Local or Global Smoothing? A Bandwidth Selector for Dependent Data

Francesco Giordano – Maria Lucia Parrella

Department of Economics and Statistics
University of Salerno

COMPSTAT 2010

Aims and prospects

This is part of a work in progress. The context and the goals of this research are:

1. **analyzing the problem of bandwidth selection** in local polynomial estimation (**LPE**) of dependent data;
2. **proposing a locally adaptive bandwidth selector**, based on the use of the neural network estimator (**NNE**);
3. **evaluating the gain in using a local bandwidth** instead of a global (fixed) bandwidth.

Here we present some theoretical and computational results....

Motivations

- Bandwidths are very crucial in LPE!
- One expects local bandwidths to perform better than global bandwidths. Is it really true when we use the ***estimated bandwidths***?
- The available selection procedures may present some drawbacks:
  - often only for global bandwidths;
  - often not for dependent data;
  - sometimes computationally expensive;
  - sometimes extremely biased and/or variable;
  - sometimes not fully automatic.

Recent research in bandwidth selection

Generally, bandwidth selection may be used as a tool to deal with different problems in several contexts. See, for example:

- Gao and Gijbels (JASA, 2008): **size and power** in nonparametric kernel testing.
- Gluhovsky and Gluhovsky (JASA, 2007): **smooth conditions** in kernel regression.
- Lafferty and Wassermann (AS, 2008): **variable selection** in high dimension problems.
- Prewitt and Lohr (JRSS, 2006): **multicollinearity** in local regression.

Our framework

Our proposal applies to different setups. In this talk we consider, in particular, the following framework:

- **DGP**: real strictly stationary process $\{X_t\}$
- **Model**:

$$X_t = m(X_{t-1}) + \sigma(X_{t-1})\epsilon_t, \qquad \epsilon_t \sim i.i.d.(0,1)$$

- **Context**: estimation of the volatility function by LPE

$$\sigma^2(x) = Var\left[X_t | X_{t-1} = x\right], \qquad x \in \mathbb{R}.$$

- **Note**: for simplicity, here we consider $m(x) \equiv 0$.

The local polynomial estimator of volatility

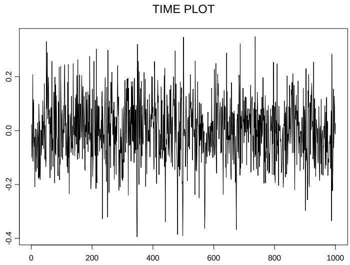
The volatility function $\sigma^2(x)$ may be estimated by local polynomials:

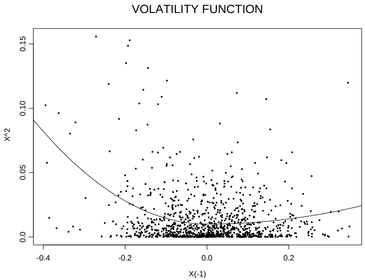
$$\hat{\sigma}^2(x;h) = \sum_{t=1}^{n} X_t^2 W_{K,h,p}(x - X_{t-1}).$$

- $W_{K,h,p}(\cdot)$ are the weights, derived from a local approximation of $\sigma^2(x)$ through a polynomial of order p;
- K is the *Kernel function*;
- h is the *bandwidth*, which regulates the smoothness of $\hat{\sigma}^2(x;h)$.

An illustrative example of LPE

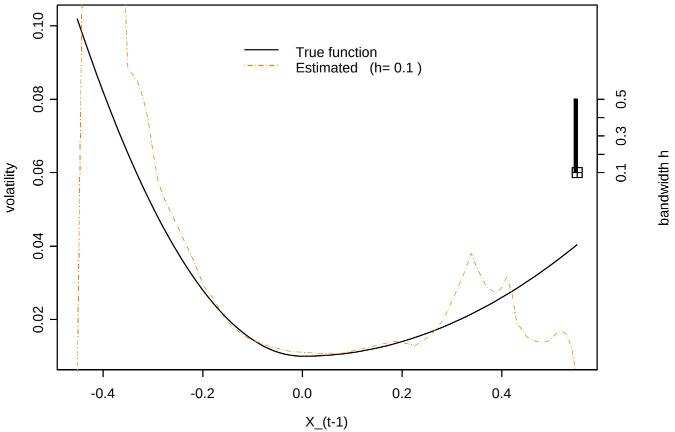
$$X_t = \sqrt{0.01 + 0.1X_{t-1}^2 + 0.35X_{t-1}^2\mathbb{I}(X_{t-1} < 0)} * \varepsilon_t$$





