Topics in Applied Econometrics for Public Policy

Master in Economics of Public Policy, BSE

Handout 3: Nonparametric Series Regression

Laura Mayoral

IAE and BSE

Barcelona, Spring 2023

1. Introduction

- Previous handout: nonparametric Kernel regression.

- Several methods: local constant, local linear, Lowess, K-NN, etc.

- Recall these are local average methods, (i.e., averages of the dependent variable) where the weights employed are Kernel weights.

- Now: a new class of nonparametric regression methods: nonparametric series regression.

- Goal: Same as before, estimate the conditional expectation.

Series Regression

- Model: Consider two random variables (y, x) who are related by,

$$y = m(x) + e \quad (1)$$

where $\mathbb{E}[e|x] = 0$ and $\mathbb{E}[e^2|x] = \sigma^2(x)$.

- Goal: estimate m(.), unspecified

- Idea: approximate m(x) by a flexible function.

- We focus on linear functions (other possibilities also exist but linear functions are simple and work well)

- In particular
  - polynomials
  - splines

Series Regression, II

- Linear series regression models take the form

$$y = X_K' \beta_K + e_K \tag{2}$$

- where X_K is a vector of regressors obtained by transforming x in different ways
- β_K is a coefficient vector.

- We examine next two popular series regression estimators
- Polynomials
- Splines

Polynomial Regression

- Conditional expectation:

- approximated by a polynomial in x of degree p:

$$m_K(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p$$

.

- number of parameters to be estimated is $K = p + 1$

- Simple approach: estimate b_k by OLS

- p: controls the degree of flexibility of a polynomial regression. Tradeoff
  - A large p provides a lot of flexibility
  - But it can become too noisy

Example

- (from Hansen's book, chapter 20)
- Log wages on experience for women with college education (education= 16), separately for white women and Black women

