On Fractile Transformation of Covariates in Regression[1]

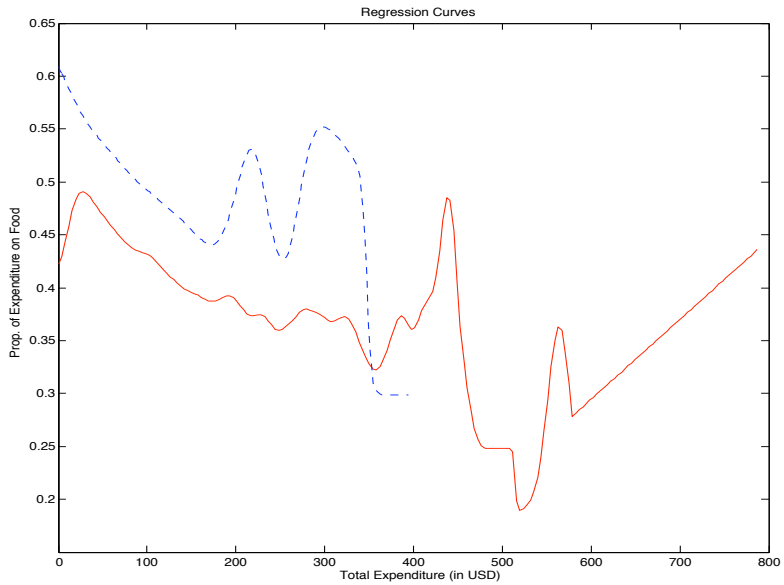
Bodhisattva Sen
Department of Statistics
Columbia University, New York

ERCIM'10
11 December, 2010

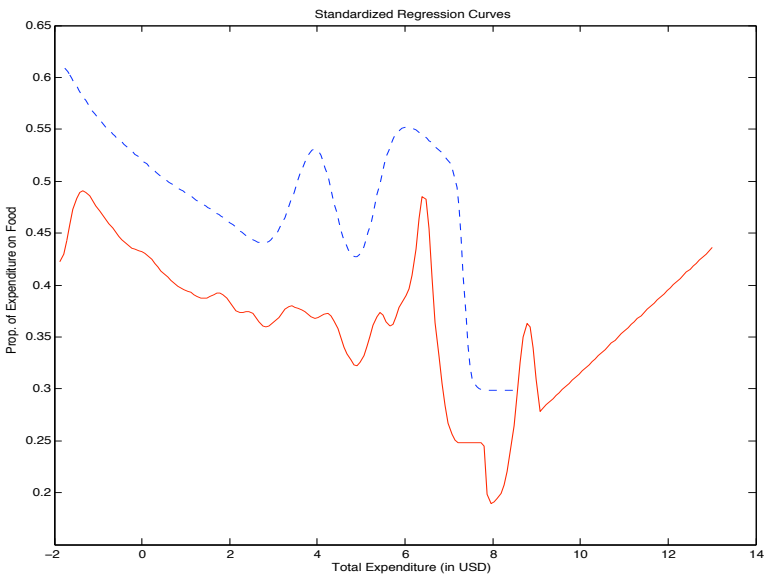
[1] Joint work with Probal Chaudhuri, Indian Statistical Institute, Calcutta

Example 1

- Household Expenditure and Income Data
- Investigate the *inequality in income* and *compare* the *economic condition* of Poland (blue) and Bulgaria (red)
- X = total expenditure; Y = proportion of expenditure on food as a fraction of X per capita per household