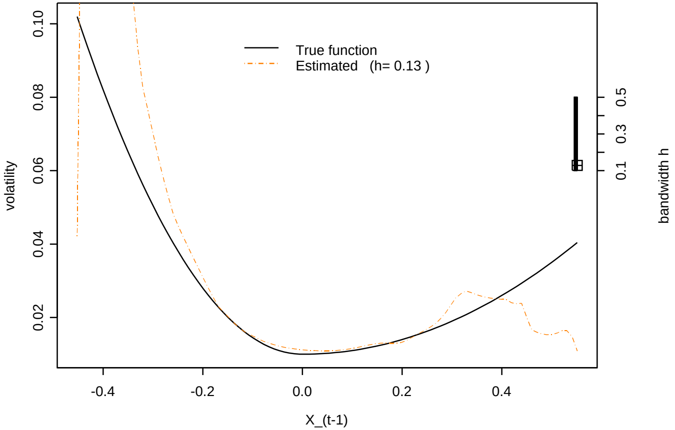
An illustrative example of LPE

$$\sigma^2(x) = 0.01 + 0.1x^2 + 0.35x^2\mathbb{I}(x < 0)$$



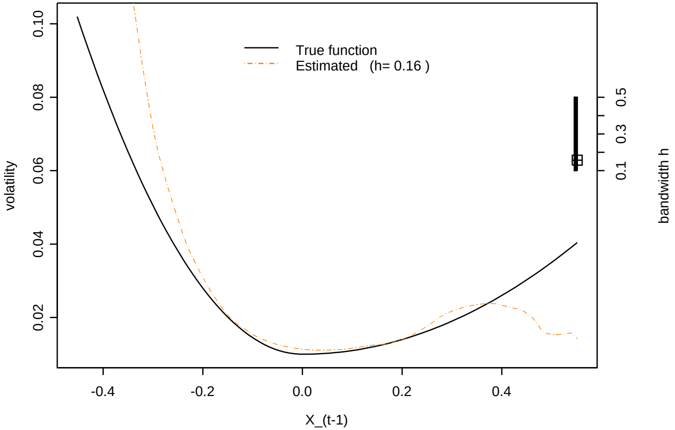
An illustrative example of LPE

$$\sigma^2(x) = 0.01 + 0.1x^2 + 0.35x^2\mathbb{I}(x < 0)$$



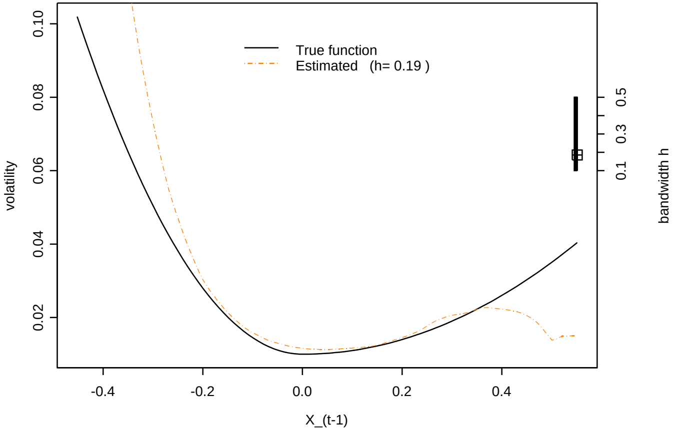
An illustrative example of LPE

$$\sigma^2(x) = 0.01 + 0.1x^2 + 0.35x^2\mathbb{I}(x < 0)$$



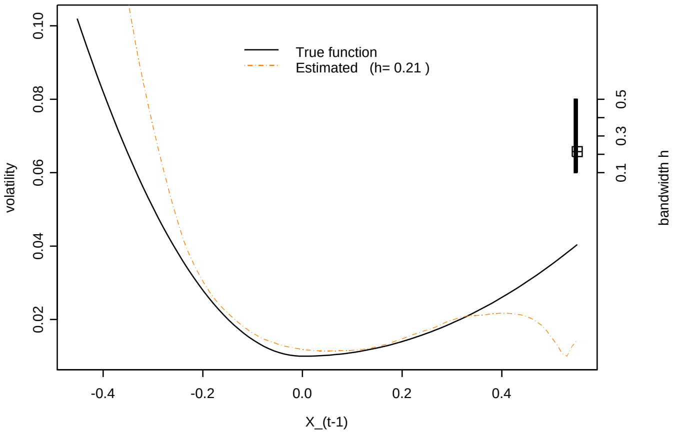
An illustrative example of LPE

$$\sigma^2(x) = 0.01 + 0.1x^2 + 0.35x^2\mathbb{I}(x < 0)$$



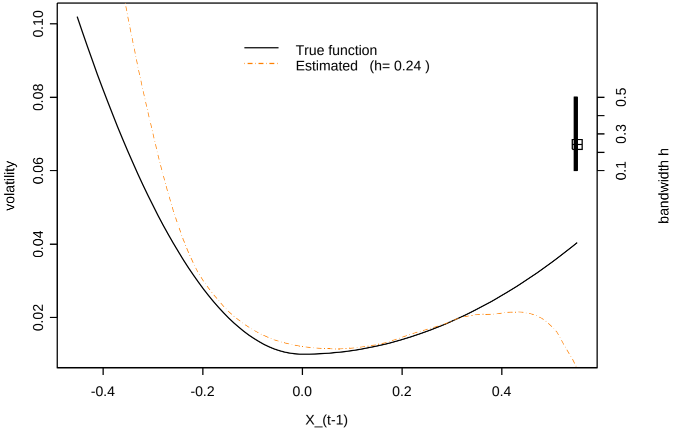
An illustrative example of LPE

$$\sigma^2(x) = 0.01 + 0.1x^2 + 0.35x^2\mathbb{I}(x < 0)$$



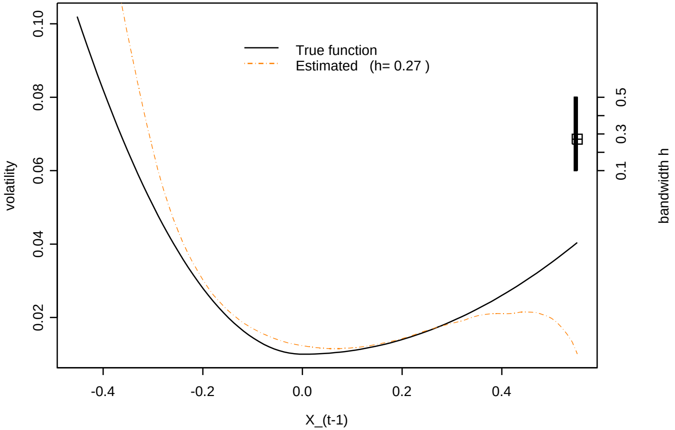
An illustrative example of LPE

$$\sigma^2(x) = 0.01 + 0.1x^2 + 0.35x^2\mathbb{I}(x < 0)$$



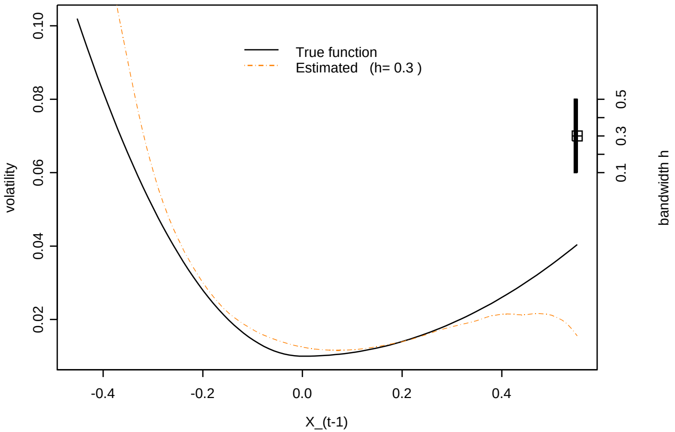
An illustrative example of LPE

$$\sigma^2(x) = 0.01 + 0.1x^2 + 0.35x^2\mathbb{I}(x < 0)$$



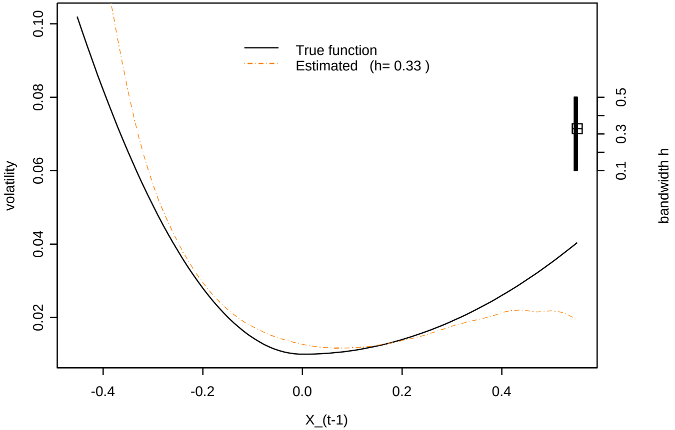
An illustrative example of LPE

$$\sigma^2(x) = 0.01 + 0.1x^2 + 0.35x^2\mathbb{I}(x < 0)$$



