbf{x}_3 - 2\mathbf{x}_5)^2 + \varepsilon$$

where $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \varepsilon \sim N(0,1)$ *are IID. In the model. This is a single-index model with*

$$\alpha_0 = (1/3, 0, 2/3, 0, -2/3)^\top$$

and

$$\phi(v) = v^2.$$

50 random samples are drawn from the model.

Using the OPG method, the estimated parameters is

$$\hat{\alpha} = (0.336880614, 0.009664347, 0.665012391, 0.013110368, -0.666336769)^\top$$

and the link funciton is shown in Figure 2.

Using the minimization method, the estimated parameters is

$$\hat{\alpha} = (0.313257837, 0.021758737, 0.673364884, 0.002026994, -0.669306887)^\top$$

and the link funciton is shown in Figure 3.

References

Härdle, W. and Stoker, T. M. (1989) Investigating smooth multiple regression by method of average derivatives. *J. Amer. Stat. Ass.* **84**, 986-995.

Ichimura, H. (1993) Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics* **58**, 71-120.

Nishiyama, Y. and Robinson, P. M. (2005) The bootstrap and the Edgeworth correction for semiparametric averaged derivatives. *Econometrica*, **73**, 903-948.

Xia, Y (2005) On efficiencies of estimations for single index models. *Econometric Theory.*

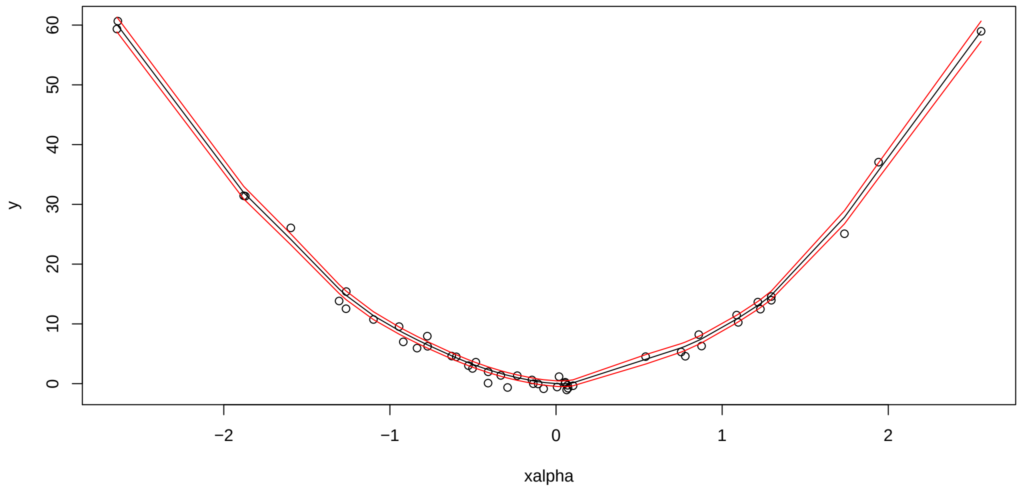

Figure 2: The estimation using OPG method. the dots are the observed values of Y and the curve is the fitted link function. **(sim.R) (c2c2.R)**

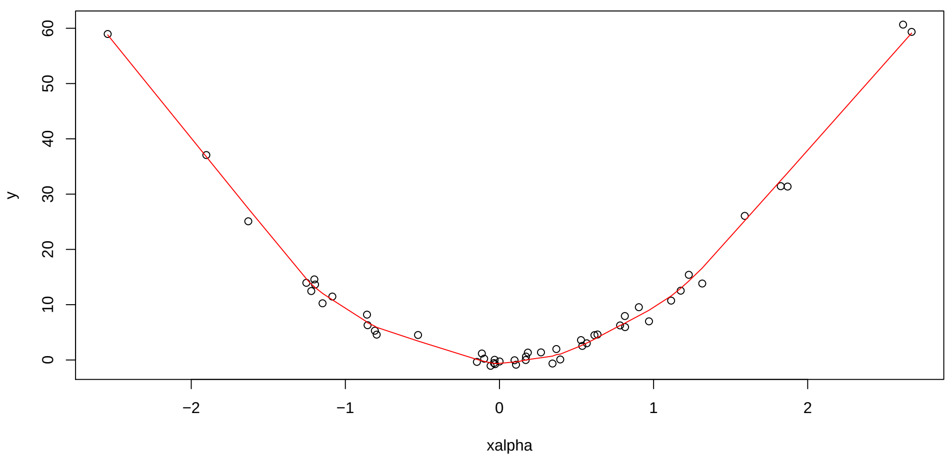

Figure 3: The estimation using minimization method (ppr code). the dots are the observed values of Y and the curve is the fitted link function. **(c2c2.R)**