

# Nonparametric Methods in Econometrics using



David T. Jacho-Chávez<sup>1</sup>

<sup>1</sup>Department of Economics  
London School of Economics

Cemmap 2006

## Packages:

- ▶ `Ecdat` – Data sets for econometrics
- ▶ `KernSmooth` – Functions for kernel smoothing for Wand & Jones (1995)

## Data: Earnings

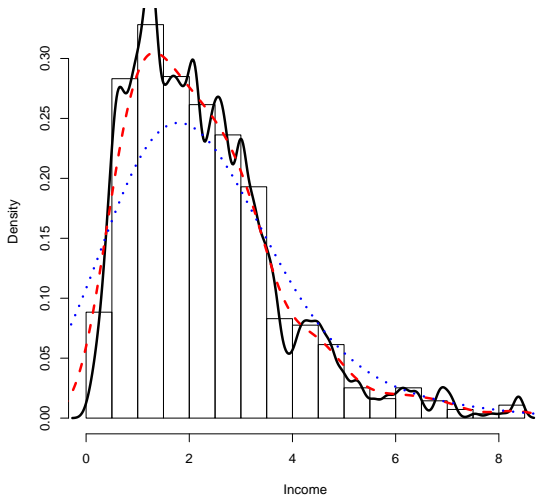
Description: 4266 observations from a cross-section (1988–1989) from USA

- ▶ `age` (age groups) – A factor with levels (g1,g2,g3)
- ▶ `y` (average annual earnings) – In 1982 US dollars

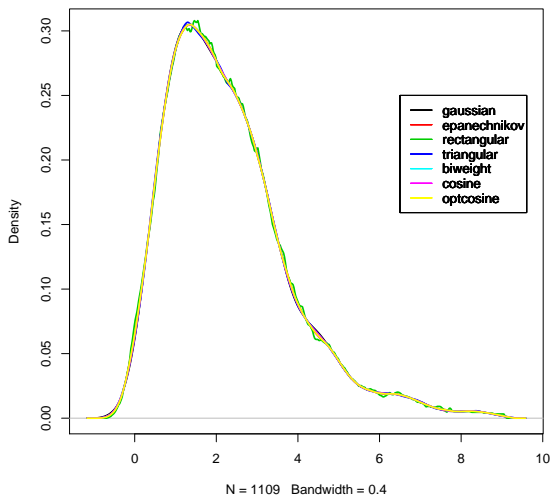
Source: Mills, Jeffery A. and Sourushe Zandvakili (1997) “Statistical Inference via Bootstrapping for Measures of Inequality”, *Journal of Applied Econometrics*, 12(2), pp. 133-150

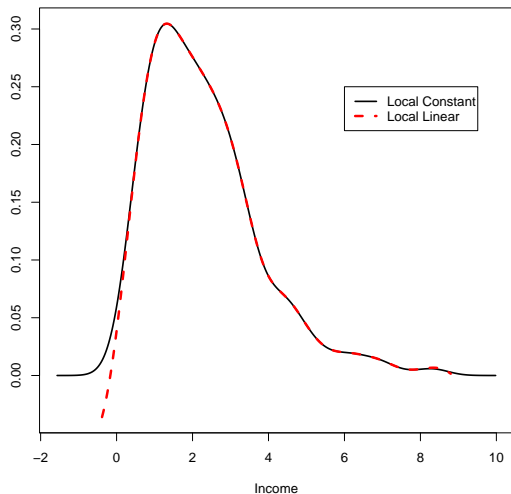
```
package: stats
density(x, bw = "nrd0", adjust = 1,
        kernel = c("gaussian", "epanechnikov",
                   "rectangular",
                   "triangular", "biweight",
                   "cosine", "optcosine"),
        weights = NULL, window = kernel, width,
        give.Rkern = FALSE,
        n = 512, from, to, cut = 3,
        na.rm = FALSE, ...)
```

### Sensitivity: Bandwidth

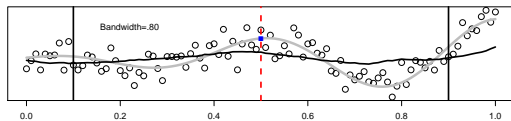
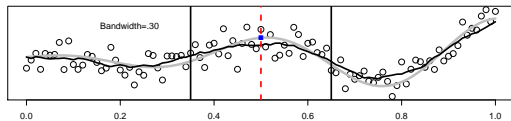
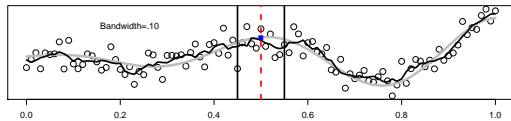