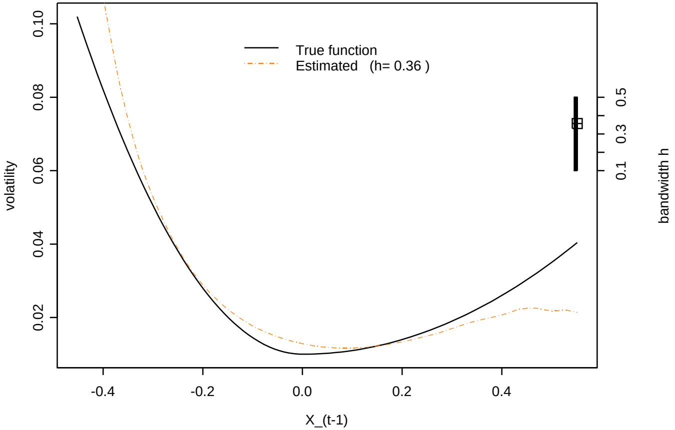
An illustrative example of LPE

$$\sigma^2(x) = 0.01 + 0.1x^2 + 0.35x^2\mathbb{I}(x < 0)$$



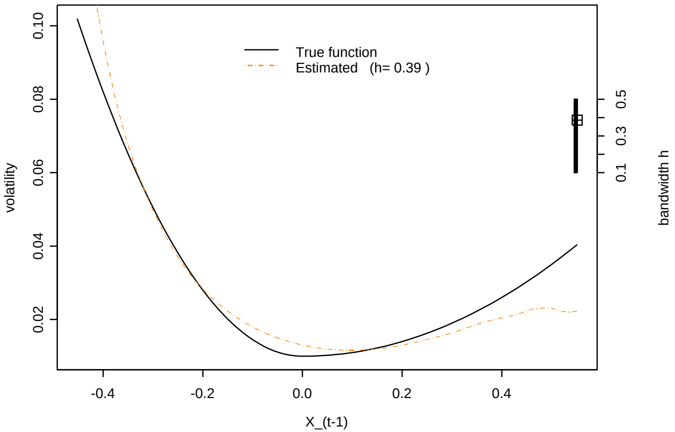
An illustrative example of LPE

$$\sigma^2(x) = 0.01 + 0.1x^2 + 0.35x^2\mathbb{I}(x < 0)$$



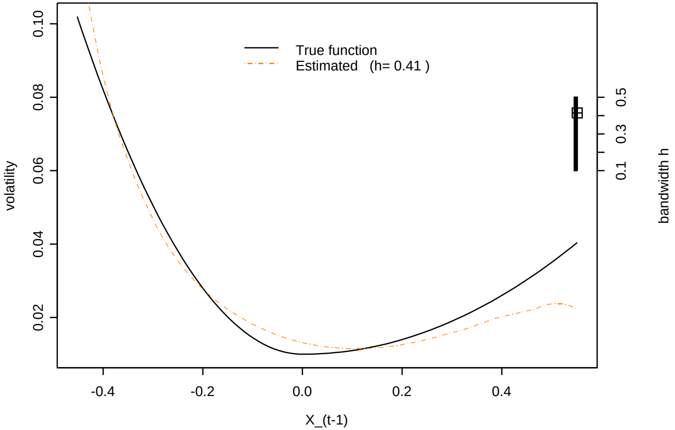
An illustrative example of LPE

$$\sigma^2(x) = 0.01 + 0.1x^2 + 0.35x^2\mathbb{I}(x < 0)$$



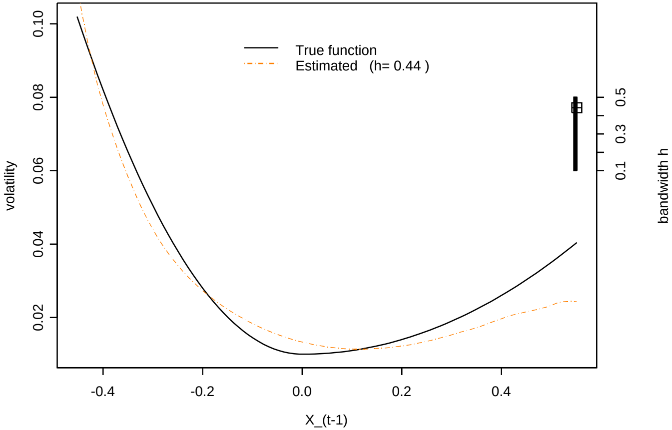
An illustrative example of LPE

$$\sigma^2(x) = 0.01 + 0.1x^2 + 0.35x^2\mathbb{I}(x < 0)$$



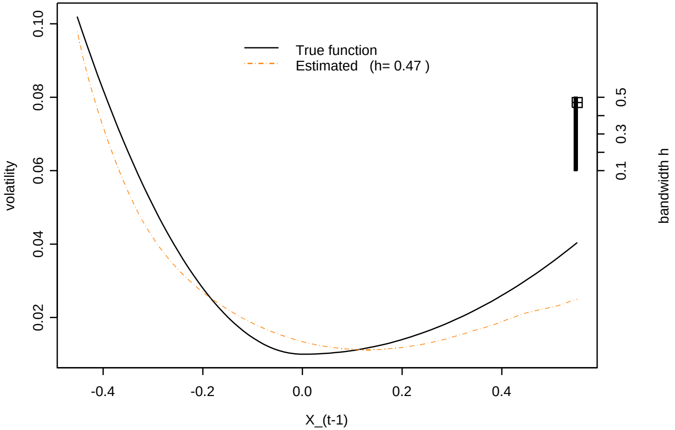
An illustrative example of LPE

$$\sigma^2(x) = 0.01 + 0.1x^2 + 0.35x^2 \mathbb{I}(x < 0)$$



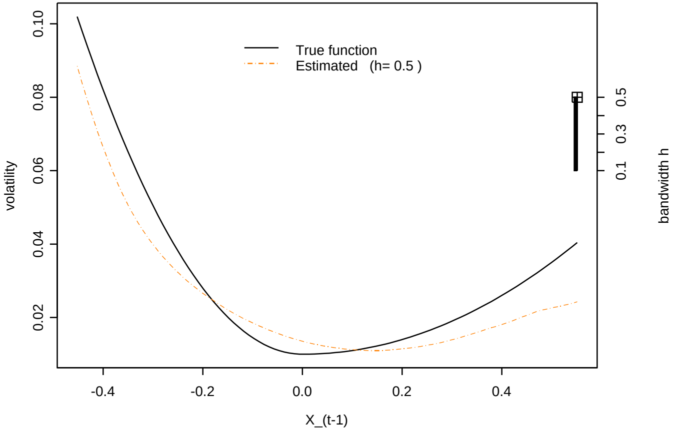
An illustrative example of LPE

$$\sigma^2(x) = 0.01 + 0.1x^2 + 0.35x^2\mathbb{I}(x < 0)$$



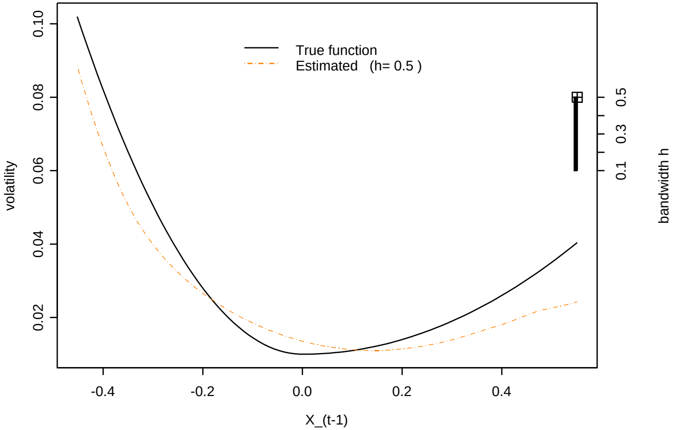
An illustrative example of LPE

$$\sigma^2(x) = 0.01 + 0.1x^2 + 0.35x^2\mathbb{I}(x < 0)$$



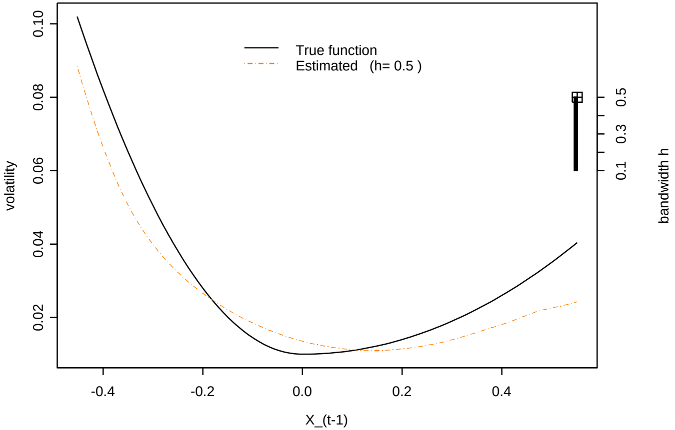
An illustrative example of LPE

$$\sigma^2(x) = 0.01 + 0.1x^2 + 0.35x^2\mathbb{I}(x < 0)$$



