

# Nonparametric Methods

Robert A. Miller

Structural Econometrics

November 2025

# Empirical Distribution Function

## Definition

- Define the right continuous cumulative probability distribution function  $F_x : \mathbb{R}^K \rightarrow [0, 1]$  as:

$$F_x(x) \equiv F_x(x_1, \dots, x_K) \equiv \Pr \{X_1 \leq x_1, \dots, X_K \leq x_K\}$$

- Denote by  $x_n \equiv (x_{1n}, \dots, x_{Kn}) \in \mathbb{R}^K$  for  $n \in \{1, \dots, N\}$  a draw of observations from  $F(x)$ .
- Without making any parametric assumptions, one could estimate  $F(x)$  with the *empirical distribution function* defined as:

$$\begin{aligned} F_x^{(N)}(x) &\equiv F_x^{(N)}(x_1, \dots, x_K) \\ &\equiv \frac{1}{N} \sum_{n=1}^N \mathbf{1} \{x_{1n} \leq x_1, \dots, x_{Kn} \leq x_K\} \end{aligned}$$

- In words the estimator  $F^{(N)}(x^*)$  is the sample proportion of observations with a value less than or equal to  $x^*$ .
- Note the derivative of  $F^{(N)}(x^*)$  is zero wherever it exists, and  $F^{(N)}(x^*)$  itself is punctuated by jump points (at each observation).

# Empirical Distribution Function

## Empirical measures of conditional expectations

- The joint distribution of  $(y, x)$  can be estimated in the same way.
- We also define the empirical conditional distribution function as:

$$F_{y|x}^{(N)}(y|x) \equiv \frac{F_{y,x}^{(N)}(y,x)}{F_x^{(N)}(x)} = \frac{\sum_{n=1}^N \mathbf{1}\{y_n \leq y, x_{1n} \leq x_1, \dots, x_{Kn} \leq x_K\}}{\sum_{n=1}^N \mathbf{1}\{x_{1n} \leq x_1, \dots, x_{Kn} \leq x_K\}}$$

- Hence the estimated expected value of  $y$  conditional on  $x \leq x^*$  is:

$$E_{y|x}^{(N)}(y|x \leq x^*) = \frac{\sum_{n=1}^N y_n \mathbf{1}\{x_{1n} \leq x_1, \dots, x_{Kn} \leq x_K\}}{\sum_{n=1}^N \mathbf{1}\{x_{1n} \leq x_1, \dots, x_{Kn} \leq x_K\}}$$

- This result can be extended to obtain expressions for

$$E_{y|x}^{(N)}(y|x^* < x \leq x^{**}) \text{ and so forth.}$$

# Kernel Estimator

A uniform kernel for a univariate probability density function

- Suppose  $F_x(x)$  is differentiable with *pdf*  $f_x(x)$
- In the univariate case ( $K = 1$ ) approximate  $f_x(x)$  with:

$$\begin{aligned}f_x^{(N)}(x) &\equiv \left[ F_x^{(N)}(x + \phi^{(N)}) - F_x^{(N)}(x - \phi^{(N)}) \right] / 2\phi^{(N)} \\ &= \frac{1}{2N\phi^{(N)}} \sum_{n=1}^N \mathbf{1} \left\{ x - \phi^{(N)} < x_n \leq x + \phi^{(N)} \right\} \\ &= \frac{1}{N\phi^{(N)}} \sum_{n=1}^N k \left( \frac{x_n - x}{\phi^{(N)}} \right)\end{aligned}$$

where  $k(z) \equiv \frac{1}{2}$  if  $|z| \leq 1$  and  $k(z) \equiv 0$  if  $|z| > 1$ .

- Defined this way  $f_x^{(N)}(x)$  is called a uniform kernel estimator because its *kernel*,  $k(z)$ , is a uniform pdf.
- The rationale for the dependence of  $\phi^{(N)}$ , its *bandwidth*, on  $N$  is that  $\phi^{(N)} \rightarrow 0$  but  $N\phi^{(N)} \rightarrow \infty$  as  $N \rightarrow \infty$ .