Sensitivity: 7 different kernels, same bandwidth

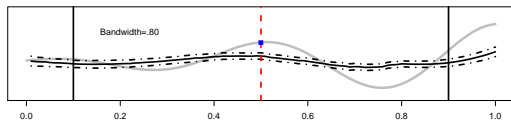
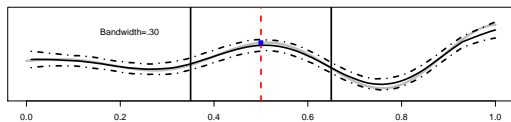
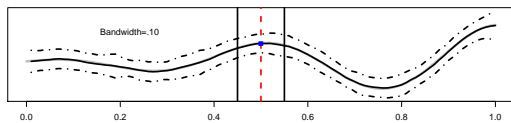




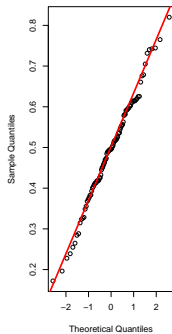
```
package: stats  
ksmooth(x, y,  
        kernel = c("box", "normal"),  
        bandwidth = 0.5,  
        range.x = range(x),  
        n.points = max(100, length(x)), x.points)
```



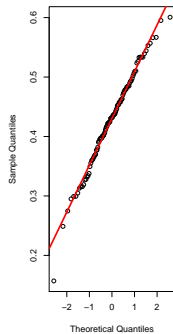




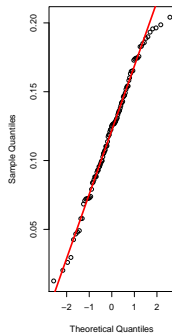
Normal Q-Q Plot:  $m(0.5)$   
 $h=0.1$



Normal Q-Q Plot:  $m(0.5)$   
 $h=0.3$



Normal Q-Q Plot:  $m(0.5)$   
 $h=0.8$



## Packages:

- ▶ `Ecdat` – Data sets for econometrics
- ▶ `locfit` – Local Regression, Likelihood and Density Estimation

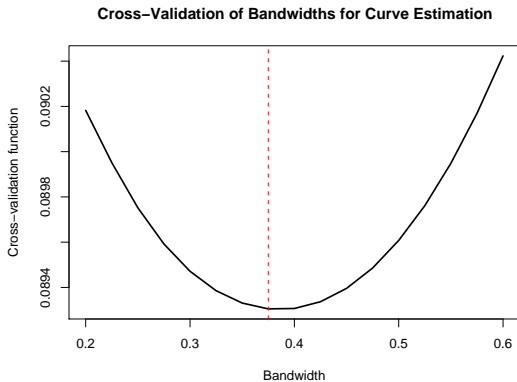
## Data: Housing

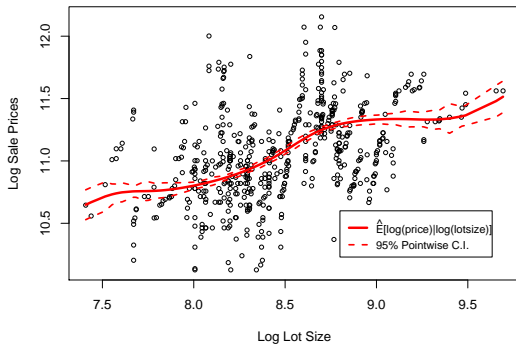
Description: 546 observations from a cross-section (1987) regarding sales prices of houses in the city of Windsor (Canada)

- ▶ *price* – Sale price of a house
- ▶ *lotsize* – The lot size of a property in square feet

Source: Anglin, P.M. and R. Gencay (1996) “Semiparametric estimation of a hedonic price function”, *Journal of Applied Econometrics*, 11(6), 633-648

```
package: locfit
density.lf(x, n = 50, window = "gaussian",
           width, from, to,
           cut = if(iwindow == 4.) 0.75 else 0.5,
           ev = lfgrid(mg = n, ll = from, ur = to),
           deg = 0,
           family = "density", link = "ident", ...)
```





## Packages:

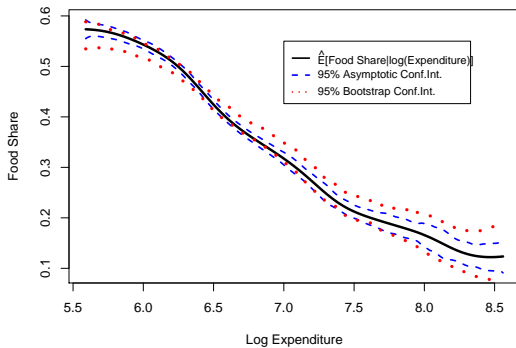
- ▶ `Ecdat` – Data sets for econometrics
- ▶ `locfit` – Local Regression, Likelihood and Density Estimation

Data: `Southafrica`

Description: Data taken from Living Standards Measurement Survey  
– <http://www.worldbank.org/lsms/>

- ▶ `ltexp` – Log(Total monthly household expenditure)
- ▶ `FoodShr` – Share of total expenditure on food

Source: Yatchew, Adonis (2003), *Semiparametric Regression for the Applied Econometrician*, Cambridge University Press, First Edn