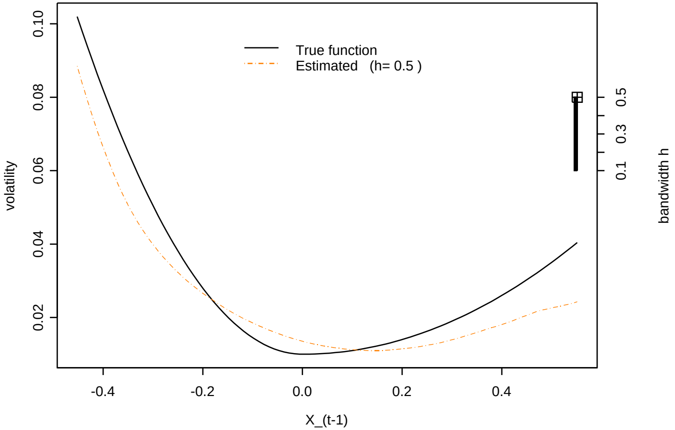
An illustrative example of LPE

$$\sigma^2(x) = 0.01 + 0.1x^2 + 0.35x^2\mathbb{I}(x < 0)$$



An illustrative example of LPE

$$\sigma^2(x) = 0.01 + 0.1x^2 + 0.35x^2\mathbb{I}(x < 0)$$



The plug-in optimal **local** bandwidth

- It is derived by minimizing the asymptotic mean square error

$$
\begin{aligned}
h_L^{opt}(x) &= \arg\min_h AMSE\{\hat{\sigma}^2(x;h)\} \\
&= \left\{ \frac{C_{p,K} \times v(x)}{n \times d^2(x) \times f_X(x)} \right\}^{1/(2p+3)}
\end{aligned}
$$

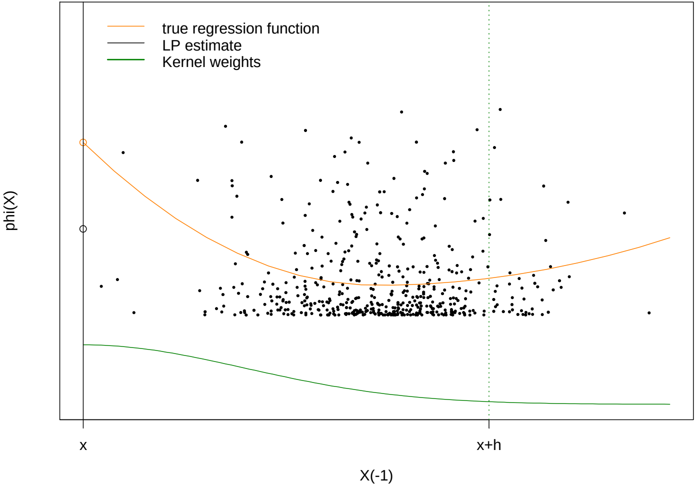
The plug-in optimal **local** bandwidth

- It is derived by minimizing the asymptotic mean square error

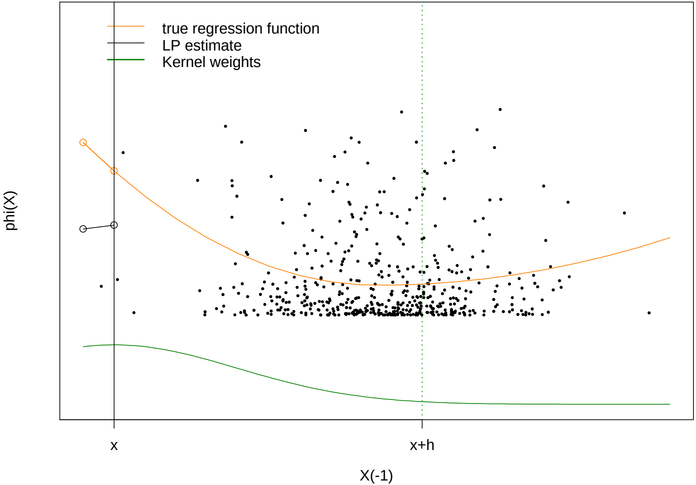
$$\begin{aligned} h_L^{opt}(x) &= \arg\min_h AMSE\{\hat{\sigma}^2(x;h)\} \\ &= \left\{ \frac{C_{p,K} \times v(x)}{n \times d^2(x) \times f_X(x)} \right\}^{1/(2p+3)} \end{aligned}$$

- The only unknown components (to estimate and plug-in) are
  - $v(x)$, the conditional variance of X_t^2;
  - $d(x)$, the $(p+1)$-th derivative of $\sigma^2(x)$;
  - $f_X(x)$, the design density.

