

Topics in Applied Econometrics for Public Policy

Master in Economics of Public Policy, BSE

Handout 4: Semiparametric Regression

Laura Mayoral

IAE and BSE

Barcelona, Spring 2023

1. Introduction

- Previous handouts: regression models without any structure.
- This gives a lot of **flexibility** but it also has some limitations :
 - Sometimes, theory may place some structure on the data. We might want to incorporate this information in the model
 - We can only include in the analysis a relative small set of variables (curse of dimensionality)
 - ...but incorporating many variables might be needed to avoid endogeneity of regressors
- This lecture: **Semiparametric methods**

Semiparametric models: examples (from Cameron&Trivedi)

■ Many semiparametric models and many methods to estimate them. This is only a short intro to these methods.

Table 9.2. *Semiparametric Models: Leading Examples*

Name	Model	Parametric	Nonparametric
Partially linear	$E[y \mathbf{x}, \mathbf{z}] = \mathbf{x}'\boldsymbol{\beta} + \lambda(\mathbf{z})$	$\boldsymbol{\beta}$	$\lambda(\cdot)$
Single index	$E[y \mathbf{x}] = g(\mathbf{x}'\boldsymbol{\beta})$	$\boldsymbol{\beta}$	$g(\cdot)$
Generalized partial linear	$E[y \mathbf{x}, \mathbf{z}] = g(\mathbf{x}'\boldsymbol{\beta} + \lambda(\mathbf{z}))$	$\boldsymbol{\beta}$	$g(\cdot), \lambda(\cdot)$
Generalized additive	$E[y \mathbf{x}] = c + \sum_{j=1}^k g_j(x_j)$	–	$g_j(\cdot)$
Partial additive	$E[y \mathbf{x}, \mathbf{z}] = \mathbf{x}'\boldsymbol{\beta} + c + \sum_{j=1}^k g_j(z_j)$	$\boldsymbol{\beta}$	$g_j(\cdot)$
Projection pursuit	$E[y \mathbf{x}] = \sum_{j=1}^M g_j(\mathbf{x}'_j \boldsymbol{\beta}_j)$	$\boldsymbol{\beta}_j$	$g_j(\cdot)$
Heteroskedastic linear	$E[y \mathbf{x}] = \mathbf{x}'\boldsymbol{\beta}; V[y \mathbf{x}] = \sigma^2(\mathbf{x})$	$\boldsymbol{\beta}$	$\sigma^2(\cdot)$

are identified. For example, see the discussion of single-index models. In addition to estimation of $\boldsymbol{\beta}$, interest also lies in the marginal effects such as $\partial E[y|\mathbf{x}, \mathbf{z}]/\partial \mathbf{x}$.

Roadmap

1. Partially Linear Models
2. Single Index Models
3. Summary, other models exist . . .

Partially Linear Model

- Partially Linear Model: conditional mean is a linear regression function plus an unspecified nonlinear component.

$$E[y|x, z] = x\beta + \lambda(z) \quad \lambda(\cdot) \text{ unspecified.}$$

- **Model** to be estimated:

$$y = x\beta + \lambda(z) + u, \quad E(u|x, z) = 0. \quad (1)$$

- **Estimation Method: Robinson Difference Estimator**

We will obtain estimates for β and for λ in two steps

- **Step 1:** get rid of $\lambda(z)$ and estimate β (only)
- **Step 2:** Use the estimates of β in a model that will allow us to obtain an estimate for $\lambda(\cdot)$

Robinson Difference Estimator

- Step 1: A) get rid of $\lambda(\cdot)$, B) estimate β

A) get rid of lambda:

- Take conditional expectations (by z) on both sides of Model 1 (and notice that $E(u|z) = 0$):

$$E[y|z] = E[x|z]' \beta + \lambda(z) \quad (2)$$

- Subtract the two equations –eq (1) and eq (2)– and obtain the model:

$$y - E[y|z] = (x - E[x|z])' \beta + u \quad (3)$$

- The conditional moments $E[y|z]$ and $E[x|z]$ are unknown \Rightarrow Replace them by non parametric estimators \hat{m}_{y_i} and \hat{m}_{x_i}

- Robinson's difference estimator:
- Step 1: B) estimate β in the model

$$y - \hat{m}_{y_i} = (x - \hat{m}_{x_i})' \beta + u \quad (4)$$

- The resulting estimator of β is consistent and A.N (assuming u is *i.i.d.*):

$$\sqrt{N}(\hat{\beta}_{PL} - \beta) \xrightarrow{d} N(0, \sigma^2 \left(\text{plim} \frac{1}{N} \sum_{i=1}^N (x_i - E[x_i|z_i])(x_i - E[x_i|z_i])' \right)^{-1})$$

- Notes:

1. Cost of non-specifying λ ? higher variance (efficiency loss) [But no loss if $E(x|z)$ is linear!]

2. The distribution assumes homokedasticity (u is *i.i.d.*). Use Eicker-White standard errors to make it robust to heteroskedasticity

■ Notes (cont).:

3. To estimate the variance: replace $(x_i - E[x_i|z_i])$ by $(x_i - \hat{m}_{x_i})$

4. How to compute \hat{m}_{x_i} and \hat{m}_{y_i} ?

■ Robinson: Kernel estimates with convergence no slower than $N^{-1/4}$.

