Advanced Public Economics — Applied Tutorial

Session III: Estimating Shape of Optimal Tax Rates

Daniel Weishaar

Roadmap for Session III

Estimate the **shape of optimal tax rates** over the income distribution.

① Estimate income distributions in a non-parametric way (kernel density).

② Estimate shape of optimal tax rates based on structural model.

Brief Recap of the Formula

Optimal Tax Rates — Based on Primitives

Following Diamond (1998), under the assumption of quasi-linear utility function with iso-elastic effort costs, i.e.,

$$u(c(\omega), y(\omega), \omega) = c(\omega) - k(y(\omega), \omega), \qquad k(y(\omega), \omega) = \left(1 + \frac{1}{\varepsilon}\right)^{-1} \left(\frac{y(\omega)}{\omega}\right)^{1+\frac{1}{\varepsilon}},$$

the **optimal non-linear tax rate** satisfies the condition

$$\frac{T'(y(\omega))}{1 - T'(y(\omega))} = (1 - G(\omega))\frac{1 - F(\omega)}{f(\omega)\omega}\frac{1}{\varepsilon}$$

Types or productive abilities generally hard to observe, but we can rewrite the condition in terms of observable statistics, in particular the income distribution (Saez, 2001).

Optimal Tax Rates — Based on Observables

Since $F_y(y') = F(\omega_0(y'))$ and $f_y(y') = f(\omega_0(y'))\omega_0'(y')$, the formula reads

$$\frac{T'(y)}{1 - T'(y)} = (1 - \mathcal{G}(y)) \frac{1 - F_y(y)}{f_y(y)} \frac{\omega_0'(y)}{\omega_0(y)} \frac{1}{\varepsilon}$$

For now, further assume that exogenous welfare weights are Rawlsian, which yields the revenue-maximizing non-linear tax rate

$$\frac{T'(y)}{1 - T'(y)} = \frac{1 - F_y(y)}{f_y(y)} \frac{\omega_0'(y)}{\omega_0(y)} \frac{1}{\varepsilon}$$

$\Rightarrow$ **Question:** How do we bring this formula to the data?

Optimal Tax Rates — (Inverse) Hazard Rate

$$\frac{T'(y)}{1 - T'(y)} = \frac{1 - F_y(y)}{f_y(y)} \frac{\omega_0'(y)}{\omega_0(y)} \frac{1}{\varepsilon}$$

Hazard rate requires estimation of cumulative distribution function (CDF) and probability density function (PDF). Different methods for estimation:

- ▶ **Parametric:** assume a functional form for the distribution (log-normal, Pareto,...) and select parameters to fit the data, e.g., through Maximum-Likelihood.
  - ▶ We estimated the Pareto parameter for the top of the distribution last time.
  - ▶ But! Empirically observed distribution might feature irregular patterns that cannot be captured by one functional form.

- ▶ **Non-Parametric:** Start with the histogram of incomes and estimate the density using a **kernel density**. Intuitively, the kernel density estimator smooths the information from histogram of incomes.

Optimal Tax Rates — Kernel Density Estimator

We follow the explanations provided in a data science blogpost, see also visualization.

Goal: Estimate an unknown PDF $f(x)$ without assuming a parametric form.

Key idea: Build the density from identical "building blocks" (kernels), one placed at each observation. Each block is a smooth PDF centered at a data point.

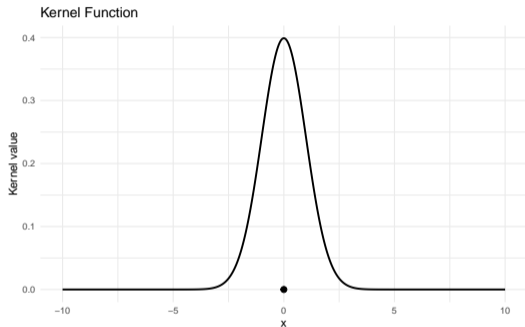# Optimal Tax Rates — Kernel Density Estimator, Toy Example