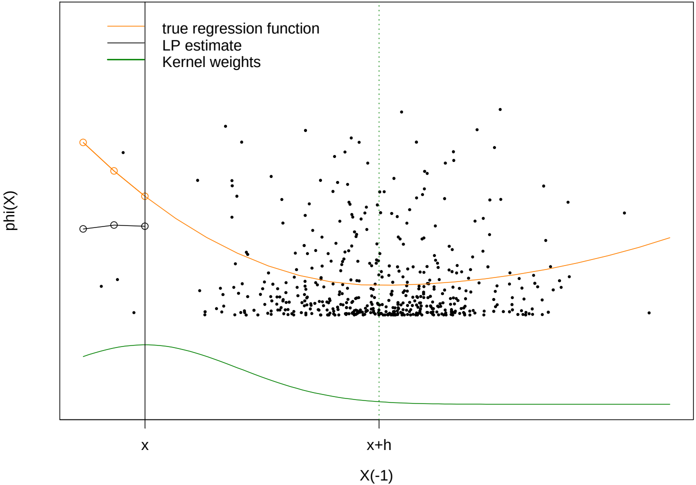
An illustrative example: local (variable) bandwidth



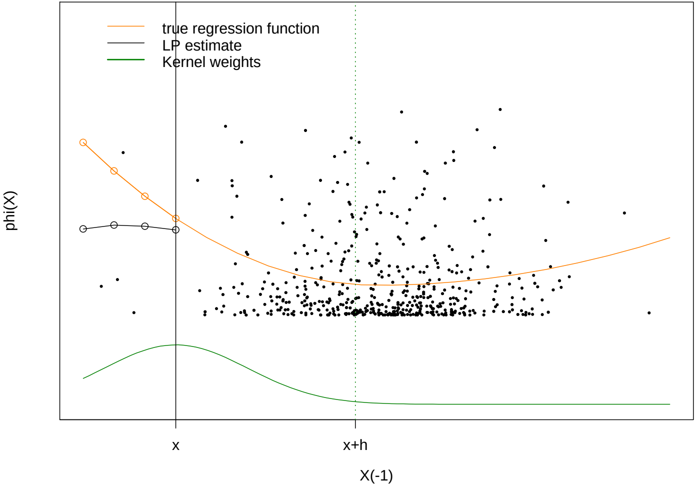
An illustrative example: local (variable) bandwidth



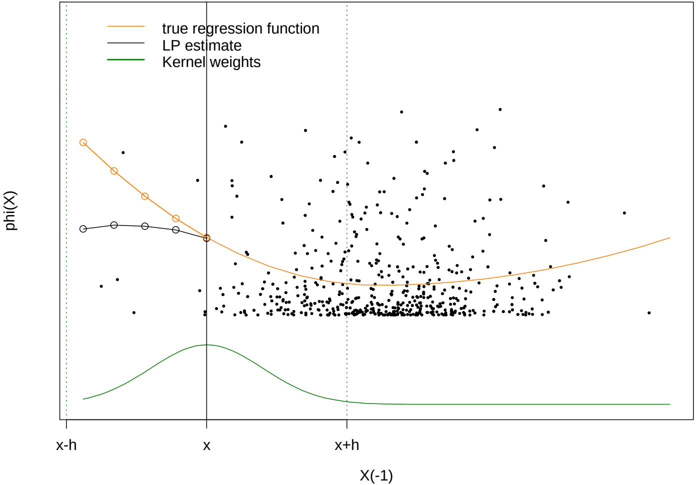
An illustrative example: local (variable) bandwidth



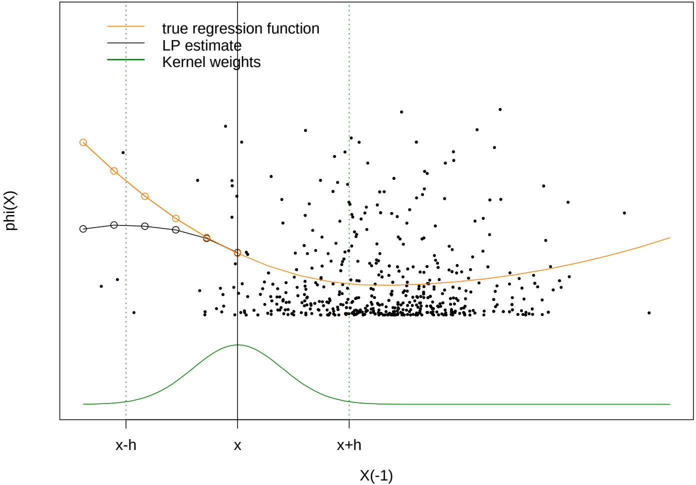
An illustrative example: local (variable) bandwidth



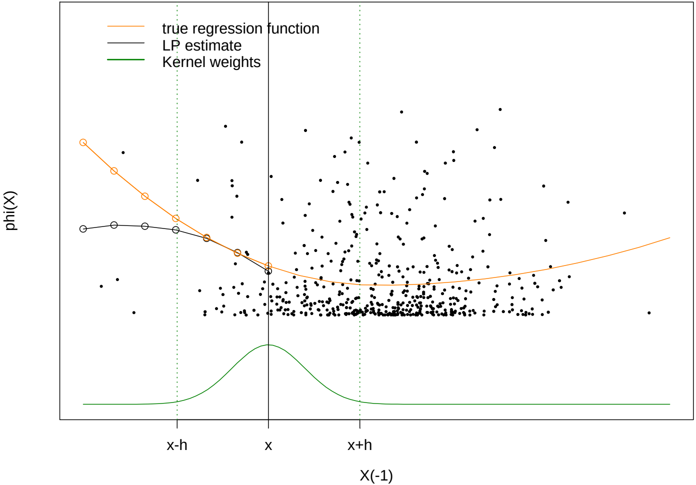
An illustrative example: local (variable) bandwidth



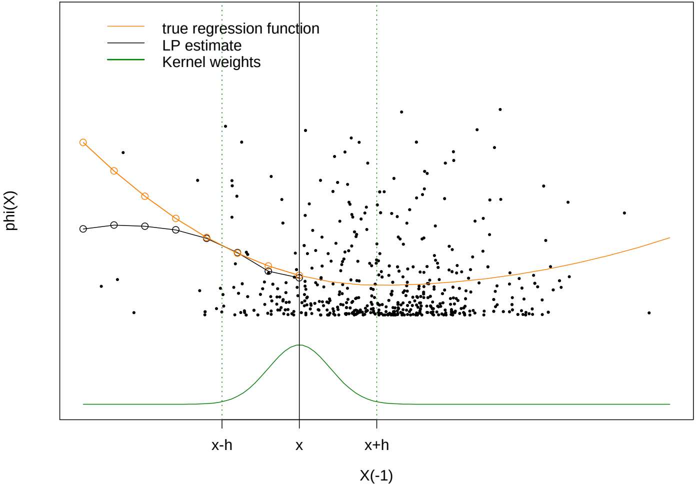
An illustrative example: local (variable) bandwidth