# Kernel Estimator

Estimating the survivor function and hazard rate

- For example, denoting by  $p(t)$  the probability of stopping at  $t \in \{1, 2, \dots, T\}$ , where  $T \leq \infty$ , the survivor function and discrete hazard are, respectively:

$$1 - F(t) \text{ and } h(t) \equiv \frac{p(t)}{1 - \sum_{s=1}^{t-1} p(s)}$$

- In a sample of  $\{t_n\}_{n=1}^N$  they can be estimated by:

$$1 - F_N(t) \equiv \frac{1}{N} \sum_{n=1}^N \mathbf{1}\{t_n \geq t\} \text{ and } h_N(t) \equiv \frac{\sum_{n=1}^N \mathbf{1}\{t_n = t\}}{\sum_{n=1}^N \mathbf{1}\{t_n \geq t\}}$$

- The continuous time  $t \in (0, T)$  analogues of the survivor function take the same form.
- When  $t \in (0, T)$  the hazard rate and its nonparametric estimator are:

$$h(t) \equiv \frac{F'(t)}{1 - F(t)} \text{ and } h(t) = \frac{\frac{1}{\phi^{(N)}} \sum_{n=1}^N k\left(\frac{t_n - t}{\phi^{(N)}}\right)}{\sum_{n=1}^N \mathbf{1}\{t_n \geq t\}}$$

# Kernel Estimator

## Multivariate densities

- More generally we can define a kernel estimator  $f_x^{(N)}(x)$  for the multivariate density  $f_x(x)$  of  $x \in \mathbb{R}^K$  as:

$$f_x^{(N)}(x) \equiv \frac{1}{N \left(\phi^{(N)}\right)^K} \sum_{n=1}^N k_x \left( \frac{x_n - x}{\phi^{(N)}} \right)$$

where:

$$k_x(z_1, \dots, z_K) \equiv \prod_{k=1}^K k(z_k)$$

and  $k(z) : \mathbb{R} \rightarrow \mathbb{R}^+$  is a symmetric pdf with mean zero and finite variance:

$$k(z) \equiv k(-z) \quad \int k(z) dz = 1 \quad 0 < \int z^2 k(z) dz < \infty$$

# The Conditional Expectation Function

## Estimating the conditional expectation function

- We can form a kernel estimator for the nonparametric regression function of  $y$  on  $x$ :

$$E[y|x] \equiv \int y f_{y|x}(y|x) dy \equiv \int y \frac{f_{y,x}(y,x)}{f_x(x)} dy \quad (1)$$

by substituting in the kernel estimators for  $f(x)$ :

$$\hat{f}_x^{(N)}(x) = N^{-1} \left( \phi^{(N)} \right)^{-K} \sum_{n=1}^N k_x \left( \frac{x_n - x}{\phi^{(N)}} \right)$$

and  $f(y, x)$ :

$$\hat{f}_{y,x}^{(N)}(y, x) = N^{-1} \left( \phi^{(N)} \right)^{-(K+1)} \sum_{n=1}^N k_y \left( \frac{y - y_n}{\phi^{(N)}} \right) k_x \left( \frac{x_n - x}{\phi^{(N)}} \right)$$

into the right side of (1).

# The Conditional Expectation Function

Weighted averages to estimate conditional expectations

- As the next slide shows, this implies the estimator is the weighted average of  $\{y_n\}_{n=1}^N$ , given by:

$$E^{(N)} [y | x] \equiv \int y \frac{f_{y,x}^{(N)}(y, x)}{f_x^{(N)}(x)} dy = \sum_{n=1}^N w^{(N)}(x, x_n) y_n$$

where the weights  $w^{(N)}(x, x_n)$  are defined as:

$$w^{(N)}(x, x_n) \equiv k_x \left( \frac{x_n - x}{\phi^{(N)}} \right) / \sum_{n=1}^N k_x \left( \frac{x_n - x}{\phi^{(N)}} \right)$$

- Note that  $\phi^{(N)} \rightarrow 0$  as  $N \rightarrow \infty$  and the weight on observations distant from  $x$  disappear (since the variance of the kernel pdf is finite).
- If, as in most practical applications,  $k(z)$  is single peaked, then the weight on every observation declines with  $N$ .

# The Conditional Expectation Function

Where do the weights come from?