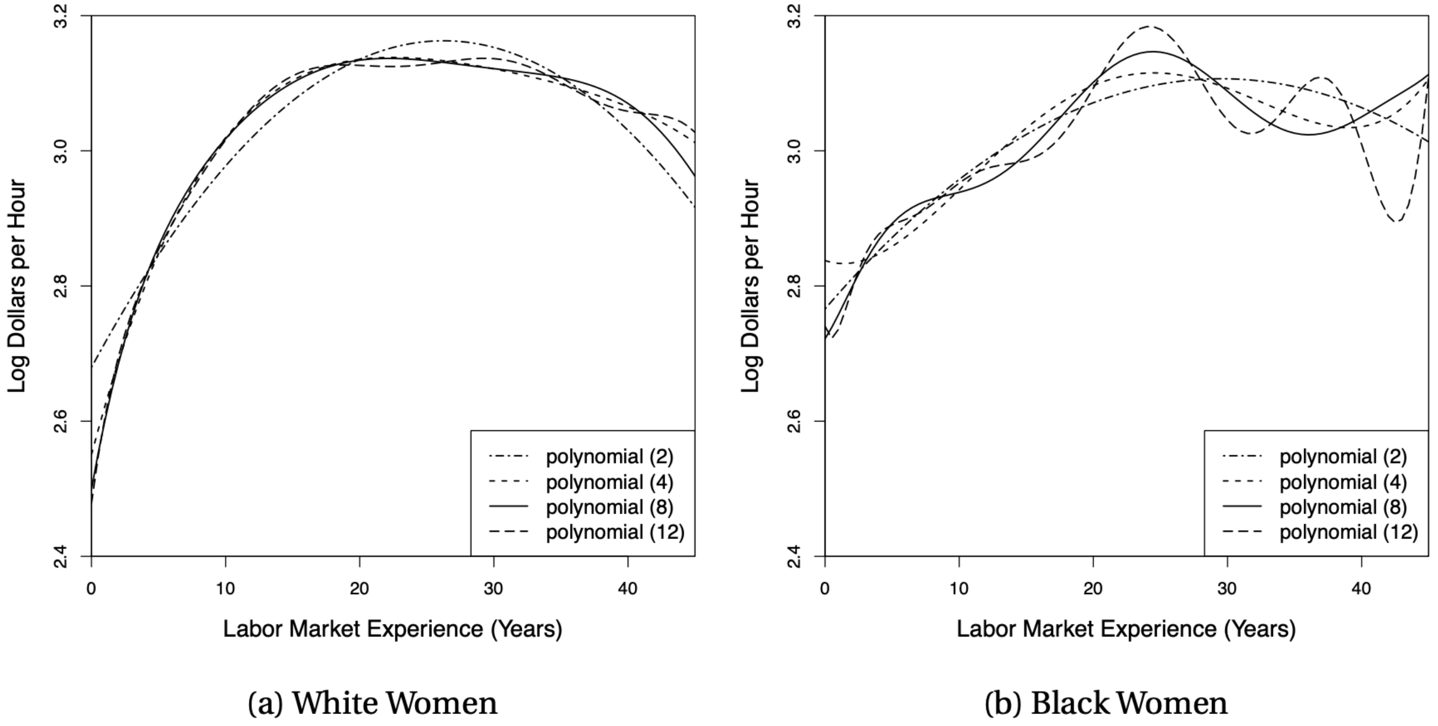


Figure 20.1: Polynomial Estimates of Experience Profile, College-Educated Women

- Difference between the two plots: might be due to the fact that the sub-sample of Black women has much fewer observations

- Then, the mean function is much less precisely estimated, giving rise to the erratic plots

Orthogonal polynomials

- The different regressors $(x^1, x^2, \ldots, x^j \ldots)$ can be highly correlated

- Then the OLS estimator can be difficult to compute (as it needs to invert a near-singular matrix)

- One solution: orthogonalize the polinomial.

- Goal of orthogonal polynomials: get rid of the problem of the inversion of $X_k' X_k$

- How they work: they produce regressors that are close to being orthogonal and have similar variances, which implies that the resulting matrix of orthogonal regressors $X_k^{*\prime} X_k^*$ is diagonal and with similar diagonal values (the variances).

- Then, use this vector of orthogonal regressors rather than X_k.

- There exist different ways of doing these orthogalizations, for instance: 1) sample orthogonalization and) use orthogonal polynomials

- The most popular orthogonal polynomials are:

- Hermite polynomial, Laguerre Polynomial etc.

(See Hansen, chapter 20 for further details)

Implementation in STATA: npregress series

From STATA help:

- npregress series: performs nonparametric series estimation

- Like linear regression, nonparametric regression models the mean of the outcome conditional on the covariates, but unlike linear regression, it makes no assumptions about the functional form of the relationship between the outcome and the covariates.

- Output: average marginal effect

- log wages on years of education.
- stata command: npregress series lnhwage educatn, polynomial
- Output: average effect
- Polynomial order: chosen by cross-validation

```
Minimizing cross-validation criterion

Iteration 0:  Cross-validation criterion =  .6118535
Iteration 1:  Cross-validation criterion =  .5714533

Computing average derivatives

Polynomial-series estimation                  Number of obs      =        177
Criterion: cross-validation                   Polynomial order   =          3
```

lnhwage	Effect	Robust std. err.	z	P>|z|	[95% conf.	interval]
educatn	.1488311	.0195814	7.60	0.000	.1104523	.1872099

Note: Effect estimates are averages of derivatives.

- Different output if regressor is continuous or discrete. Education has 14 different values. Now we enter it in the model as discrete