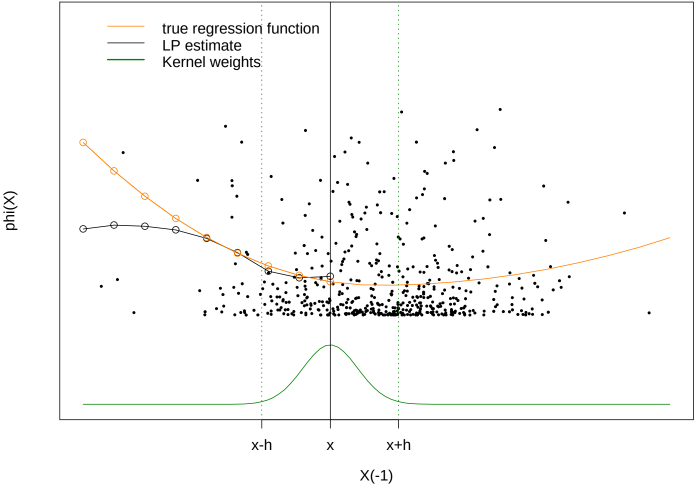
An illustrative example: local (variable) bandwidth



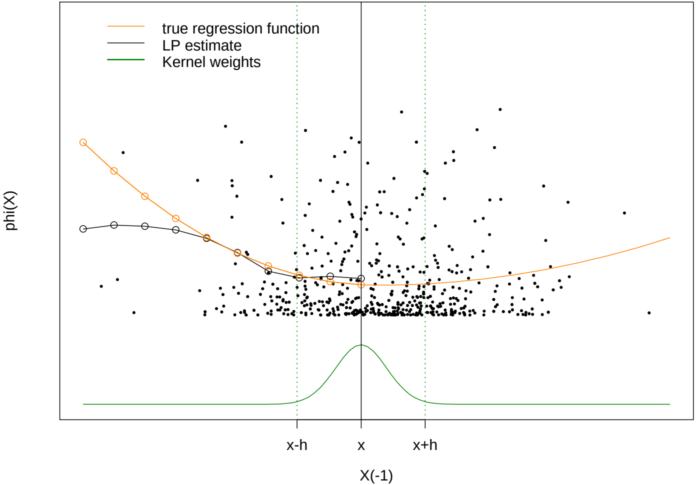
An illustrative example: local (variable) bandwidth



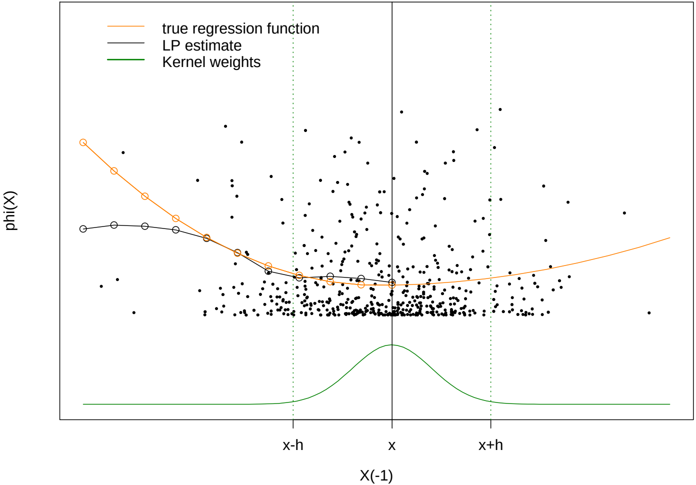
An illustrative example: local (variable) bandwidth



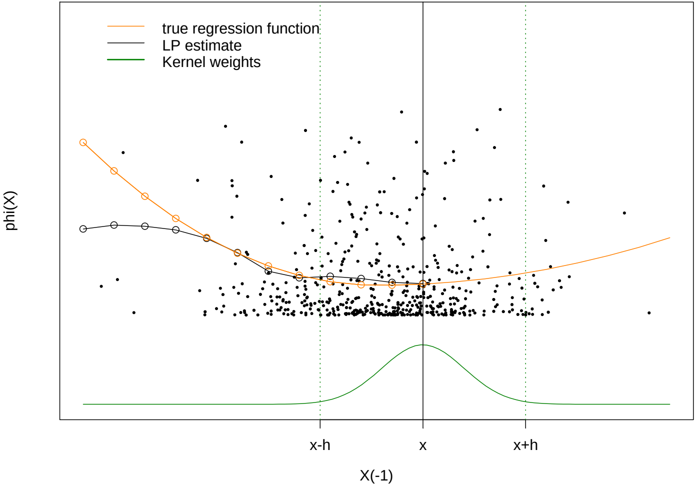
An illustrative example: local (variable) bandwidth



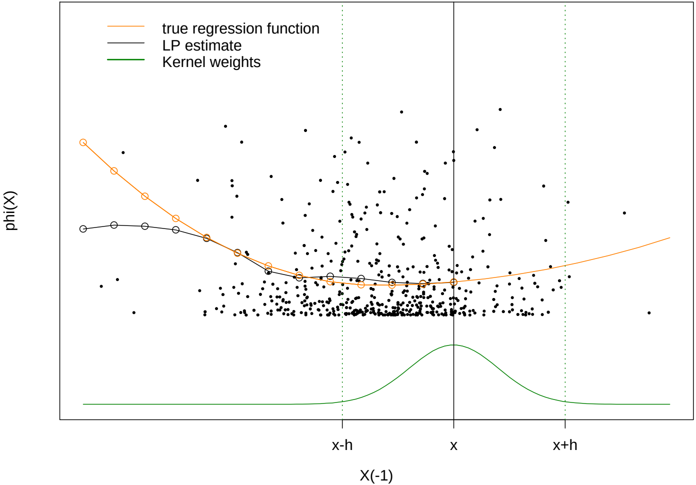
An illustrative example: local (variable) bandwidth