Usual regression functions

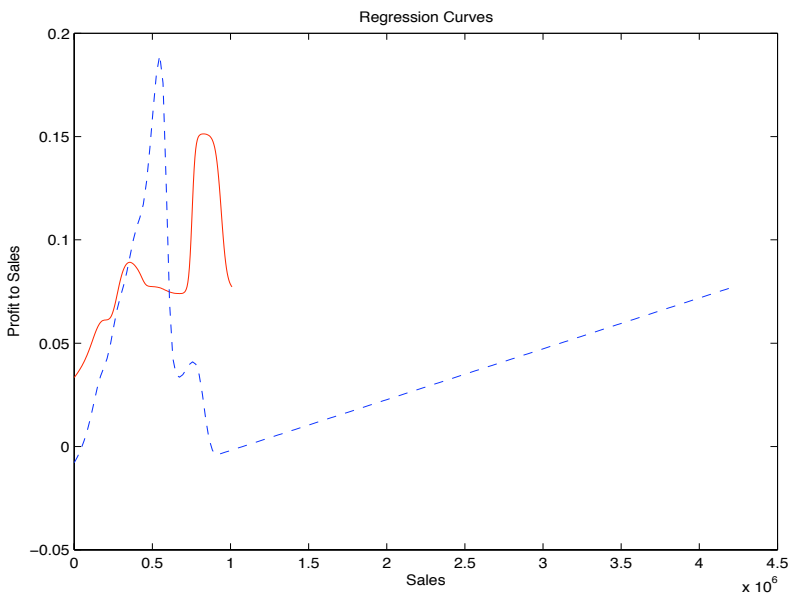


Standardized reg. functions

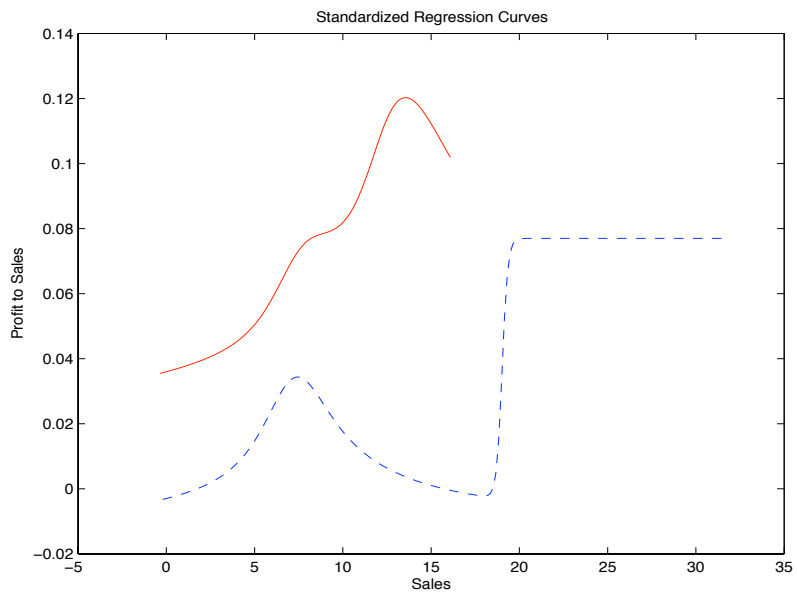
Example 2

- Data on the sales (in Indian rupees) and profit (as a fraction of sales) for companies over different years

- Compare the *Y = profitability* of the companies against *X = sales* for years 1997 (red) and 2003 (blue)



Usual regression functions



Standardized reg. functions

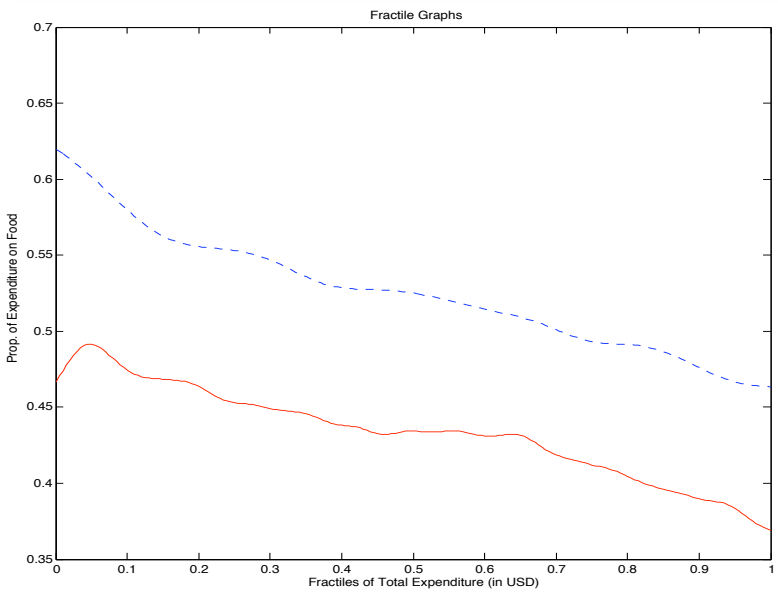
Problem: Comparison of two regression functions