■ Step 2: Estimate λ

■ Recall that $\lambda(z) = E(y|z) - E(x|z)'\beta$.

■ Estimate $\lambda(z)$ as

$$\hat{\lambda}(z) = \hat{m}_{y_i} - \hat{m}'_{x_i}\beta$$

Summarizing: Robinson difference estimator

- Model: $E[y_i|x_i, z_i] = x_i'\beta + \lambda(z_i)$, unspecified $\lambda(\cdot)$
- Steps:
 1. Kernel regress y on z and get residual $y - \hat{y}$.
 2. Kernel regress x on z and get residual $x - \hat{x}$.
 3. OLS regress $y - \hat{y}$ on $x - \hat{x}$, get $\hat{\beta}$
 4. Combine the estimates in 1) 2) and 3) to get $\hat{\lambda}(z) = \hat{m}_{y_i} - \hat{m}'_{x_i}\hat{\beta}$

An example

- Same data as before: now, wage on marital status and education.

Let's look first at the OLS:

```
regress lnhwage educatn married, vce(robust) noheader
```

lnhwage	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
educatn	.1009005	.0219979	4.59	0.000	.0574834	.1443176
married	.4198385	.1545864	2.72	0.007	.1147326	.7249443
_cons	.6149712	.3219871	1.91	0.058	-.0205319	1.250474

- Robison's estimator:
- STATA command: semipar
- married enters linearly, we allow education to enter nonparametrically

semipar lnhwage married, nonpar(educatn) robust ci title("Partial linear")

- with robust standard errors, confidence intervals...

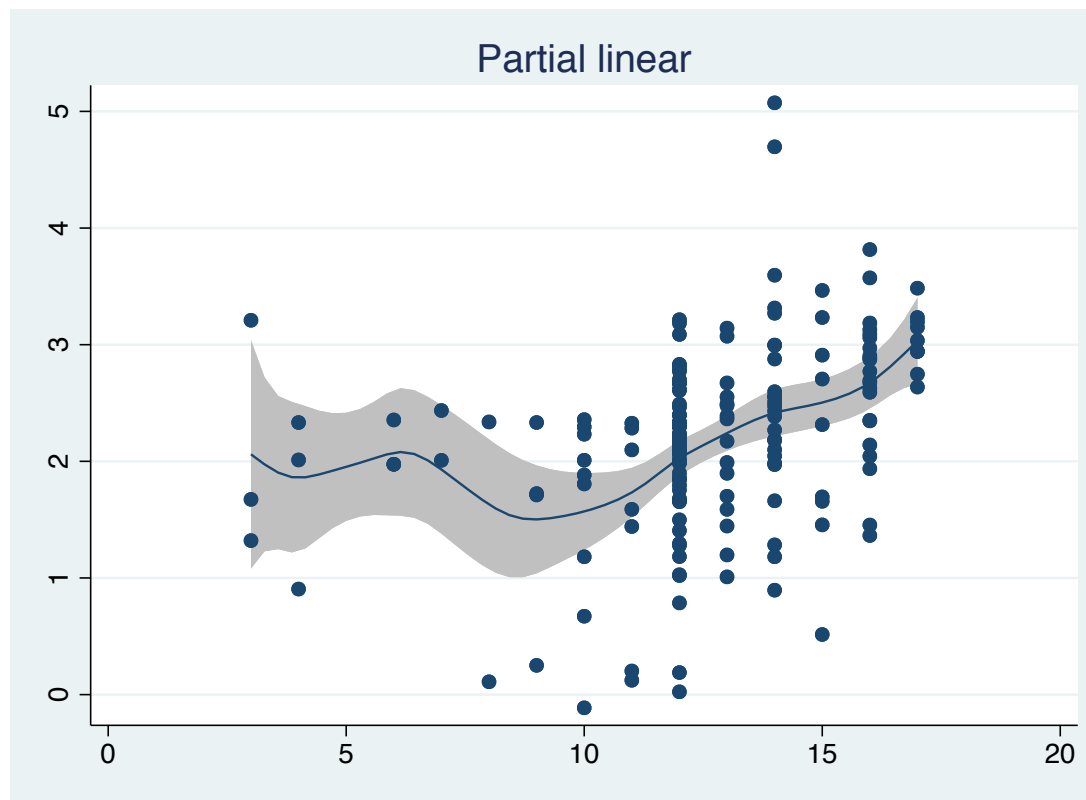
Output has two parts: 1) the parametric component

```
. semipar lnhwage married, nonpar(educat) robust ci title("Partial linear")
```

```
Number of obs = 177
R-squared      = 0.0442
Adj R-squared  = 0.0388
Root MSE      = 0.7134
```

lnhwage	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
married	.357994	.146041	2.45	0.015	.069777	.646211