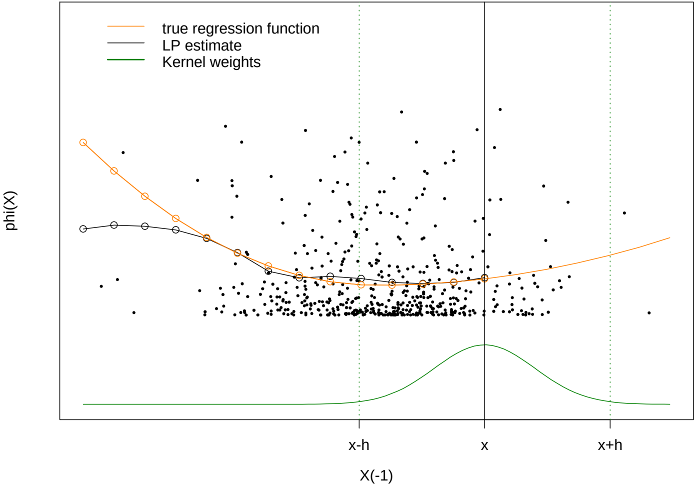
An illustrative example: local (variable) bandwidth



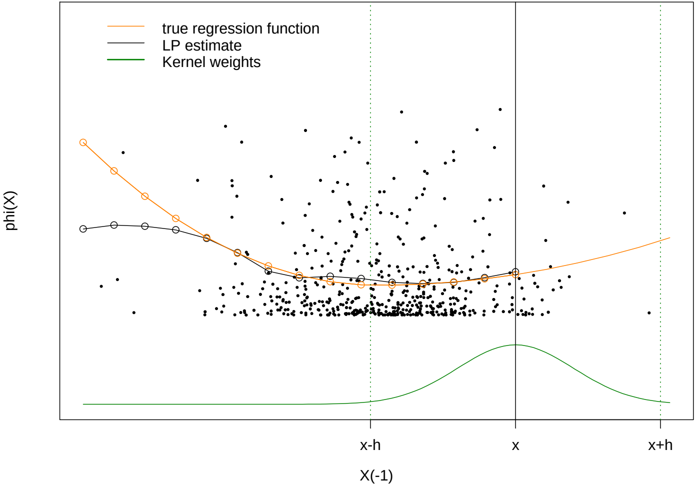
An illustrative example: local (variable) bandwidth



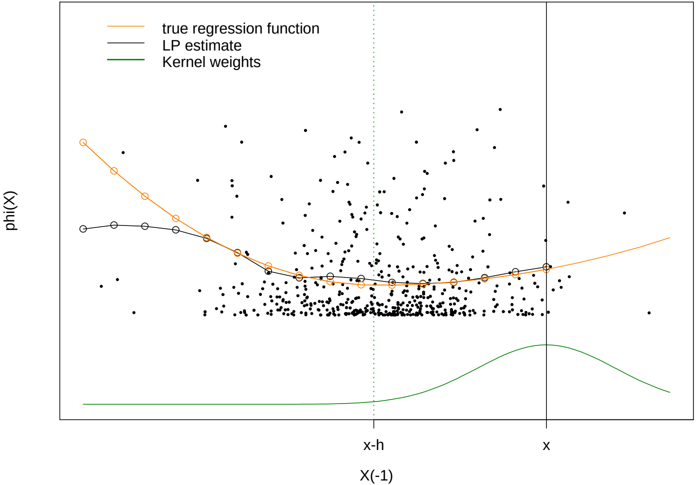
An illustrative example: local (variable) bandwidth



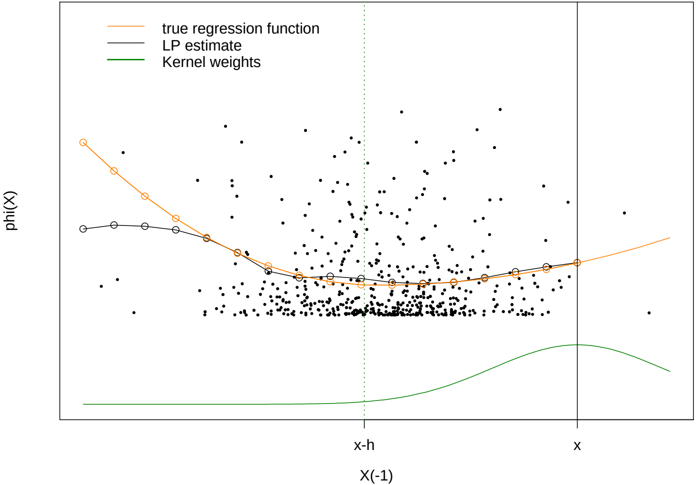
An illustrative example: local (variable) bandwidth



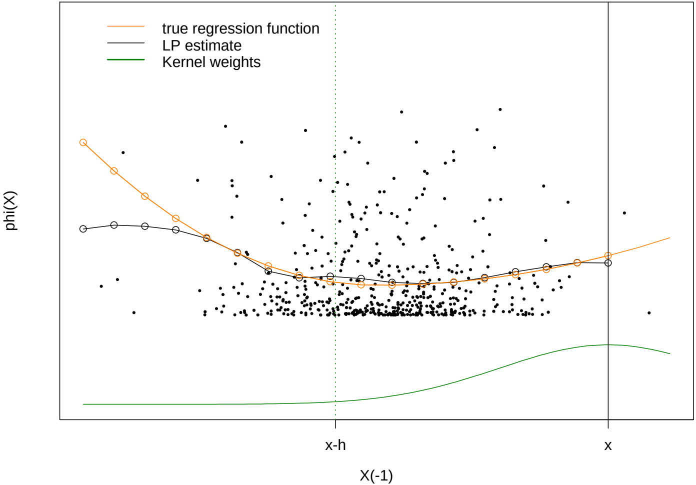
An illustrative example: local (variable) bandwidth



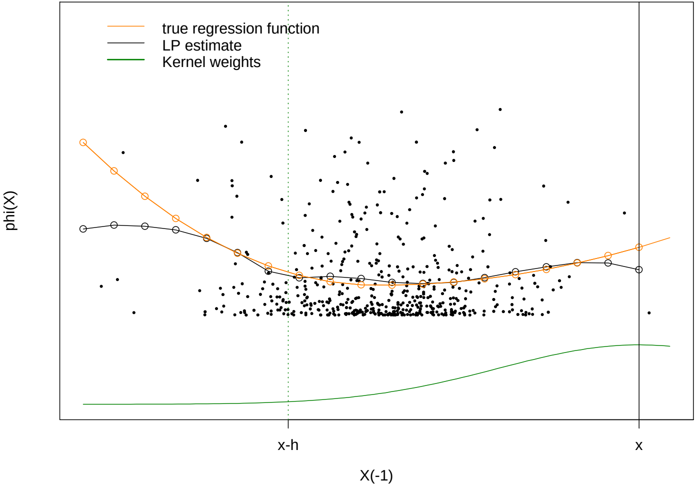
An illustrative example: local (variable) bandwidth