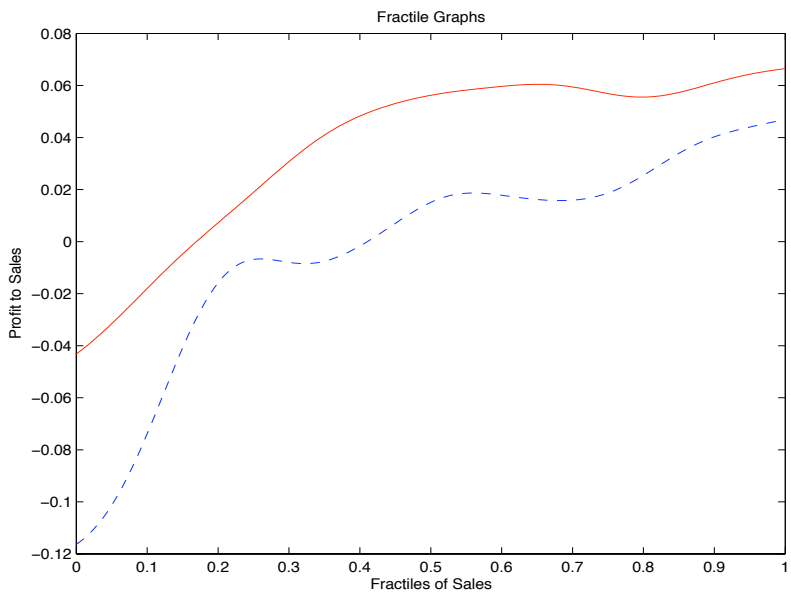
- Two bivariate populations (X_1, Y_1) and (X_2, Y_2)
- We usually look at $\mu_i(x) = E(Y_i|X_i = x)$, $i = 1, 2$

- Instead, compare the *fractile regression* functions

$$m_i(t) = E\{Y_i|F_i(X_i) = t\},\ t \in (0, 1)$$

where F_i is the c.d.f. of X_i





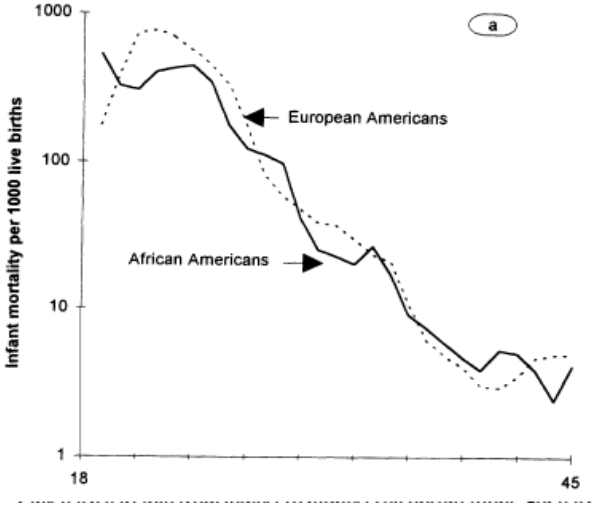
Fractile regression functions in Examples 1 and 2

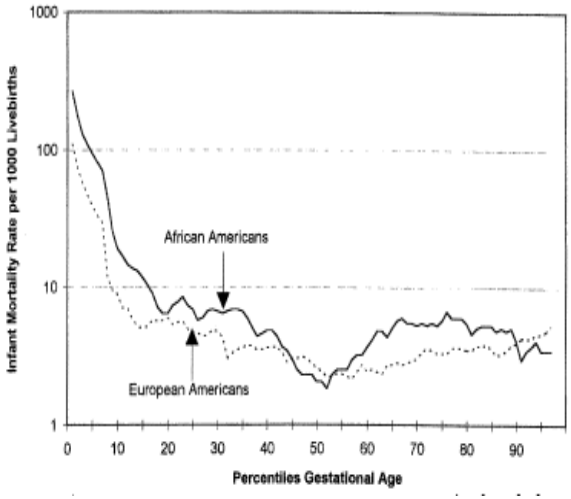
Other applications of fractile regression

- Hertz-Picciotto and Din-Dzietham (*Epidemiology*, 1998) compare the infant mortality of African and European Americans with gestational age

- Nordhaus (*PNAS*, 2006) compares the dependence of log of "output density" with key geographic variables

- Fractile regression enables us to *simultaneously* compare the effect of different covariates on one response variable