- To obtain the formula use the facts that:

$$\int zk(z) dz = 0 \text{ and } \int k(z) dz = 1$$

and the change in variables  $z = (y - y_n) / \phi^{(N)}$  to obtain:

$$\begin{aligned} \int y \frac{f_{y,x}^{(N)}(y, x)}{f_x^{(N)}(x)} dy &= \int y \frac{\sum_{n=1}^N k_y \left( \frac{y-y_n}{\phi^{(N)}} \right) k_x \left( \frac{x_n-x}{\phi^{(N)}} \right)}{\phi^{(N)} \sum_{n=1}^N k_x \left( \frac{x_n-x}{\phi^{(N)}} \right)} dy \\ &= \sum_{n=1}^N w^{(N)}(x, x_n) \int \frac{y}{\phi^{(N)}} k_y \left( \frac{y-y_n}{\phi^{(N)}} \right) dy \\ &= \sum_{n=1}^N w^{(N)}(x, x_n) \int (y_n + \phi^{(N)} z) k_y(z) dz \\ &= \sum_{n=1}^N w^{(N)}(x, x_n) y_n \end{aligned}$$

# Applying the Kernel Estimator

Heteroskedasticity with an unknown functional form

- One of the most common applications of semiparametric methods is for the linear model:

$$y_n = x_n' \beta_0 + \epsilon_n$$

where:

- $y$  is the dependent variable and  $x$  is a  $K \times 1$  explanatory variable vector
- $\beta_0$  is a  $K \times 1$  parameter vector to be estimated
- $\epsilon$  is an independently distributed disturbance with  $E[\epsilon_n | x_n] = 0$  and

$$E[\epsilon_n^2 | x_n] = \psi(x)$$

for some unknown function  $\psi(x)$ .

# Applying the Kernel Estimator

## Controlling for heteroskedasticity

- We proceed stepwise:

- ① Run least squares of  $y$  on  $x$  to obtain  $\beta_{OLS}^N$ .
- ② Form  $\widehat{\epsilon}_n^2 \equiv (y_n - x_n' \beta_{OLS}^N)^2$
- ③ Estimate  $\psi(x_n)$  with:

$$\psi^{(N)}(x_n) = \frac{\sum_{m=1}^N \widehat{\epsilon}_m^2 k_x\left(\frac{x_n - x_m}{\phi^{(N)}}\right)}{\sum_{m=1}^N k_x\left(\frac{x_n - x_m}{\phi^{(N)}}\right)}$$

- ④ Run least squares on the linear model:

$$\frac{y_n}{\sqrt{\psi^{(N)}(x_n)}} = \frac{x_n'}{\sqrt{\psi^{(N)}(x_n)}} \beta_0 + \frac{\epsilon_n}{\sqrt{\psi^{(N)}(x_n)}}$$

to obtain a heteroskedastically corrected estimate of  $\beta_0$ , which we denote by  $\beta_{HET}^N$ .

# Applying the Kernel Estimator

Comparing this estimator with the GLS analogue

- If  $h(x_n)$  was known, then we could obtain  $\beta_{GLS}^N$ , the GLS estimator, denoted by  $\beta_0$  by running the OLS regression:

$$\frac{y_n}{\sqrt{\psi(x_n)}} = \frac{x_n'}{\sqrt{\psi(x_n)}} \beta_0 + \frac{\epsilon_n}{\sqrt{\psi(x_n)}}$$

- We previously  $\beta_{GLS}^N$  is unbiased, but  $\beta_{HET}^N$  is biased, because:

$$\begin{aligned} & E \left[ \left[ \psi^{(N)}(x_n) \right]^{-1/2} \epsilon_n \left| \left[ \psi^{(N)}(x_n) \right]^{-1/2} x_n' \right. \right] \\ & \neq E \left[ \left[ \psi(x_n) \right]^{-1/2} \epsilon_n \left| \left[ \psi(x_n) \right]^{-1/2} x_n' \right. \right] \\ & = 0 \end{aligned}$$

- The assumptions above imply  $\beta_{OLS}^N$  is unbiased, begging the question whether  $\beta_{HET}^N$  or  $\beta_{OLS}^N$  is preferred.

# Implementation Details

## Choice of nonparametric method

- As a practical matter there are essentially three types of choices researchers make when implementing a nonparametric estimator:
  - 1 Should a kernel estimator or an alternative related estimator (mentioned below) be used?
  - 2 If kernel estimator is chosen, what is the functional form of the kernel?
  - 3 If a kernel estimator is chosen, what bandwidth should be picked?
- Alternative nonparametric estimators to kernel estimation include:
  - nearest neighbor, choosing the number of observations near the point of interest, rather than the bandwidth.
  - local linear regression, fitting a linear (or polynomial) regression through observations that are close to the point of interest.
- Note that if:
  - local regression is chosen, analogous questions to the second two must be addressed. (How many polynomial coefficients and how "local"?)
  - choosing nearest neighbor is "like" choosing a uniform kernel with a different approach about "what points to include".

