1 Semiparametric Partially Linear Models

Consider the semiparametric partially linear model

$$Y_i = X_i\beta + g(Z_i) + u_i, \; i = 1, 2, \cdots, n,$$

where X_i is a $1 \times p$ vector, β is a $p \times 1$ vector of unknown parameters, and $Z_i \in R^q$. The data is i.i.d. with $E[u_i|X_i, Z_i] = 0$ and $E[u_i^2|X_i, Z_i] = \sigma^2(X_i, Z_i)$. The motivation is to obtain a root-n consistent estimator of β and an estimator of $g(x)$.

In the model, some components (i.e. $X_i\beta$) are parametric while the remailing components (i.e. $g(\cdot)$ and the distribution of u) are left unspecified.

Partially linear models have many applications. Engle, Granger, Rice and Weiss (1986) were among the frst to consider the partially linear model. They analyzed the relationship between temperature and electricity usage. They used data based on the monthly electricity sales y_i for four cities, the monthly price of electricity x_1, income x_2, and average daily temperature t. They modeled the electricity demand y as the sum of a smooth function g of monthly temperature t, and a linear function of x_1 and x_2, as well as with 11 monthly dummy variables $x_3, \cdots, x_{13}$. That is, their model was

$$y = \sum_{j=1}^{13} \beta_j x_j + g(t) + \epsilon,$$

where g is a smooth function.

Identification:

X cannot contain a constant (i.e. β cannot contain an intercept). Otherwise, the intercept could not be identified separately from the unknown function $g(\cdot)$.

None of the components of X can be a deterministic function of Z. Otherwise, the corresponding part of this component in $X\beta$ cannot be identified separately from the unknown function $g(\cdot)$.

Generally, the identification condition is: The matrix $\Phi \equiv E[(X - E(X|Z))'(X - E(X|Z))] > 0$.

Estimation of the Parametric Part:

- **Robinson's Estimator.** see Robinson (1988), "Root-n consistent semiparametric regression", Econometrica 56, 931-954.

Since

$$Y_i = X_i\beta + g(Z_i) + u_i,$$
$$E[Y_i|Z_i] = E[X_i|Z_i]\beta + g(Z_i),$$

we obtain

$$Y_i - E[Y_i|Z_i] = (X_i - E[X_i|Z_i])\,\beta + u_i.$$

Applying the least squares method, we get an infeasible estimator of β :

$$\hat{\beta}_{\text{inf}} = \left[\sum_{i=1}^{n} (X_i - E[X_i|Z_i])'\,(X_i - E[X_i|Z_i])\right]^{-1} \sum_{i=1}^{n} (X_i - E[X_i|Z_i])'\,(Y_i - E[Y_i|Z_i])\,.$$

By the Lindeberg-Levy CLT,

$$\sqrt{n}\left(\hat{\beta}_{\text{inf}} - \beta\right) \to N\left(0, \Phi^{-1}\Psi\Phi^{-1}\right),$$

where

$$\Phi \equiv E[(X - E(X|Z))'(X - E(X|Z))] > 0$$

and

$$\Psi \equiv E[\sigma^2(X_i, Z_i)\,(X_i - E[X_i|Z_i])'\,(X_i - E[X_i|Z_i])].$$

To present a feasible estimator, we replace the unknown conditional expectations $E[X_i|Z_i]$ and $E[Y_i|Z_i]$ in $\hat{\beta}_{\text{inf}}$ with their consistent kernel estimators

$$\hat{X}_i \equiv \hat{E}[X_i|Z_i] = \frac{n^{-1}\sum_{j=1}^{n} X_j K_h(Z_i, Z_j)}{n^{-1}\sum_{j=1}^{n} K_h(Z_i, Z_j)}$$

and

$$\hat{Y}_i \equiv \hat{E}[Y_i|Z_i] = \frac{n^{-1}\sum_{j=1}^{n} Y_j K_h(Z_i, Z_j)}{n^{-1}\sum_{j=1}^{n} K_h(Z_i, Z_j)},$$

respectively. Here $K_h(Z_i, Z_j) = \prod_{s=1}^{q} h_s^{-1} k\left((Z_{si} - Z_{sj})/h_s\right)$. Then we obtain a feasible estimator of β :

$$\hat{\beta} = \left[\sum_{i=1}^{n} \left(X_i - \hat{X}_i\right)' \left(X_i - \hat{X}_i\right)\right]^{-1} \sum_{i=1}^{n} \left(X_i - \hat{X}_i\right)' \left(Y_i - \hat{Y}_i\right) 1\{\hat{f}(Z_i) \geq b\},$$

where $b \equiv b_n > 0$ satisfies $\lim_{n\to\infty} b_n = 0$. In application we can ignore the trimming parameter b. The feasible estimator $\hat{\beta}$ has the same asymptotic properties as its infeasible counterpart $\hat{\beta}_{\text{inf}}$, i.e.