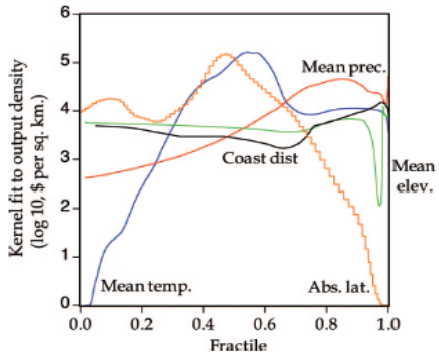


Fig. 2. Fractile plot for key geographic variables. The figure shows the fractile plots for key variables (mean temperature, mean precipitation, mean distance from coast, mean elevation, and absolute value of latitude). Fractiles rank each variable from lowest to highest cell observations. For each variable, we have fitted a kernel density function to the bivariate relationship between the $\log_{10}$ (output density) and the geographic variable. Zero values of output are included as equal 0 ($n = 17{,}796$).

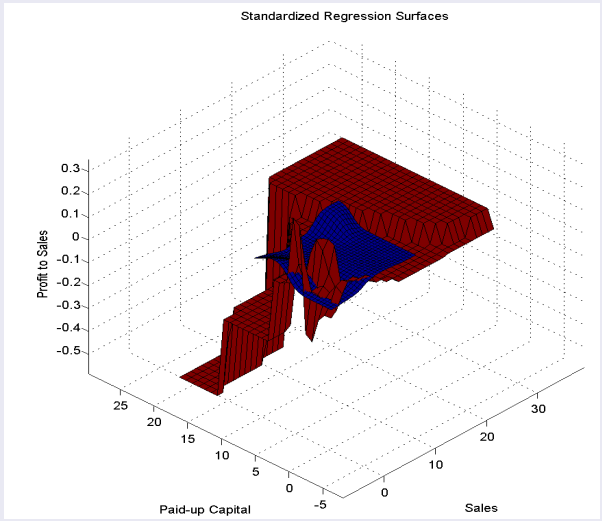
Why the fractile transformation $X_1 \mapsto F_1(X_1)$?

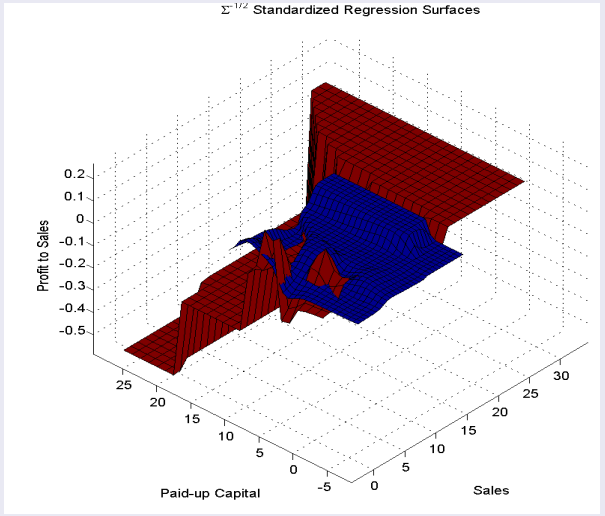
- Transformed covariates $F_1(X_1)$ and $F_2(X_2)$ both have a *Unif*$(0,1)$ distribution; thereby adjusting for covariate skewness/*data sparsity*
- *Distribution-free* nonparametric standardization
- Compare $m_1(t)$ and $m_2(t)$, the means of Y_1 and Y_2 at the t-th *quantile* of the covariates
- Makes the fractile regression functions *invariant* under all strictly increasing transformations of the covariate, e.g., if $X_2 = \phi(X_1)$, $Y_1 = Y_2$, then

$$E\{Y_1|F_1(X_1)\} = E\{Y_2|F_2(X_2)\}$$

- Mahalanobis (*Econometrica*, 1960), Sen and Chaudhuri (*JASA*, 2010), ...

Extension to multi-dimension





Questions:

- How do we standardize the *distribution* of the covariates that will enable a more meaningful comparison of the regression functions?

- Suppose $(\mathbf{X}_1, Y)$ and $(\mathbf{X}_2, Y)$ in $\mathbb{R}^{d+1}$ for $d \geq 1$, $\mathbf{X}_2 = g(\mathbf{X}_1)$ and $g : \mathbb{R}^d \mapsto \mathbb{R}^d$ is an (unknown) invertible function. How to standardize the covariates and conclude that the two regression functions are *essentially the same*?