npregress series lnhwage i.educatn, polynomial

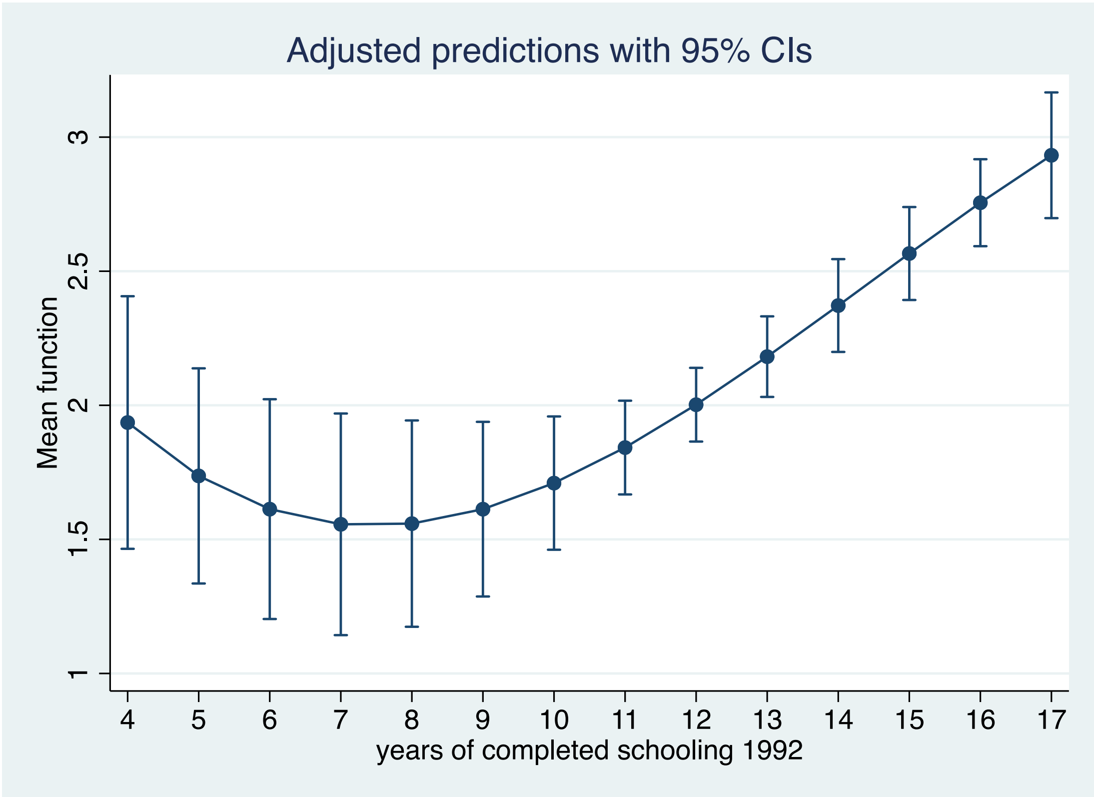
```
. npregress series lnhwage i.educatn, polynomial

Computing approximating function

Minimizing cross-validation criterion

Iteration 0:  Cross-validation criterion =  .6576989

Computing average derivatives

Polynomial-series estimation              Number of obs      =            177
Criterion: cross-validation               Polynomial order   =              1
```

lnhwage	Effect	Robust std. err.	z	P>|z|	[95% conf. interval]	
educatn						
(4 vs 3)	-.43711	.5819696	-0.75	0.453	-1.57775	.7035295
(6 vs 3)	.0336978	.504947	0.07	0.947	-.9559802	1.023376
(7 vs 3)	-.0246969	.4939832	-0.05	0.960	-.9928862	.9434923
(8 vs 3)	-1.022982	1.072799	-0.95	0.340	-3.125629	1.079666
(9 vs 3)	-.6525539	.6075096	-1.07	0.283	-1.843251	.5381431
(10 vs 3)	-.6353192	.5691324	-1.12	0.264	-1.750798	.4801598
(11 vs 3)	-.7832856	.6340793	-1.24	0.217	-2.026058	.459487
(12 vs 3)	-.0904436	.5003293	-0.18	0.857	-1.071071	.8901838
(13 vs 3)	-.0444978	.515248	-0.09	0.931	-1.054365	.9653696
(14 vs 3)	.3913532	.5309523	0.74	0.461	-.6492942	1.432001
(15 vs 3)	.0298562	.5860523	0.05	0.959	-1.118785	1.178498
(16 vs 3)	.5577673	.5102761	1.09	0.274	-.4423555	1.55789
(17 vs 3)	.853556	.5015454	1.70	0.089	-.1294549	1.836567

Note: Effect estimates are averages of contrasts of factor covariates.

Estimated function at different data points

npregress series lnhwage educatn, polynomial margins, at(educatn=(4 5 6 7 8 9 10 11 12 13 14 15 16 17)) marginsplot)



Splines

- A spline is a piecewise polynomial.

- Order of polynomial: pre-selected to be linear, quadratic, or cubic.

- The flexibility of the model: determined by the number of polynomial segments.

- The join points between the segments are called knots.

- If there's 1 knot, there are two segments, etc.

Splines, II

- How to construct a spline?

- Choose p (order of the polynomial), typically p=1, 2 or 3

  - A quadratic or cubic spline is useful when it is desired to impose smoothness

  - a linear spline is useful when it is desired to allow for sharp changes in slope.

- Choose number of knots

Examples

- Example 1: a linear spline with one knot τ:

(we allow the slope to change once)

$$m_K(x) = \beta_0 + \beta_1 x + \beta_2 (x - \tau)\mathbb{1}_{(x \geq \tau)}$$

Notice that;

- for $x < \tau$, $m_K(x) = \beta_0 + \beta_1 x$ is linear with slope β_1;

- for $x \geq \tau$, $m_K(x)$ is linear with slope $\beta_1 + \beta_2$; and the function is continuous at $x = \tau$.