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \to N\left(0, \Phi^{-1}\Psi\Phi^{-1}\right),$$

under the following conditions:

(i) $(Y_i, X_i, Z_i), i = 1, , 2, \cdots, n$ are i.i.d. observations, Z_i admits a PDF $f \in G_{v-1}^{\infty}$, $g(\cdot) \in G_v^4$, and $E[X|z] \in G_v^4$. Here $v > 1$ is an integer and G_v^{α} denotes the class of smooth functions such that if $g \in G_v^{\alpha}$, then g is v times differentiable with its partial derivative functions (up to order $v(\geq 0)$) satisfying some Lipschitz-type conditions: $|g(z) - g(z')| \leq H_g(z)|z - z'|$, where $H_g(z)$ is continuous having finite αth moment;

(ii) $E[u|x, z] = 0, E[u^2|x, z] = \sigma^2(x, z)$ is continuous in z, both X and u have finite fourth moments;

(iii) Kernel function $k(\cdot)$ is a bounded vth order kernel, and $k(t) = O((1 + |t|)^{-v-1})$;

(iv) As $n \to \infty$, $n(h_1 \cdots h_q)^2 b^4 \to \infty$, $nb^{-4} \sum_{s=1}^{q} h_s^{4v} \to 0$.

Notes:

1. Condition (i): smoothness and moment conditions. $g(\cdot)$ and $E[X|z]$ are vth order differentiable. This and Condition (iii) imply that the bias of the kernel estimator is of order $O(\sum_{s=1}^{q} h_s^v)$. Condition (iv) (ignoring b) is equivalent to

$$\sqrt{n}\left[\sum_{s=1}^{q} h_s^{2v} + (nh_1 \cdots h_q)^{-1}\right] \to 0 \text{ as } n \to \infty.$$

   We have known that $O\left(\sum_{s=1}^{q} h_s^{2v} + (nh_1 \cdots h_q)^{-1}\right)$ is the order of the nonparametric MSE and $\hat{\beta}_{\text{inf}}$ is a root-n consistent estimator of β. For $\hat{\beta}$ to be a root-n consistent estimator of β, Condition (iv) is required.

2. Assume that $v = 2$ and $h_s = h$. Condition (iv) becomes

$$\sqrt{n}\left[\sum_{s=1}^{q} h_s^{4} + (nh_1 \cdots h_q)^{-1}\right] \sim \sqrt{n}\left[h^4 + (nh^q)^{-1}\right] = o(1),$$

   which requires that $q < 4$ (or that $q \leq 3$ since q is a positive integer). Therefore, one need to use a higher kernel if $q \geq 4$.

3. One undesirable feature of the Robinson's semiparametric estimator is the use of a trimming function which requires one to choose a nuisance parameter b.

- **Density-weighted Estimator:** avoid a random denominator in the kernel estimator.

Denote $f_i = f(Z_i)$. Since

$$\begin{aligned} Y_i - E[Y_i|Z_i] &= (X_i - E[X_i|Z_i])\,\beta + u_i, \\ (Y_i - E[Y_i|Z_i])\,f_i &= (X_i - E[X_i|Z_i])\,\beta f_i + u_i f_i, \end{aligned}$$

one can give an infeasible estimator $\hat{\beta}_{\mathrm{inf},f}$ of β using the least squares method by regressing $(Y_i - E[Y_i|Z_i])\,f_i$ on $(X_i - E[X_i|Z_i])\,f_i$.Then

$$\sqrt{n}\left(\hat{\beta}_{\mathrm{inf},f} - \beta\right) \rightarrow N\left(0, \Phi_f^{-1}\Psi_f\Phi_f^{-1}\right),$$

where

$$\Phi_f \equiv E[(X_i - E(X_i|Z_i))'(X_i - E(X_i|Z_i))f_i^2] > 0$$

and

$$\Psi_f \equiv E[\sigma^2(X_i, Z_i)\,(X_i - E[X_i|Z_i])'\,(X_i - E[X_i|Z_i])\,f_i^4].$$

An feasible estimator $\hat{\beta}_f$ of β is obtained by replacing $E[X_i|Z_i]$, $E[Y_i|Z_i]$ and f_i in $\hat{\beta}_{\mathrm{inf},f}$ with their consistent kernel estimators, respectively. Since there does not exist a random denominator f_i, Condition (iv) is replaced by those without the nuisance parameter b, that is,

$$\text{As } \; n \rightarrow \infty, \;\; n(h_1 \cdots h_q)^2 \rightarrow \infty \text{ and } \sum_{s=1}^{q} h_s^{4v} \rightarrow 0.$$

