Nonlinearities in the development process: A nonparametric approach

Théophile T. Azomahou
UNU-MERIT, Maastricht University

Globelics Academy 2008
Tampere, Finland
Parametric Vs. nonparametric models

Accounting for nonlinearities in parametric and nonparametric regressions

Parametric linear regression model

\[ y = \alpha + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \varepsilon \]

- Aim: estimate the parameters \( \alpha \) and \( \beta \)s.

- Advantage: basic econometrics, easy to perform, friendly user software available (methods: least squares, maximum likelihood, IV, GMM, ...)

- Drawback: misspecification problem (DGP is assumed to be known).
Parametric Vs. nonparametric models

Accounting for nonlinearities in parametric and nonparametric regressions

Nonparametric (or semiparametric) regression

\[ y = f(x) + \varepsilon \]

- **Aim**: estimate the function \( f(x) \), without assuming a particular form for \( f(x) \).
- **Advantage**: robust to misspecification, DGP unknown (methods: kernel, cubic spline, k-nn, series, local linear, etc).
- **Drawback**: possibly data consuming, lack of friendly user procedures.
Examples (cont’d)

1. The case of CO$_2$ emissions (EKC)

- Parametric estimation: inverted U-shaped relationship

- Nonparametric estimation: monotonically increasing function
Examples

2. Life expectancy and income growth


- Nonparametric: convex-concave relationship between life expectancy at birth and income per capita.
Outline of the presentation

1. Set up
2. Kernel density estimation
3. Nonparametric regression
4. Semiparametric estimation
5. Application: technology frontier
Kernel density estimation

The core of the method: The kernel density estimation

Density as distribution

- The density of a variable describes the distribution of the values that the random variable takes.
  - Fully parametric distribution assumes about the form of the density.
  - Canonical example: if $X \sim N(\mu, \sigma^2)$, then

$$
\hat{f}(x) = f(x \mid \hat{\mu}, \hat{\sigma}^2) = \frac{1}{\hat{\sigma}} \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{x - \hat{\mu}}{\hat{\sigma}} \right)^2 \right].
$$

for some estimation of $\hat{\mu}$ and $\hat{\sigma}$ (mean and variance obtained from a given sample).

- Problem: narrow distributional assumption about the density.
Kernel density estimation

Histogram as a crude density estimator

Example of distribution of sales over 1,270 firms

- Descriptive statistics:

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Std.Dev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales</td>
<td>0.3428</td>
<td>0.08919</td>
<td>0.2361</td>
<td>0.5664</td>
</tr>
</tbody>
</table>

- Histogram:
The distribution seems to be **bimodal**, but no particular functional form seems natural.
Kernel density estimation

From histogram to kernel

Frequency

\[ \hat{f}(x) = \frac{1}{n} \frac{\text{frequency in bin}_x}{\text{width of bin}_x} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} \textbf{1}(x - \frac{h}{2} < x_i < x + \frac{h}{2}) \]

where:

- \( x_k \) is the midpoint of the \( k \)th bin and \( h \) is the width of the bin. The distance to the left and right boundaries of the bins are \( h/2 \).
- \( \textbf{1}(\text{statement}) \) denotes an indicator function.
- The frequency count in each bin is the number of observations in the sample which fall in the range \( x_k \pm h/2 \). Collecting terms gives the formula.
- \( \text{bin}_x \) denotes the bin which has \( x \) as its midpoint.
Kernel density estimation

From histogram to kernel

Rearrange the event in the indicator function to produce an equivalent form: the (naive) density kernel estimator.

\[
\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} 1 \left( -\frac{1}{2} < \frac{x_i - x}{h} < \frac{1}{2} \right)
\]

\[
= \hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h} K \left[ \frac{x_i - x}{h} \right]
\]

where:

- \( K[z] = 1[-1/2 < z < 1/2] \).

- This form of the estimator counts the number of points that are within \( \frac{1}{2} \) bin width of \( x_k \).
Kernel density estimation

i) From histogram to kernel: why is it naive?
   ▶ This estimator is neither smooth nor continuous (crudeness of $K[z]$).
   ▶ Its shape is partly determined by where the leftmost and rightmost terminals of the histogram are set.
   ▶ The shape of the histogram will be crucially dependent on the bandwidth, itself.

ii) How to overcome the crudeness of the weighting function $K[z]$?
   ▶ Rosenblatt (1956): substitute for the naive estimator some other weighting function which is continuous and integrates to one.