- β_2 is the change in the slope at τ.

- Example 2: A linear spline with two knots $\tau_1 < \tau_2$:

(The knots allow the slope to change twice)

$$m_K(x) = \beta_0 + \beta_1 x + \beta_2 (x - \tau_1)\mathbb{1}(x \geq \tau_1) + \beta_3 (x - \tau_2)\mathbb{1}(x \geq \tau_2)$$

.

- Example 3: quadratic spline with one knot is (we allow the coefficient of x^2 to change once)

$$m_K(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 (x - \tau)^2 \cdot 1(x \geq \tau)$$

- In general, a pth-order spline with N knots $\tau_1 < \tau_2 < \cdots < \tau_N$ is

$$m_K(x) = \sum_{j=0}^{N+p-1} \beta_j x^j + \sum_{k=1}^{N} \beta_{p+k} (x - \tau_k)^p \cdot 1x \geq \tau_k$$

- Important: select the number and location of knots.
- As usual, many options for doing this
- Simplest: evenly spaced

Example 1 (Hansen, Chapter 20)

- Graph plots log wages on experience for Black women (394 obs.)

- quadratic spline (smooth changes)

- four equally-spaced knots at experience levels of 10, 20, 30, and 40 (7 coefficients)

- For comparison: 6th order polynomial regression (also 7 coefficients).

(a) Experience Profile

- Interpretation example 1:

- the spline is a quadratic over each segment, but the first two segments (experience levels between 0-10 and 10-20 years) are essentially linear.

- Most of the curvature occurs in the third and fourth segments (20-30 and 30-40 years) where the estimated regression function peaks and twists into a negative slope.

- The estimated regression function is smooth.

Example 2 (Hansen, Chapter 20)

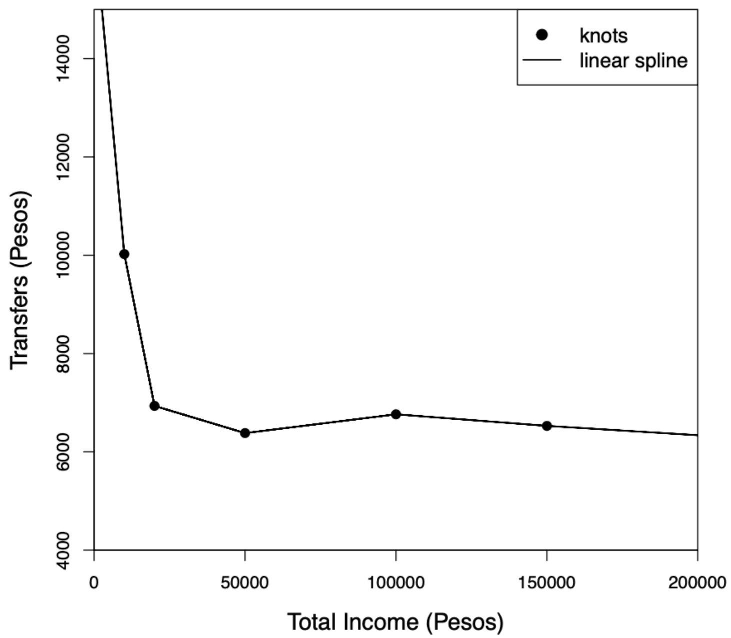
- A model of altruistic transfers: transfers of extended family. vs. income family.

- Model predicts that extended families will make gifts (transfers) when the recipient family's income is low, but will not make transfers if the recipient family's income exceeds a threshold.

- A pure altruistic model predicts that the regression of transfers received on family income should have a slope of 1 up to this threshold and be flat above this threshold.

(sharp changes)

- linear spline with knots at 10000, 20000, 50000, 100000, and 150000 pesos.