Then we have

$$\sqrt{n}\left(\hat{\beta}_f - \beta\right) \rightarrow N\left(0, \Phi_f^{-1}\Psi_f\Phi_f^{-1}\right).$$

Estimation of Nonparametric Component

After obtaining a root-n consistent estimator of β ($\hat{\beta}$ or $\hat{\beta}_f$), since $g(Z_i) = E[Y_i - X_i\beta|Z_i]$, a consistent estimator of $g(z)$ is given by

$$\hat{g}(z) = \frac{\sum_{i=1}^{n}(Y_i - X_i\hat{\beta})K_h(z, Z_i)}{\sum_{i=1}^{n} K_h(z, Z_i)} \tag{1}$$

or

$$\hat{g}(z) = \frac{\sum_{i=1}^{n}(Y_i - X_i\hat{\beta}_f)K_h(z, Z_i)}{\sum_{i=1}^{n} K_h(z, Z_i)}.$$

Asymptotically, $\hat{g}(z)$ is equivalent to the following infeasible estimator

$$\tilde{g}(z) = \frac{\sum_{i=1}^{n}(Y_i - X_i\beta)K_h(z, Z_j)}{\sum_{i=1}^{n} K_h(z, Z_j)}.$$

$\hat{g}(z)$ and $\tilde{g}(z)$ have the same convergence rate and asymptotic distribution.

Note: The choice of h_s's for estimating $g(\cdot)$ can be quite different from those for estimation β. To obtain a root-n consistent estimator of β, a higher order kernel function is needed if $q \geq 4$. However, there is no need to use a higher order kernel when estimating $g(\cdot)$, regardless of the value of q. A nonnegative second order kernel is good enough to estimate $g(z)$, and the bandwidths can be chosen by least squares cross-validation by minimizing

$$\sum_{i=1}^{n} \left[Y_i - X_i\hat{\beta} - \hat{g}_{-i}(Z_i, h) \right]^2,$$

where

$$\hat{g}_{-i}(Z_i, h) = \frac{\sum_{j\neq i}^{n} (Y_j - X_j\hat{\beta}) K_h(Z_i, Z_j)}{\sum_{j\neq i}^{n} K_h(Z_i, Z_j)}$$

is the leave-one-out estimator in (1).

Since the LSCV bandwidth h is of order $O_p(n^{-1/(q+4)})$, condition (iv) is satisfied if $q \leq 3$. In this case, $\hat{\beta}$ and $\hat{h}$ can be chosen simultaneously by solving the minimization problem:

$$\min_{\beta, h} \sum_{i=1}^{n} \left[Y_i - X_i\beta - \hat{g}_{-i}(Z_i, h) \right]^2.$$

An Empirical Study: see Blundell, Duncan and Pendakur (1998), "Semiparametric estimation and consumer demand", Journal of Applied Econometrics 13, 435-462.

Study the Engle curve, i.e. the relationship between budget shares and total expenditure, which is modelled as

$$w_{ij} = z_i\alpha_j + g_j(\ln x_i) + u_{ij}$$

where w_{ij} is the budget share of the jth good for individual i, z_i is a finite vector of observable exogenous regressors, $\ln x_i$ is the log of total expenditure, the unobservable u_{ij} satisfy that $E[u_{ij}|\ln x_i, z_i] = 0$, $E[u_{ij}^2|\ln x_i, z_i] = \sigma^2(\ln x_i, z_i)$.

Example 5 (Semiparametric Partially Linear Model, see ex5) The data generating processes are:

Design 1: $Y_i = 1 + X_i + \sin(8Z_i + 5) + u_i,\ i = 1, 2, \cdots, n$, where $X_i \sim N(0, 1)$ and $Z_i \sim U[0, 1]$, $u_i \sim N(0, 0.16Z_i)$;

Design 2: $Y_i = 1 + X_i + \sin(8Z_i + 5) + u_i,\ i = 1, 2, \cdots, n$, where $X_i = 2Z_i^{1/2} + 1$ and $Z_i \sim U[0, 1]$, $u_i \sim N(0, 0.16Z_i)$.

Design 3: $Y_i = 1 + X_i + \sin(8Z_i + 5) + u_i,\ i = 1, 2, \cdots, n$, where $X_i = \lambda_i Z_i^{1/2} + 1$ with $\lambda_i \sim N(0, 1)$, $Z_i \sim U[0, 1]$, and $u_i \sim N(0, 0.16Z_i)$.

In each of the three designs, $\beta = 1$ and $g(z) = 1 + \sin(8z + 5)$.The sample size is $n = 400$. The sample are independent. In the estimation, the bandwidths for X and Z are chosen according to the reference normal rule-of-thumb.

What the difference between the estimation results in the Designs 1, 2 and 3? Why the difference?