# Implementation Details

## Some trade-offs

- Practical wisdom suggests that:
  - on the first question the estimators produce comparable results.
  - similarly the choice of the kernel weighting density is rarely a critical issue: independent normal kernels are often selected.
- All these estimators are sensitive to choice of the bandwidth (number of nearest neighbors, and so on):
  - setting the bandwidth on a uniform kernel to cover all the observations yields the unconditional mean
  - setting the bandwidth so tight that one observation is covered doesn't smooth at all
- More generally bandwidth choices balance:
  - bias (from including observations that are too distant from the point of interest)
  - against variance (from including too few observations).

# Implementation Details

Selecting the bandwidth for estimating a density at a point

- One approach is to choose a bandwidth  $\phi$  to minimize an MSE approximation, which (from Lecture 5) can be expressed as:

$$E \left( \left\{ f_x^{(N)}(x) - E \left[ f_x^{(N)}(x) \right] \right\}^2 \right) + \left\{ E \left[ f_x^{(N)}(x) \right] - f_x(x) \right\}^2$$

- One can show:

$$E \left[ f_x^{(N)}(x) \right] - f_x(x) \approx \frac{\phi^2}{2} \left[ f_x'''(x) \right] \int z^2 k(z) dz$$

$$E \left( \left\{ f_x^{(N)}(x) - E \left[ f_x^{(N)}(x) \right] \right\}^2 \right) \approx \frac{1}{N\phi} \left[ f_x(x) \int z^2 k(z) dz \right]$$

- In practice we might:
  - 1 obtain first round estimates of  $f_x^{(N)}(x)$  with "any" bandwidth.
  - 2 solve the minimization problem with the approximation.
  - 3 use the solution in a second round estimator of  $f_x^{(N)}(x)$ .

# Implementation Details

## Least squares cross validation

- Sometimes the overall fit of  $f_x^{(N)}(x)$  to  $f_x(x)$  is of more interest.
- We could minimize the approximate integrated squared error:

$$\begin{aligned} & \arg \min_{\phi} \int \left[ f_x^{(N)}(x) - f_x(x) \right]^2 dx \\ &= \arg \min_{\phi} \left\{ \int \left[ f_x^{(N)}(x) \right]^2 dx - 2E \left[ f_x^{(N)}(x) \right] \right\} \end{aligned}$$

by minimizing the difference of:

$$\begin{aligned} \int \left[ f_x^{(N)}(x) \right]^2 dx &= \frac{1}{(N\phi)^2} \int \left[ \sum_{n=1}^N k \left( \frac{x_n - x}{\phi} \right) \right]^2 dx \\ 2E \left[ f_x^{(N)}(x) \right] &\approx \frac{2}{N(N-1)\phi} \left[ \sum_{n=1}^N \sum_{m=1, m \neq n}^N k \left( \frac{x_m - x_n}{\phi} \right) \right] \end{aligned}$$

with respect to  $\phi$  using numerical methods.

# Imposing Semiparametric Restrictions

## Estimating the support of a distribution function

- To estimate the support of a distribution function, a very different set of procedures is used.
- Suppose  $F(x)$  is a cumulative distribution function with support  $[\underline{x}, \bar{x}]$  for unknown  $\underline{x}$  and  $\bar{x}$  satisfying  $-\infty \leq \underline{x} < \bar{x} \leq \infty$ .
- Assume  $x_n$  is independently drawn from  $F(x)$  for  $n \in \{1, \dots, N\}$ .
- Estimate  $\bar{x}$  with  $\bar{x}^{(N)} \equiv \max\{x_1, \dots, x_N\}$ .
- Similarly estimate  $\underline{x}$  with  $\underline{x}^{(N)} \equiv \min\{x_1, \dots, x_N\}$ .
- Note  $\bar{x}^{(N)}$  is a monotone increasing sequence bounded above by  $\bar{x}$ .
- Also  $P(x_n \leq x^*) = F(x^*) < 1$  for all  $x^* < \bar{x}$   
 $\Rightarrow P(\bar{x}^{(N)} \leq x^*) = F(x^*)^{(N)} \rightarrow 0$  for all  $x^* < \bar{x}$ .
- These observations motivate the choice of  $\bar{x}^{(N)}$  as an estimator for  $\bar{x}$ .
- Analogous arguments apply to the properties of  $\underline{x}^{(N)}$ .