(b) Effect of Income on Transfers

Splines in STATA

npregress series lnhwage educatn, spline

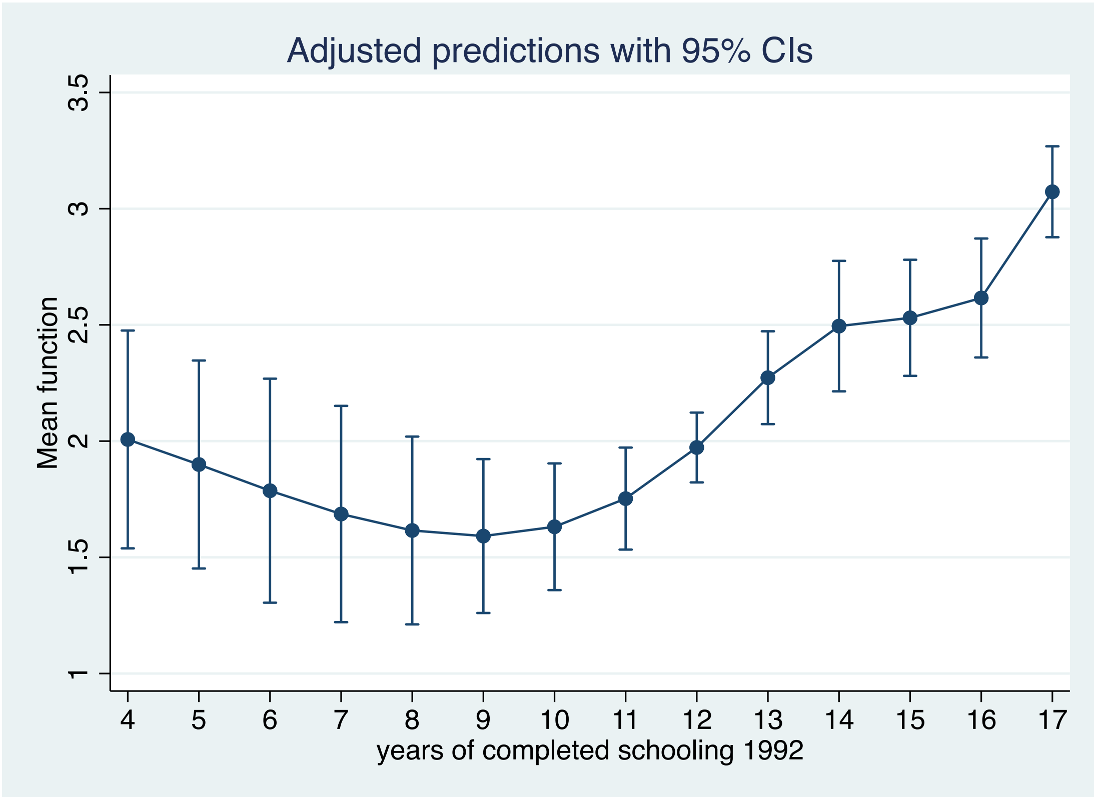
Note: unless specified otherwise, cubic spline and number of knots chosen by cross validation

In this example: cubic spline, 3 knots...how many parameters?

```
. npregress series lnhwage educatn, spline
warning: you have entered variable educatn as continuous but it only has 14 distinct values.  The e
        substantially if you inadvertently include a discrete variable as continuous

Computing approximating function

Minimizing cross-validation criterion

Iteration 0:  Cross-validation criterion =  .5835807
Iteration 1:  Cross-validation criterion =  .5803272

Computing average derivatives

Cubic-spline estimation                     Number of obs     =        177
Criterion: cross-validation                 Number of knots   =          3

------------------------------------------------------------------------------
             |               Robust
     lnhwage |     Effect   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
     educatn |   .2100777   .0385008     5.46   0.000     .1346175    .2855378
------------------------------------------------------------------------------
Note: Effect estimates are averages of derivatives.
```

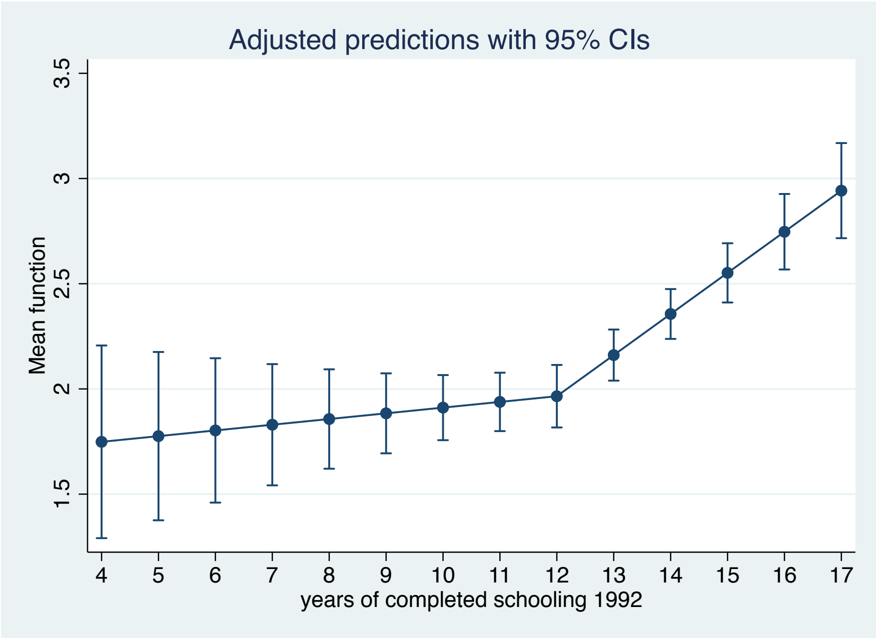
- Conditional expectation at different points

