Nonparametric Methods

Robert A. Miller

Structural Econometrics

November 2021

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

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

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

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

- 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:

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

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

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

- This result can be extended to obtain expressions for $E_{y|x}^{(N)}(y|x^* < x \leq x^{**})$ 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:

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

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 $\phi^{(N)} \to 0$ but $N\phi^{(N)} \to \infty$ as $N \to \infty$.

Kernel Estimator

Estimating the survivor function and hazard rate

- For example, denoting by $p(t)$ the probability of stopping at $t \in \{1, 2, \ldots, T\}$, 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:

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

- 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:

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

- 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:

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

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

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

The Conditional Expectation Function

Estimating the conditional expectation function

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

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

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

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

and $f\left(y,x\right)$:

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

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:

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

- Note that $\phi^{(N)} \to 0$ as $N \to \infty$ 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:

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

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

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

Applying the Kernel Estimator

Heteroskedasticity with an unknown functional form

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

$$y_n = x_n' \beta_0 + \epsilon_n \tag{2}$$

  where:

  - y is the dependent variable and x is a $K \times 1$ explanatory variable vector
  - β_0 is a $K \times 1$ parameter vector to be estimated
  - ϵ is an independently distributed disturbance with $E\left[\epsilon_n \left| x_n \right.\right] = 0$ and

$$E\left[\epsilon_n^2 \left| x_n \right.\right] = \psi\left(x\right) \tag{3}$$

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

Applying the Kernel Estimator

Controlling for heteroskedasticity

- We proceed stepwise:
  1. Run least squares of y on x to obtain β^N_{OLS}.
  2. Form $\widehat{\epsilon}^2_n \equiv \left(y_n - x'_n \beta^N_{OLS}\right)^2$
  3. Estimate $\psi(x_n)$ with:

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

  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 β_0, which we denote by β^N_{HET}.

Applying the Kernel Estimator

Comparing this estimator with the GLS analogue

- If $h(x_n)$ was known, the we could obtain β_{GLS}^N, the GLS estimator, denoted by β_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 β_{GLS}^N is unbiased, but β_{HET}^N is biased, because:

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

- The assumptions above imply β_{OLS}^N is unbiased, begging the question whether β_{HET}^N or β_{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 though 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 ϕ 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:

$$\int \left[f_x^{(N)}(x) \right]^2 dx = \frac{1}{(N\phi)^2} \int \left[\sum_{n=1}^{N} k\left(\frac{x_n - x}{\phi} \right) \right]^2 dx$$

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

with respect to ϕ using numerical methods.

- 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 $[\underline{x}, \overline{x}]$ for unknown $\underline{x}$ and $\overline{x}$ satisfying $-\infty \leq \underline{x} < \overline{x} \leq \infty$.
- Assume x_n is independently drawn from $F(x)$ for $n \in \{1, \ldots, N\}$.
- Estimate $\overline{x}$ with $\overline{x}^{(N)} \equiv \max\{x_1, \ldots, x_N\}$.
- Similarly estimate $\underline{x}$ with $\underline{x}^{(N)} \equiv \min\{x_1, \ldots, x_N\}$.
- Note $\overline{x}^{(N)}$ is a monotone increasing sequence bounded above by $\overline{x}$.
- Also $P(x_n \leq x^*) = F(x^*) < 1$ for all $x^* < \overline{x}$
$\Rightarrow P\left(\overline{x}^{(N)} \leq x^*\right) = F(x^*)^{(N)} \rightarrow 0$ for all $x^* < \overline{x}$.
- These observations motivate the choice of $\overline{x}^{(N)}$ as an estimator for $\overline{x}$.
- 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 ϵ_n is distributed normal with mean μ and variance σ^2 and:

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

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

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

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 $\theta_0 \equiv \left(\beta_0, \mu_0, \sigma_0, \overline{y}_0, \underline{y}_0\right)$.
- Rather than estimating θ_0 simultaneously, we estimate:
  1. $\overline{y}_0$ with $\overline{y}^{(N)}$ and $\underline{y}_0$ with $\underline{y}^{(N)}$ where $\overline{y}^{(N)} \equiv \max\{y_1, \ldots, y_N\}$ and $\underline{y}^{(N)} \equiv \min\{y_1, \ldots, y_N\}$.
  2. $(\beta_0, \mu_0, \sigma_0)$ with ML (or NLS) as if $\overline{y}_0 = \overline{y}^{(N)}$ and $\underline{y}_0 = \underline{y}^{(N)}$:

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

Imposing Semiparametric Restrictions

Nonparametric monotone conditional expectations functions (Brunk,1958)

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

$$\max_{x}\left\{\frac{\sum_{n=1}^{N} 1\left\{x < x_n \leq x^*\right\} y_n}{\sum_{n=1}^{N} 1\left\{x < x_n \leq x^*\right\}}\right\} \tag{4}$$

or:

$$\min_{x}\left\{\frac{\sum_{n=1}^{N} 1\left\{x^* < x_n \leq x\right\} y_n}{\sum_{n=1}^{N} 1\left\{x^* < x_n \leq x\right\}}\right\} \tag{5}$$

- 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.

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

$$y_n = x_n' \beta_0 + g(z_n) + \epsilon_n \tag{6}$$

  where:

  - y is the dependent variable:
  - x and z are explanatory variables with no common components
  - g has an unknown functional form
  - β_0 and $g(z)$ are the parameters to be estimated

- We make the standard assumption that:

$$E[\epsilon_n | x_n] = E[\epsilon_n | z_n] = 0 \tag{7}$$

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

$$\begin{aligned} \widetilde{y}_n &\equiv y_n - E\left[y_n \,|z_n\right] \\ \widetilde{x}_n' &\equiv x_n' - E\left[x_n' \,|z_n\right] \end{aligned}$$

1. Nonparametrically estimate $E\left[y_n \,|z_n\right]$ and $E\left[x_n' \,|z_n\right]$.
2. Estimate β_0 by regressing $\widetilde{y}_n$ on $\widetilde{x}_n$, and exploiting the identity:

$$\widetilde{y}_n = \widetilde{x}_n' \beta_0 + \epsilon_n \tag{8}$$

3. Substitute the estimates of β_0, $E\left[y_n \,|z_n\right]$ and $E\left[x_n' \,|z_n\right]$ into:

$$g\left(z_n\right) = E\left[y_n \,|z_n\right] - E\left[x_n' \,|z_n\right]\beta_0$$

Imposing Semiparametric Restrictions

Identities behind inference in partially linear models

- From (6) and (7):

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

- Then (8) follows from subtracting this equation from (6):

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

- Multiplying (8) by $\widetilde{x}_n$ and taking expectations yields:

$$E\left[\widetilde{x}_n\widetilde{y}_n\right] = E\left[\widetilde{x}_n\widetilde{x}_n'\right]\beta_0$$

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

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

and:

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

- Single index models are defined by the equation:

$$y_n = g\left(x_n'\beta_0\right) + \epsilon_n \tag{9}$$

  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.
  - β_0 is a $k \times 1$ parameter vector to be estimated too.
- Yet again we assume $E\left[\epsilon_n \left| x_n \right.\right] = 0$.
- This assumption implies $E\left[y_n \left| x_n \right.\right] = g\left(x_n'\beta_0\right)$.

Imposing Semiparametric Restrictions

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

- If $g(\cdot)$ was known, β_0 could be estimated with NLS.
- Alternatively if β_0 was known, $g(x_n'\beta_0)$ could be estimated with a nonparametric regression.
- The following estimator combines these ideas, choosing β to solve:

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

  - with $k \times 1$ FOC:

$$\frac{1}{N}\sum_{n=1}^{N} \widehat{g}(x_n'\beta)\,\widehat{g}'(x_n'\beta)\,x_n = \frac{1}{N}\sum_{n=1}^{N} y_n'\widehat{g}(x_n'\beta)\,x_n$$

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

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

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

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