# Imposing Semiparametric Restrictions

## Truncated linear regression

- Suppose:

$$y_n = x_n\beta + \epsilon_n$$

where  $\epsilon_n$  is distributed normal with mean  $\mu$  and variance  $\sigma^2$  and:

$$d_n = \begin{cases} 2 & \text{if } y_n > \bar{y} \\ 1 & \text{if } \underline{y} < y_n \leq \bar{y} \\ 0 & \text{if } y_n < \underline{y} \end{cases}$$

- This model is parametrized by  $\theta \equiv (\beta, \mu, \sigma, \bar{y}, \underline{y})$ .
- Conditional on  $x_n$  the probability density function of  $1\{d_n = 1\}y_n$  is:

$$\exp\left[-\left(\frac{y_n - x_n\beta}{\sigma}\right)^2\right] / \int_{\underline{y}}^{\bar{y}} \exp\left[-\left(\frac{y - x_n\beta}{\sigma}\right)^2\right] dy$$

# Imposing Semiparametric Restrictions

An estimator for the truncated linear regression

- Suppose we observe  $(y_n, x_n)$  for a truncated sample of  $N$  observations in which  $d_n = 1$  and  $\theta_0 \equiv (\beta_0, \mu_0, \sigma_0, \bar{y}_0, \underline{y}_0)$ .
- Rather than estimating  $\theta_0$  simultaneously, we estimate:
  - 1  $\bar{y}_0$  with  $\bar{y}^{(N)}$  and  $\underline{y}_0$  with  $\underline{y}^{(N)}$  where  $\bar{y}^{(N)} \equiv \max \{y_1, \dots, y_N\}$  and  $\underline{y}^{(N)} \equiv \min \{y_1, \dots, y_N\}$ .
  - 2  $(\beta_0, \mu_0, \sigma_0)$  with ML (or NLS) as if  $\bar{y}_0 = \bar{y}^{(N)}$  and  $\underline{y}_0 = \underline{y}^{(N)}$ :

$$\begin{aligned} & \arg \min_{\beta, \mu, \sigma} \prod_{n=1}^N \left\{ \frac{\exp \left[ \frac{-1}{2\sigma^2} (y_n - x_n \beta)^2 \right]}{\int_{\underline{y}^{(N)}}^{\bar{y}^{(N)}} \exp \left[ \frac{-1}{2\sigma^2} (y_n - x_n \beta)^2 \right] dy} \right\} \\ &= \arg \min_{\beta, \mu, \sigma} \sum_{n=1}^N \left[ (y_n - x_n \beta)^2 - \ln \left\{ \int_{\underline{y}^{(N)}}^{\bar{y}^{(N)}} \exp \left[ \frac{-1}{2\sigma^2} (y_n - x_n \beta)^2 \right] dy \right\} \right] \end{aligned}$$

# Imposing Semiparametric Restrictions

Nonparametric monotone conditional expectations functions (Brunk,1958)

- To further develop this line of enquiry, suppose  $E[y|x] \equiv g(x)$  and  $g(x)$  is (weakly) increasing in all its arguments.
- Suppose, as before,  $(y_n, x_n)$  are independent for  $n \in \{1, \dots, N\}$ .
- To estimate  $g(x^*)$  at  $x^*$  for the partial sums:

$$\max_x \left\{ \frac{\sum_{n=1}^N \mathbf{1}\{x < x_n \leq x^*\} y_n}{\sum_{n=1}^N \mathbf{1}\{x < x_n \leq x^*\}} \right\} \quad (2)$$

or:

$$\min_x \left\{ \frac{\sum_{n=1}^N \mathbf{1}\{x^* < x_n \leq x\} y_n}{\sum_{n=1}^N \mathbf{1}\{x^* < x_n \leq x\}} \right\} \quad (3)$$

- Thus the criterion function determining  $x$ , justified by the monotonicity assumption, determines the bandwidth automatically, contrasting with the discretion of bandwidth choice for an unrestricted nonparametric regression function estimator.