- ... and the non-parametric component: Plot of $\lambda(z)$ against z where z is education



Trimming

- **Trimming**: we can apply trimming to estimate the nonparametric component.
- In kernel estimation: estimates are not good in areas of the support of z with low density values
- Why? low density=few values to compute the local average
- **Trimming** consists of excluding data points for which the density $f(z) < b$, for some positive value b
- the command `semipar` allows for the introduction of trimming (default is no trimming)

Summarizing

- Partially linear model: additive model with a parametric part and a non-parametric one
- Estimation: Robinson's two step estimator
- Stata: semipar
- Advantages of this method:
 - The model allows "any" form of the unknown
 - $\hat{\beta}$ is \sqrt{n} -consistent

Single Index Models

- Model:

$$E[y_i|x_i] = g(x_i'\beta)$$

where $g(\cdot)$ is not specified.

- Many standard nonlinear (parametric) models such as logit, probit, and Tobit are of single-index form. (In these cases $g(\cdot)$ is known)
- But we can also estimate this model leaving $g(\cdot)$ unspecified and estimate it non-parametrically.

More on interpretation: marginal effects

- For single-index models the effect on the conditional mean of a change in the j th regressor using calculus methods is

$$\frac{\partial E[y|x]}{\partial x_j} = g'(x'\beta)\beta_j, \quad (5)$$

where $g'(z) = \frac{\partial g(z)}{\partial z}$.

- Then, relative effects of changes in regressors are given by the ratio of the coefficients since

$$\frac{\partial E[y|x] / \partial x_j}{\partial E[y|x] / \partial x_k} = \frac{\beta_j}{\beta_k}, \quad (6)$$

because the common factor $g'(x'\beta)$ cancels.

- Thus, if β_j is two times β_k , then a one-unit change in x_j has twice the effect as a one-unit change in x_k .

Estimation with unspecified $g(\cdot)$

- Identification: β can only be identified up to location and scale
- That is, we estimate $a + b\beta_i$
- This is still useful to compute relative marginal effects!
- Ichimura semiparametric least squares: choose β and $g(\cdot)$ that minimizes

$$\operatorname{argmin}_{\beta} \sum_{i=1}^N w(x_i) f(y_i - \hat{g}(x_i' \beta))^2,$$

where $\hat{g}(\cdot)$ is the leave-one-out NW estimator and $w(\cdot)$ is a trimming function that drops outlying x values.

- β can only be estimated up to scale in this model,
- but still useful as ratio of coefficients equals ratio of marginal effects in a single-index models.

Example

- Same data as before. Now: wages on hours worked and education

- STATA command: `sls` (Semiparametric Least Squares from Ichimura, 1993)

- Install it first:

```
ssc install sls
```

- Run both `ols` and `sls` and compare results

lnhwage	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
hours	.0001365	.0000839	1.63	0.106	-.0000292	.0003022
educatn	.1071543	.0206339	5.19	0.000	.0664293	.1478793
_cons	.6437424	.3068995	2.10	0.037	.0380175	1.249467

. sls lnhwage hours educatn, trim(1,99)

initial: SSq(b) = 120.10723

alternative: SSq(b) = 120.1062

rescale: SSq(b) = 98.292016

SLS 0: SSq(b) = 98.292016

SLS 1: SSq(b) = 98.195246

SLS 2: SSq(b) = 98.007811

SLS 3: SSq(b) = 98.007526

SLS 4: SSq(b) = 98.007526

pilot bandwidth

1052.001873

SLS 0: SSq(b) = 99.252078 (not concave)