## Packages:

- ▶ `JLLprod` – Nonparametric Estimation of Homothetic and Generalized Homothetic Production Functions
- ▶ `akima` – Interpolation of irregularly spaced data

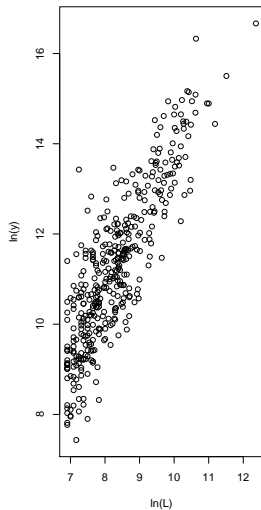
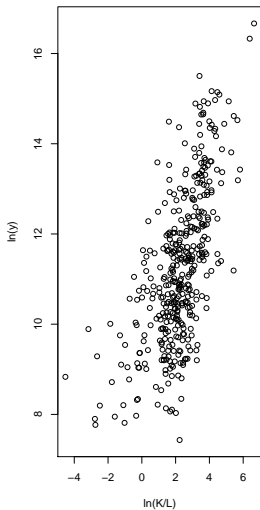
Data: `ecu`

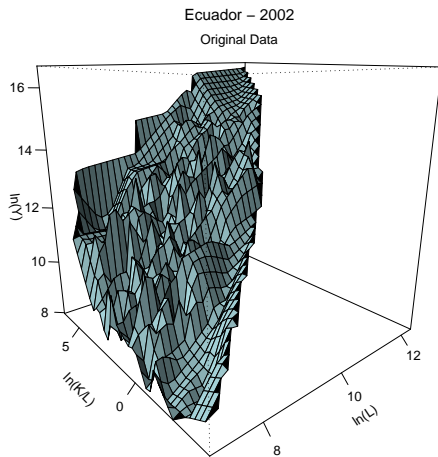
Description: 406 observations at plant level for the Petroleum, Chemical & Plastics industry in Ecuador for the year 2002

- ▶ `lny` ( $\text{Log}(y)$ ) – Output in thousands of current US dollars
- ▶ `lnk` ( $\text{Log}(K)$ ) – Capital in thousands of current US dollars
- ▶ `lnl` ( $\text{Log}(L)$ ) – The average number of employees

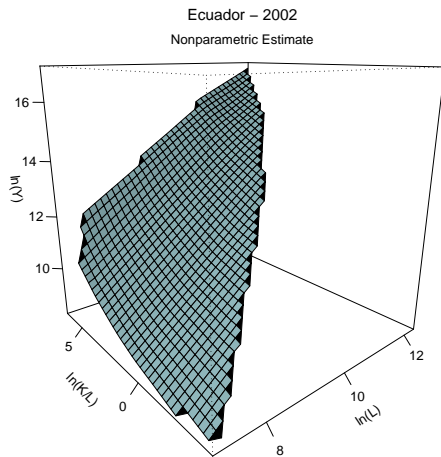
Source: Jacho-Chávez, David T., A. Lewbel and O. B. Linton (2005)  
“Identification and Nonparametric Estimation of a Transformed Additively Separable Model”, *Unpublished Manuscript*

Ecuador – 2002  
Petroleum, Chemical & Plastics, 406 Plants





```
>library(JLLprod,akima,locfit)
>data(ecu)
>fit      <- locfit(lny~lp(lnl,lnk-lnl,nn=0,h=4,
                        deg=1,scale=T), data=ecu)
>fitted <- fitted(fit)
>persp(interp(lnl,lnk-lnl,fitted),
        axes=TRUE,lty=1,lwd=1,,xlab="ln(L)",
        ylab="ln(K/L)", zlab="ln(Y)",
        ticktype="detailed", nticks=4,
        font.lab=1, font.main=1,
        col="powderblue", theta= 320,
        phi=17, shade=0.45,
        main="Ecuador - 2002")
mtext("Nonparametric Estimate")
```



## Packages:

- ▶ `Ecdat` – Data sets for econometrics
- ▶ `gam` – Generalized Additive Models

## Data: `Participation`

Description: 872 observations from a cross-section regarding labor force participation in Switzerland

- ▶ `lnnlinc` – The log of nonlabor income
- ▶ `age` – Age in years divided by 10
- ▶ `educ` – Years of formal education
- ▶ `nyc` – The number of young children (younger than 7)
- ▶ `noc` – Number of older children

Source: Gerfin, Michael (1996) “Parametric and semiparametric estimation of the binary response”, *Journal of Applied Econometrics*, 11(3), 321-340.

```
>library(Ecdat,gam)
>data(Participation)
>lab1.gam <- gam(lnnlinc ~ s(educ)+s(age)
                  +s(nyc)+s(noc),
                  data=Participation)
>par(mfrow=c(2,2),pty="s",lwd=3,las=1)
>plot.gam(lab1.gam,se=T,col="red")
```