Start with a single kernel

Kernel is a valid probability density with unit area:

$$K(x) \geq 0, \qquad \int_{-\infty}^{\infty} K(x)\,dx = 1.$$
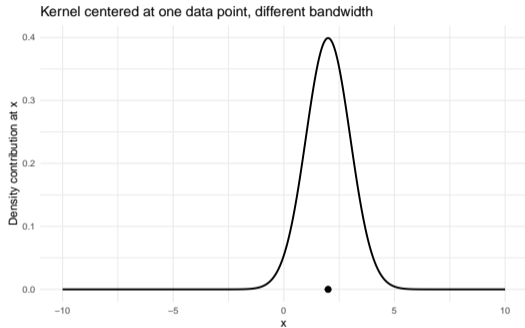
Example: Gaussian kernel

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Kernel Function

Optimal Tax Rates — Kernel Density Estimator, Toy Example

Shift the kernel to a data point x_i:

$$K(x - x_i).$$

Kernel centered at one data point, different bandwidth

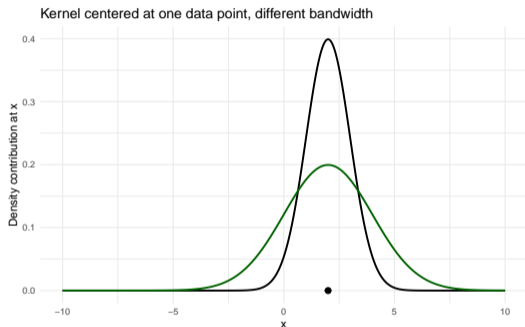# Optimal Tax Rates — Kernel Density Estimator, Toy Example

Control smoothness by scaling with the bandwidth h:

$$K\left(\frac{x - x_i}{h}\right).$$

Scaling changes the area, so we renormalize:

$$\frac{1}{h}K\left(\frac{x - x_i}{h}\right).$$

▶ large h: wide, smooth density;
▶ small h: narrow, spiky density.

Kernel centered at one data point, different bandwidth

Optimal Tax Rates — Kernel Density Estimator, Toy Example

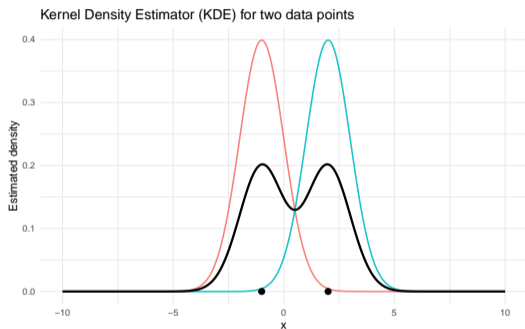**Multiple data points** For each data point x_i create a kernel contribution

$$f_i(x) = \frac{1}{h} K\left(\frac{x - x_i}{h}\right).$$

Add all contributions:

$$\tilde{f}(x) = \sum_{i=1}^{n} f_i(x).$$

This has total area n. Normalize by dividing by n to obtain a proper PDF:

$$\widehat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right).$$

Kernel Density Estimator (KDE) for two data points

Optimal Tax Rates — Kernel Density Estimator, Toy Example

Multiple data points For each data point x_i create a kernel contribution
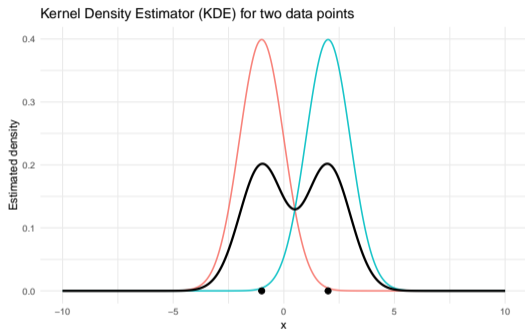
$$f_i(x) = \frac{1}{h} K\left(\frac{x - x_i}{h}\right).$$

Add all contributions:

$$\tilde{f}(x) = \sum_{i=1}^{n} f_i(x).$$

This has total area n. Normalize by dividing by n to obtain a proper PDF:

$$\widehat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right).$$

Kernel Density Estimator (KDE) for two data points

$\Rightarrow$ Hands-on exercise:

Optimal Tax Rates — Kernel Density Estimator — Specifications

The **Kernel Density Estimator (KDE)** is defined via

- ▶ **Kernel** — shape with which the pdf around each observation is approximated. Next to Gaussian, there are also other kernels (uniform, triangular, biweight, triweight, Epanechnikov). $\Rightarrow$ Choice not super important.
- ▶ **Bandwidth** — degree of smoothing $\Rightarrow$ Very important! Different methods selecting bandwidth relate to bias-variance trade-off.
    - ▶ **Small** $h \Rightarrow$ undersmoothing, high variance, many small peaks.
    - ▶ **Large** $h \Rightarrow$ oversmoothing, high bias, important features disappear.