SLS 1: SSq(b) = 97.285143

SLS 2: SSq(b) = 97.202952

SLS 3: SSq(b) = 97.201992

SLS 4: SSq(b) = 97.201988

Number of obs = 177

root MSE = .741056

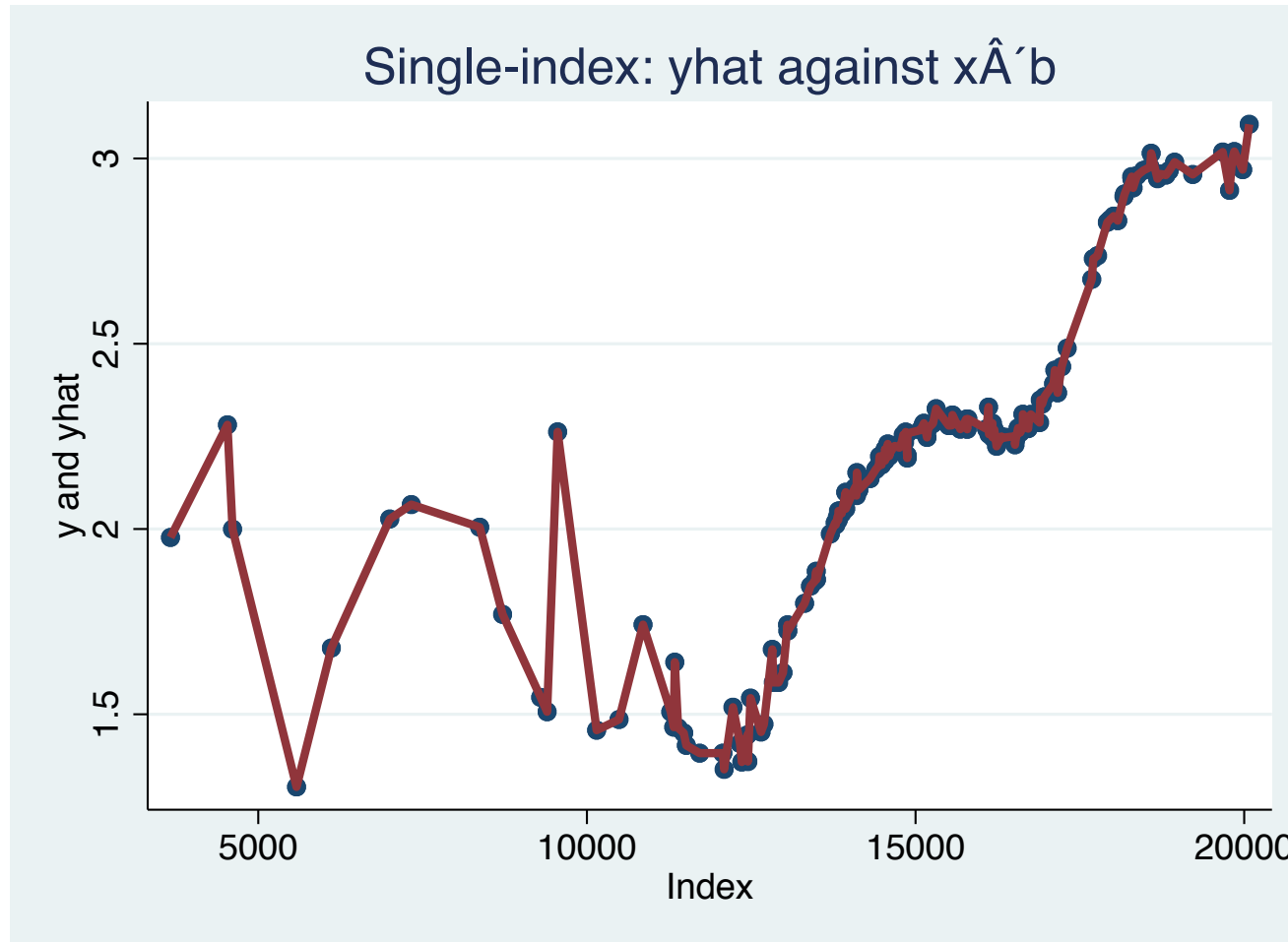
lnhwage	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
Index						
educatn	1048.102	275.9987	3.80	0.000	507.1545	1589.05
hours	1	(offset)				

■ Interpretation:

■ one more year of education has the same effect on log hourly wage as working 1048 more hours a year!

■ Compared to OLS, $0.1071453/0.0001365 = 785$.

- This graph plots the predicted conditional expectation versus $x'\beta$ (highly nonlinear)



Index models, summary

- Generalizes the linear regression model (which assumes $g(\cdot)$ constant)
- Gain over parametric models: more flexible
- Gain over fully non-parametric: only one nonparametric dimension (avoid curse of dimensionality).
- We only identified the parameters up to location and scale but
- The ratio of the coefficients provides the relative marginal effects

Takeaway

- Semiparametric methods aim to overcome some of the limitations of fully parametric and fully nonparametric methods
- Flexible, yet tractable (many variables can be included)
- Literature is very large
- Here, we've only presented a short introduction.
- Partial linear model, single index models ... but many more models are available
- See this document to learn about other semiparametric methods STATA can handle