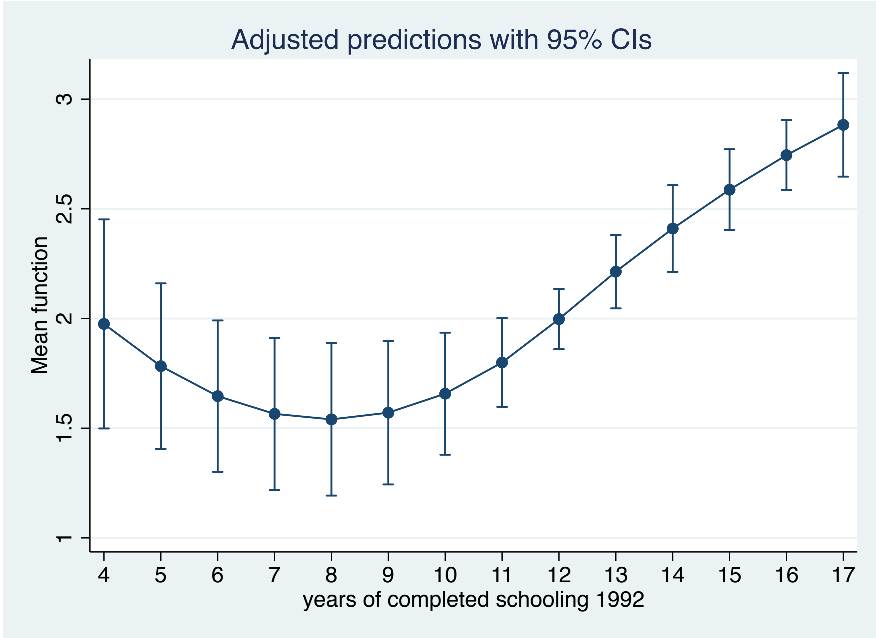
margins, at(educatn=(4 5 6 7 8 9 10 11 12 13 14 15 16 17))
marginsplot



- Restrict to linear or quadratic spline to see the difference

(in both cases, 1 knot selected by CV)





Asymptotic properties

- Consistent (if $K, N \to \infty$)

- Asymptotically normal

- Rate of convergence is $\sqrt{N}$

- But finite-samples biases still exist (the C.I. is not centered correctly centered), similar case as in Kernel regression!

- the bias term can be made asymptotically negligible if we assume that

$$Kincreaseswith$$

N$atasufficientlyfastrate.$

The Global/Local Nature of Series Regression

- Kernel regression as inherently local in nature.

- The Nadaraya-Watson, Local Linear, and Local Polynomial estimators estimate $m(x_0)$ only considering x's close to x_0.

- In contrast, series regression is typically described as global in nature: estimators are a function of the whole sample

- However, series regression estimators share the local smoothing property of kernel regression:

- As the number of series terms K increase a series estimator also becomes a local weighted average estimator.

- Thus, another interpretation

  - Both are global in nature when h is large (kernels) or K is small (series), and

  - . . . are local in nature when h is small (kernels) or K is large (series).

- See Hansen, Chapter 20, for additional details

Takeaway

- Two different ways of doing nonparametric regression
- Local averages
- This handout: Flexible functions of the regressors:
  - Splines
  - Polynomials
- Very easy to implement, quicker rate of convergence

Additional References

In case you are interested on this topic, you can check the following references:

- For a textbook treatment of series regression: see Li and Racine (2007).

- For an advanced treatment see Chen (2007).

- Two seminal contributions are Andrews (1991a) and Newey (1997).

- Recent contributions: Belloni, Chernozhukov, Chetverikov, and Kato (2015) and Chen and Christensen (2015).