Bandwidth Selection: Choose h to balance bias (too large h) and variance (too small h), i.e., we minimize mean integrated squared error (MISE) $\rightarrow$ optimal bandwidth h^*.

$$\text{MISE}(h) = \mathbb{E}\left[\int \left(\hat{f}_h(x) - f(x)\right)^2 dx\right].$$

Optimal Tax Rates — Kernel Density Estimator — Bandwidth Selection

1. Rule-of-Thumb (Silverman)
- ▶ Assumes data come from a Gaussian distribution.
- ▶ Optimal bandwidth is then function of observed characteristics of this distribution.

2. Cross-Validation Methods
- ▶ Choose h by optimizing out-of-sample fit.
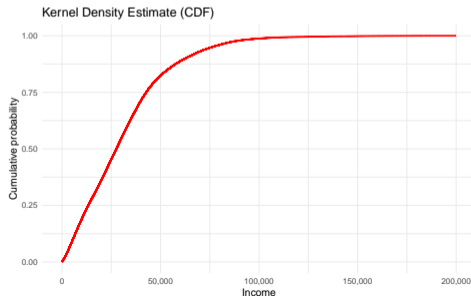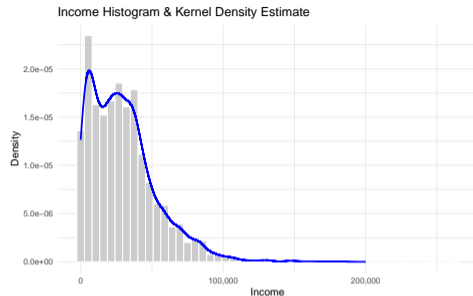
3. Adaptive (Variable) Bandwidths
- ▶ Allow h_i to vary: small h_i in dense regions, large in sparse regions.
- ▶ Useful for highly skewed or heavy-tailed data (e.g. income distributions).

Implementation in R via `density` command (1+2). Adaptive bandwidth selection (3) via packages `lokern` or `kedd`.
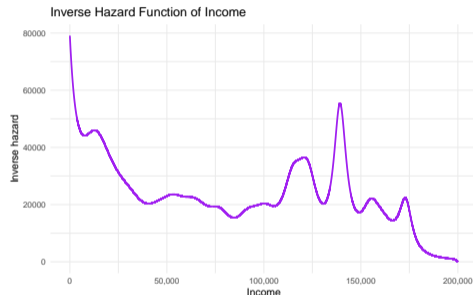
$\Rightarrow$ Hands-on exercise

Optimal Tax Rates — Hazard Rate

▶ In the hands-on exercise, we estimated the pdf and cdf for the income distribution based on the SOEP training data.

Income Histogram & Kernel Density Estimate

Kernel Density Estimate (CDF)

Optimal Tax Rates — Hazard Rate

- In the hands-on exercise, we estimated the pdf and cdf for the income distribution based on the SOEP training data.
- We can estimate the (inverse) hazard rate term in the optimal tax rate formula.

$$\frac{T'(y)}{1 - T'(y)} = \frac{1 - F_y(y)}{f_y(y)} \frac{\omega_0'(y)}{\omega_0(y)} \frac{1}{\varepsilon}$$

Gets noisy at top incomes! $\Rightarrow$ Append Pareto tail to the distribution.

Inverse Hazard Function of Income

That's it! See you next time!

References

References I

📄 DIAMOND, P. A. (1998). "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *American Economic Review* 88 (1), pp. 83–95.

📄 SAEZ, E. (2001). "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies* 68 (1), pp. 205–229.