# Imposing Semiparametric Restrictions

## Partially linear models

- Consider the following equation of augmenting a linear framework with an additive nonlinear real valued function:

$$y_n = x_n' \beta_0 + g(z_n) + \epsilon_n \quad (4)$$

where:

- $y$  is the dependent variable:
  - $x$  and  $z$  are explanatory variables with no common components
  - $g$  has an unknown functional form
  - $\beta_0$  and  $g(z)$  are the parameters to be estimated
- We make the standard assumption that:

$$E[\epsilon_n | x_n] = E[\epsilon_n | z_n] = 0 \quad (5)$$

# Imposing Semiparametric Restrictions

## Inference in partially linear models

- The key to identification and estimation in this model is that given  $z_n$  there is covariation between  $y_n$  and  $x_n$ . Let:

$$\begin{aligned}\tilde{y}_n &\equiv y_n - E[y_n | z_n] \\ \tilde{x}'_n &\equiv x'_n - E[x'_n | z_n]\end{aligned}$$

- 1 Nonparametrically estimate  $E[y_n | z_n]$  and  $E[x'_n | z_n]$ .
- 2 Estimate  $\beta_0$  by regressing  $\tilde{y}_n$  on  $\tilde{x}_n$ , and exploiting the identity:

$$\tilde{y}_n = \tilde{x}'_n \beta_0 + \epsilon_n \quad (6)$$

- 3 Substitute the estimates of  $\beta_0$ ,  $E[y_n | z_n]$  and  $E[x'_n | z_n]$  into:

$$g(z_n) = E[y_n | z_n] - E[x'_n | z_n] \beta_0$$

# Imposing Semiparametric Restrictions

Identities behind inference in partially linear models

- From (4) and (5):

$$E[y_n | z_n] = E[x_n' | z_n] \beta_0 + g(z_n)$$

- Then (6) follows from subtracting this equation from (4):

$$\tilde{y}_n \equiv y_n - E[y_n | z_n] = \{x_n' - E[x_n' | z_n]\} \beta_0 + \epsilon_n \equiv \tilde{x}_n' \beta_0 + \epsilon_n$$

- Multiplying (6) by  $\tilde{x}_n$  and taking expectations yields:

$$E[\tilde{x}_n \tilde{y}_n] = E[\tilde{x}_n \tilde{x}_n'] \beta_0$$

- If  $E[\tilde{x}_n \tilde{x}_n']$  is invertible then:

$$\beta_0 = E[\tilde{x}_n \tilde{x}_n']^{-1} E[\tilde{x}_n \tilde{y}_n]$$

and:

$$g(z_n) = E[y_n | z_n] - E[x_n' | z_n] E[\tilde{x}_n \tilde{x}_n']^{-1} E[\tilde{x}_n \tilde{y}_n]$$

# Imposing Semiparametric Restrictions

## Single index models

- Single index models are defined by the equation:

$$y_n = g(x_n' \beta_0) + \epsilon_n \quad (7)$$

where  $x_n' \beta_0$  is an index and:

- $y$  is the dependent variable.
  - $x$  is a  $k \times 1$  vector of explanatory variables.
  - $g : \mathbb{R} \rightarrow \mathbb{R}$  has an unknown functional form to be estimated.
  - $\beta_0$  is a  $k \times 1$  parameter vector to be estimated too.
- Yet again we assume  $E[\epsilon_n | x_n] = 0$ .
  - This assumption implies  $E[y_n | x_n] = g(x_n' \beta_0)$ .

# Imposing Semiparametric Restrictions

A direct estimator of single index models (Ichimura, 1993)

- If  $g(\cdot)$  was known,  $\beta_0$  could be estimated with NLS.
- Alternatively if  $\beta_0$  was known,  $g(x'_n\beta_0)$  could be estimated with a nonparametric regression.
- The following estimator combines these ideas, choosing  $\beta$  to solve:

$$\arg \min_{\beta} \sum_{n=1}^N [y_n - \hat{g}(x'_n\beta)]^2 = \arg \min_{\beta} \sum_{n=1}^N \left[ \hat{g}(x'_n\beta)^2 - 2y_n\hat{g}(x'_n\beta) \right]$$

- with  $k \times 1$  FOC:

$$\frac{1}{N} \sum_{n=1}^N \hat{g}(x'_n\beta) \hat{g}'(x'_n\beta) x_n = \frac{1}{N} \sum_{n=1}^N y_n \hat{g}'(x'_n\beta) x_n$$

- where  $g(\cdot)$  is approximated by the kernel estimator  $\hat{g}(\cdot)$ :

$$\hat{g}(x'\beta) = \sum_{n=1}^N w^{(N)}(x', x'_n\beta) y_n$$

- and weights  $w^{(N)}(x, x_n)$  are defined:

$$w^{(N)}(x'\beta, x'_n\beta) \equiv k_x \left( \frac{x'_n - x'}{\phi(N)} \beta \right) / \sum_{n=1}^N k_x \left( \frac{x'_n - x'}{\phi(N)} \beta \right)$$