Notation

- $(\mathbf{X}, Y)$ is a random vector having a continuous distribution on $\mathbb{R}^{d+1}$, $d \geq 1$, where $\mathbf{X} = (X_1, X_2, \ldots, X_d) \in \mathbb{R}^d$

- *Standardization* of the covariate: $\mathbf{T} : \mathbb{P} \times \mathbb{R}^d \to E \subset \mathbb{R}^d$ such that $\mathbf{x} \mapsto \mathbf{T}(\mathbf{P}, \mathbf{x}) \equiv \mathbf{T}(\mathbf{X}, \mathbf{x})$ is an *invertible* map from $\mathcal{X}_{\mathbf{P}}$, the support of $\mathbf{P}$, onto E, for every $\mathbf{X} \sim \mathbf{P} \in \mathbb{P}$, a class of distributions on $\mathbb{R}^d$.

- $\mathbf{T}_{ls}(\mathbf{P}, \mathbf{x}) = \Gamma(\mathbf{P})^{-1/2}\{\mathbf{x} - \mu(\mathbf{P})\}$, $\Gamma(\mathbf{P}) = diag(\sigma_1^2, \ldots, \sigma_d^2)$

- The *standardized regression function* is then defined as

$$m_{\mathbf{X}}(\mathbf{t}) = E\{Y | \mathbf{T}(\mathbf{P}, \mathbf{X}) = \mathbf{t}\} \quad \text{for } \mathbf{t} \in E.$$

- $\mathcal{G}$: group of one-one transformations acting on the space of all predictors $\mathbf{X} \in \mathbb{P}$. We say that $\mathbf{T}$ is *invariant* under $\mathcal{G}$ if $\mathbf{T}(\mathbf{g}(\mathbf{X}), \mathbf{g}(\mathbf{x})) = \mathbf{T}(\mathbf{X}, \mathbf{x})$, for all $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{g} \in \mathcal{G}$.

Fractile Standardization

- For $\mathbf{X} \sim \mathbf{P}$, define $\mathbf{R_P} : \mathbb{R}^d \mapsto (0,1)^d$, as

  $$\mathbf{R_P}(\mathbf{x}) = \left(F_1(x_1), F_{2|1}(x_2|x_1), \ldots, F_{d|1,\ldots,d-1}(x_d|x_1, \ldots, x_{d-1})\right),$$

  where $F_1(x_1) = P(X_1 \leq x_1), F_{2|1}(x_2) = P(X_2 \leq x_2|X_1 = x_1), \ldots$

- *Fractile regression*: $m_{\mathbf{X}}(\mathbf{t}) = E\{Y|\mathbf{R_P}(\mathbf{X}) = \mathbf{t}\}, \mathbf{t} \in (0,1)^d$

- *Distributional standardization*: $\mathbf{R_P}(\mathbf{X}) \sim$ Uniform$(0,1)^d$

- Multivariate analogue of $X_1 \mapsto F_1(X_1)$

Invariance

- Consider the group $\mathcal{F}, \mathbf{x} \mapsto (g_1(\mathbf{x}_1), \ldots, g_d(\mathbf{x}_d))$, where $g_i : \mathbb{R}^i \to \mathbb{R}$, is a $\uparrow$ func. in x_i for every fixed $(x_1, \ldots, x_{i-1})$, and $(g_1, \ldots, g_i) : \mathbb{R}^i \to \mathbb{R}^i$ is *invertible* for every i
- *Invariance*: for $\mathbf{g} \in \mathcal{F}$, $\mathbf{R}_{\mathbf{X}}(\mathbf{x}) = \mathbf{R}_{\mathbf{g}(\mathbf{X})}(\mathbf{g}(\mathbf{x}))$ for all $\mathbf{x} \in \mathbb{R}^d$
- $\{$all coordinate-wise increasing transformations$\} \subset \mathcal{F}$
- If we want the standardized regression function to be invariant under the group action $\mathcal{F}$, then the standardization $\mathbf{T}(\mathbf{X}, \cdot)$ has to be a *function of* $\mathbf{R}_{\mathbf{P}}$
- Furthermore, if we assume that $\mathbf{T}(\mathbf{X}, \mathbf{X}) \sim \mathit{Unif}(0,1)^d$ and $\mathbf{T}(\mathbf{X}, \cdot) \in \mathcal{F}$ then $\mathbf{T}(\mathbf{X}, \mathbf{x}) = \mathbf{R}_{\mathbf{P}}(\mathbf{x})$ for all $\mathbf{x}$, for all $\mathbf{X} \sim \mathbf{P} \in \mathcal{P}$