<table>
<thead>
<tr>
<th>Kernels</th>
<th>Formula $K[z]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epanechnikov</td>
<td>$0.75(1 - .2z^2)/2$, $2.26$ if $</td>
</tr>
<tr>
<td>Normal</td>
<td>$\phi(z)$ (normal density),</td>
</tr>
<tr>
<td>Uniform</td>
<td>$0.5$ if $</td>
</tr>
<tr>
<td>Logit</td>
<td>$\Lambda(z)[1 - \Lambda(z)]$ (logistic density)</td>
</tr>
<tr>
<td>Parzen</td>
<td>$4/3 - 8z^2 + 8</td>
</tr>
</tbody>
</table>
Kernel density estimation: Back to data (applications)

- Distribution of sales: Logistic kernel

- Distribution of GDP per cap: Epanechnikov kernel
Nonparametric regression

1. Regression function of a variable $y$ on a single variable $x$:

$$y = m(x) + \varepsilon$$

- No assumptions about distribution, homoscedasticity, serial correlation.
- The functional form is still the same for all values of $x$ but unknown.

2. Methods: smoothing techniques
   - The case of Nadaraya-Waston estimator

$$\hat{m}(x^* \mid x, h) = \frac{\sum_{i=1}^{n} \frac{1}{h} K \left[ \frac{x_i - x^*}{h} \right] y_i}{\sum_{i=1}^{n} \frac{1}{h} K \left[ \frac{x_i - x^*}{h} \right]} = \hat{f}(x),$$
Nonparametric regression

1. Easy to implement: A GAUSS procedure (only two lines!)

```gauss
proc (2) = npr(y,x);
    local reg, i, f;
    i=1;
    reg=zeros(n,1);
    f=zeros(n,1);
    do until i>n;
        f[i,1]=sumc(pdfn((x-x[i,1])/h))/(n*h);
        reg[i,1]=(sumc(pdfn((x-x[i,1])/h).*y)/(n*h))/f[i,1];
        i=i+1;
    endo;
    retp(f,reg);
endp;
```

2. Shortcomings
   ▶ Curse of dimensionality
   ▶ Slow speed of convergence
   ▶ Possibly data consuming
Semiparametric estimation: The partially linear regression

1. Consider the specification:

\[
y = f(x) + z'\beta + \varepsilon,
\]

Take a modified version of the previous:

\[
y - E(y|x) = [z - E(z|x)]\beta + [\varepsilon - E(\varepsilon|x)]
\]

2. Estimation procedure (Robinson, 1988):

- **Step 1**: Compute nonparametric estimators for \( E(y|x) \) and \( E(z|x) \) using the kernel method.
- **Step 2**: Compute an estimator for \( \beta \), \( \hat{\beta} \), by regressing \( y - E(y|x) \) on \( z - E(z|x) \). This step may be done by OLS.
- **Step 3**: Finally, obtain an estimator of \( f(x) \), \( \hat{f}(x) \), by a nonparametric regression \( E(\left(y - z'\hat{\beta}\right)|x) \).
1. Objective
   ▶ Study the growth strategy when countries are close to the technology frontier (US as reference).
   ▶ Estimation of two models: In the first model, the dependent variable is GDP growth rate per worker (as a measure of labor productivity growth), and in the second, the dependent variable is labor productivity backwardness (in logarithmic term).

2. Data
   ▶ 29 OECD countries data over the period 1960-2000.
   ▶ Sources: Penn World Table 6.1, World Development Indicators and Eurostat.
.tile3[Application: technology frontier, labor productivity and economic growth]