Computation of $\mathbf{R_P}$

- $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ i.i.d. $\mathbf{P}$
- $\mathbf{R_P}$ requires estimation of conditional distribution functions
- may use a kernel estimate of the *multivariate density* of $\mathbf{X}_1$, and then use it to get the various conditional densities

$$f_{n;1,2,\ldots,d}(\mathbf{x}) = \frac{1}{n(h_{1,n}h_{2,n}\ldots h_{d,n})} \sum_{i=1}^{n} \mathbf{K}\left(\frac{\mathbf{x} - \mathbf{X_i}}{\mathbf{h_n}}\right)$$

$$f_{n;j|1,\ldots,j-1}(x_j|x_1,\ldots,x_{j-1}) = \frac{f_{n;1,\ldots,j}(x_1,\ldots,x_j)}{f_{n;1,\ldots,j-1}(x_1,\ldots,x_{j-1})}$$

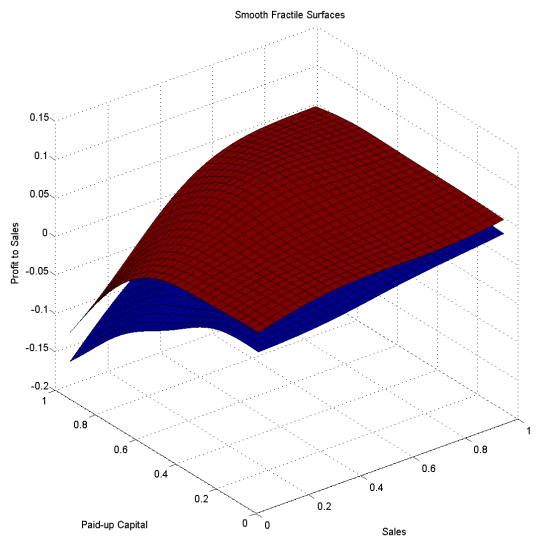
- *Standardized* covariates: $\mathbf{R}_n(\mathbf{X}_1), \mathbf{R}_n(\mathbf{X}_n), \ldots, \mathbf{R}_n(\mathbf{X}_n)$
- Under appropriate conditions, $\sup_{\mathbf{x}} \|\mathbf{R}_n(\mathbf{x}) - \mathbf{R_P}(\mathbf{x})\| \xrightarrow{P} 0$.
- Curse of *dimensionality*!

Computation of fractile regression

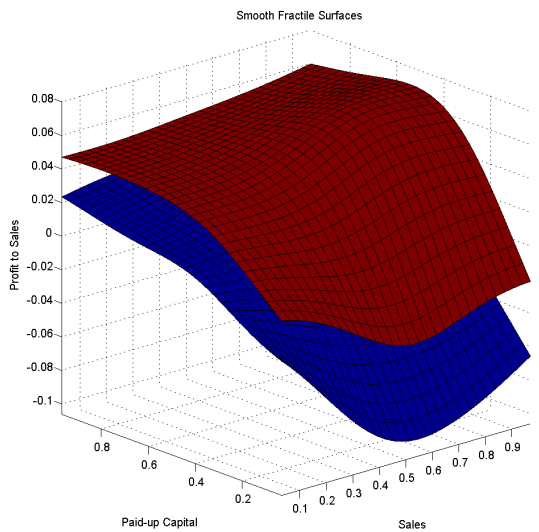
- Smooth estimate of *fractile regression*:

$$\widehat{m}_n(\mathbf{t}) = \sum_{i=1}^{n} Y_i W_{n,i}(\mathbf{t}), \quad \mathbf{t} \in (0,1)^d$$

- Nadaraya-Watson type weight: $W_{n,i}(\mathbf{t}) = \dfrac{\mathbf{K}\left(\frac{\mathbf{t}-\mathbf{R}_n(\mathbf{X}_i)}{\mathbf{h}_n}\right)}{\sum_{j=1}^{n} \mathbf{K}\left(\frac{\mathbf{t}-\mathbf{R}_n(\mathbf{X}_j)}{\mathbf{h}_n}\right)}$



Y on X_1 and X_2



Y on X_2 and X_1

Summary

- Usual comparison of regression functions not always possible and *meaningful*
- $\mathbf{R_P}$ acheives *distributional* standardization
- $\mathbf{R_P}$ has nice *invariance* properties, but computationally challanging for d large
- Alternatives: marginal standardization, centered *rank function* (multivariate distribution transform), etc.

Thank You!

Questions?