3. Descriptive statistics

<table>
<thead>
<tr>
<th>Variables</th>
<th>#Obs.</th>
<th>Mean</th>
<th>Std.Dev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor produc. growth rate</td>
<td>958</td>
<td>0.036</td>
<td>0.47</td>
<td>-0.01</td>
<td>1.02</td>
</tr>
<tr>
<td>Labor produc. backwardness</td>
<td>985</td>
<td>-0.88</td>
<td>1.2</td>
<td>-6.82</td>
<td>3.38</td>
</tr>
<tr>
<td>Primary school enrol. rate</td>
<td>603</td>
<td>0.11</td>
<td>0.01</td>
<td>0.08</td>
<td>0.14</td>
</tr>
<tr>
<td>Secondary school enrol. rate</td>
<td>615</td>
<td>0.85</td>
<td>0.23</td>
<td>0.11</td>
<td>1.48</td>
</tr>
<tr>
<td>School enrol. rate in higher educ.</td>
<td>280</td>
<td>0.36</td>
<td>0.19</td>
<td>0.04</td>
<td>0.98</td>
</tr>
<tr>
<td>Government R&amp;D expenditure</td>
<td>196</td>
<td>0.4</td>
<td>0.14</td>
<td>0</td>
<td>0.7</td>
</tr>
<tr>
<td>Industry R&amp;D expenditure</td>
<td>186</td>
<td>0.5</td>
<td>0.15</td>
<td>0</td>
<td>0.91</td>
</tr>
<tr>
<td>R&amp;D expenditure from abroad</td>
<td>182</td>
<td>0.61</td>
<td>0.52</td>
<td>0</td>
<td>3.03</td>
</tr>
</tbody>
</table>
Application: technology frontier, labor productivity and economic growth

4. Distribution of variables of interest (kernel density estimation)
Application: technology frontier, labor productivity and growth

Specification: The Generalized Additive Model (GAM) for panel data

\[ Y = \alpha + \sum_{j=1}^{p} f_j(X_j) + Z'\gamma + \epsilon \]

where:

- \( Y = (y_{i1}, \cdots, y_{iT})' \) denotes the response variable
- \( X_j = (x_{i1}, \cdots, x_{iT})' \) for \( j = 1, \cdots, p \) are non linear explanatory variables, \( i = 1, \cdots, n \) and \( t = 1, \cdots, T \)
- \( Z \) is the row vector of parametric components
- \( \alpha \) denotes the regression intercept, and \( \gamma \) the vector of parameters
- The \( f_j \) are unknown univariate functions to be estimated such that
  \[ \mathbb{E}[f_j(X_j)] = 0. \]
- Error: \( \epsilon = (\epsilon_{i1}, \cdots, \epsilon_{iT})' \) is such that
  \[ \mathbb{E}(\epsilon | X_1, \ldots, X_p, Z) = 0 \] and
  \[ \mathbb{V}(\epsilon | X_1, \ldots, X_p, Z) = \sigma^2(X_j, Z) \]
Application: technology frontier, labor productivity and economic growth

5. Estimation results (cont’d)

- Labor productivity growth and school enrollment rate in higher education

- Labor productivity backwardness and school enrollment rate in higher education
5. Estimation results

Labor productivity backwardness and the part of R&D expenditure in % of GERD funded by industries

Table: Semiparametric estimation for labor productivity backwardness

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coef.</th>
<th>Std.Err</th>
<th>df.</th>
<th>Gain&lt;sup&gt;(a)&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary school enrollment rate</td>
<td>-0.08*</td>
<td>0.04</td>
<td>2</td>
<td>0.64</td>
</tr>
<tr>
<td>Secondary school enrollment rate</td>
<td>-0.02</td>
<td>0.013</td>
<td>2</td>
<td>1.8</td>
</tr>
<tr>
<td>School enrollment rate in higher education</td>
<td>-0.02</td>
<td>0.02</td>
<td>4.99</td>
<td>20.29&lt;sup&gt;(b)&lt;/sup&gt;</td>
</tr>
<tr>
<td>Government R&amp;D expenditure</td>
<td>0.03**</td>
<td>0.01</td>
<td>3.99</td>
<td>1.24</td>
</tr>
<tr>
<td>Industry R&amp;D expenditure</td>
<td>0.02</td>
<td>0.011</td>
<td>5.99</td>
<td>20.39&lt;sup&gt;(b)&lt;/sup&gt;</td>
</tr>
<tr>
<td>R&amp;D expenditure from abroad</td>
<td>-0.04</td>
<td>0.03</td>
<td>2</td>
<td>0.17</td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.18</td>
<td>0.16</td>
<td>1</td>
<td>-</td>
</tr>
</tbody>
</table>
Regional convergence

The dynamics of transition of regional GDP in Europe

Conditionnal density of GDP per capita

Polarization of GDP per capita