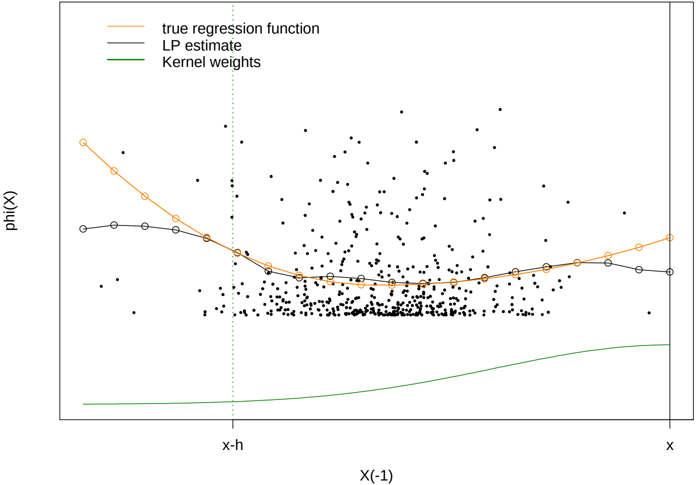
An illustrative example: local (variable) bandwidth



An illustrative example: local (variable) bandwidth



An illustrative example: local (variable) bandwidth



The plug-in optimal **global** bandwidth

- It is derived by minimizing the integrated AMSE

$$
\begin{aligned}
h_G^{opt} &= \arg\min_h \int AMSE\{\hat{\sigma}^2(x;h)\} f_X(x)dx \\
&= \left\{ \frac{C_{p,K} \times R_{\mathrm{var}}}{n \times R_{f,\mathrm{bias}}} \right\}^{1/(2p+3)}
\end{aligned}
$$

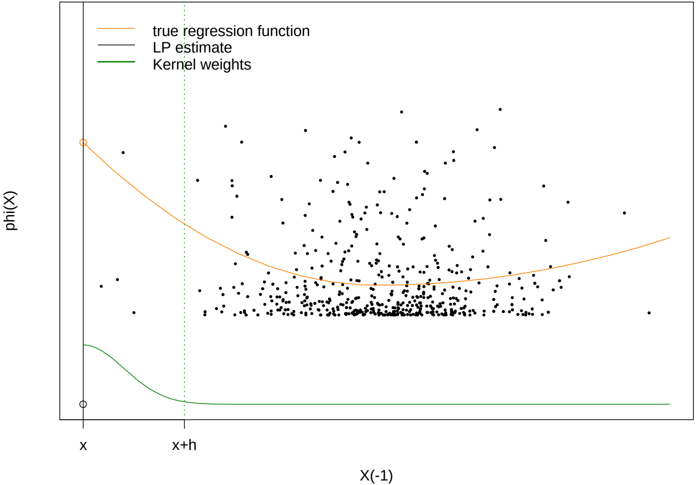
The plug-in optimal **global** bandwidth

- It is derived by minimizing the integrated AMSE

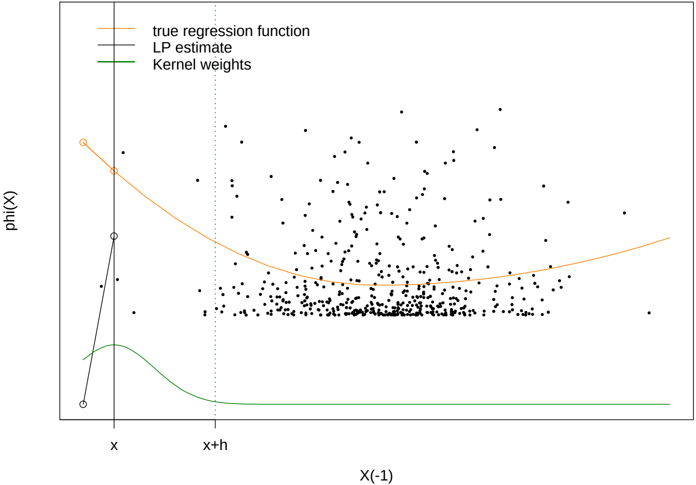
$$
\begin{aligned}
h_G^{opt} &= \arg\min_h \int AMSE\{\hat{\sigma}^2(x;h)\} f_X(x)dx \\
&= \left\{ \frac{C_{p,K} \times R_{\mathrm{var}}}{n \times R_{f,\mathrm{bias}}} \right\}^{1/(2p+3)}
\end{aligned}
$$

- The only unknown components (to estimate and plug-in) are:
  - $R_{\mathrm{var}} = \int v(x)dx$;
  - $R_{f,\mathrm{bias}} = \int d^2(x) f_X(x)dx$.

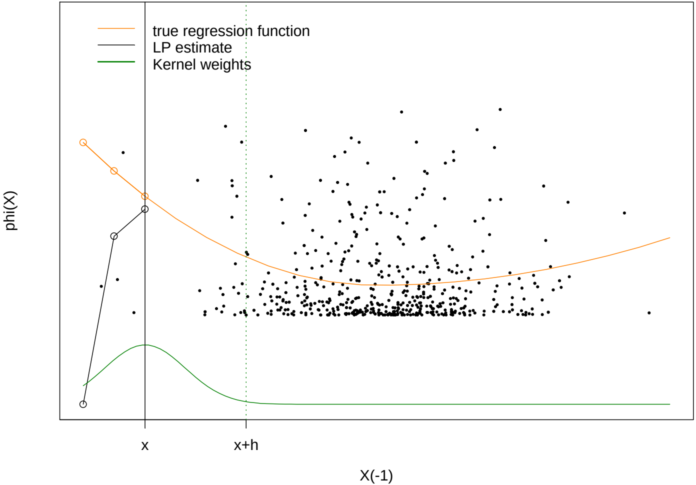
An illustrative example: global (fixed) bandwidth



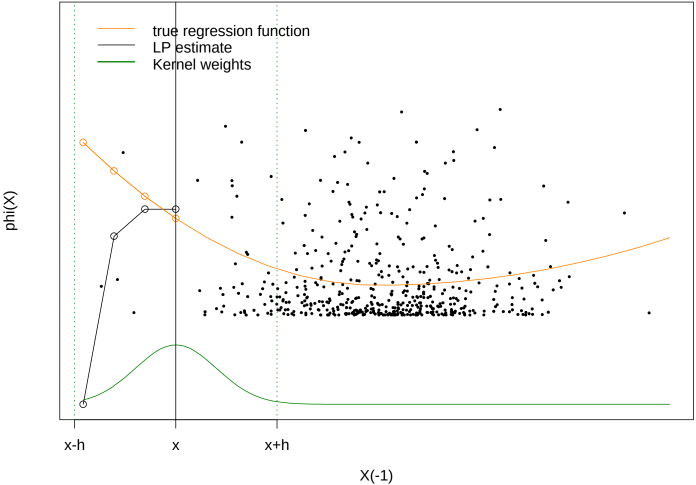
An illustrative example: global (fixed) bandwidth



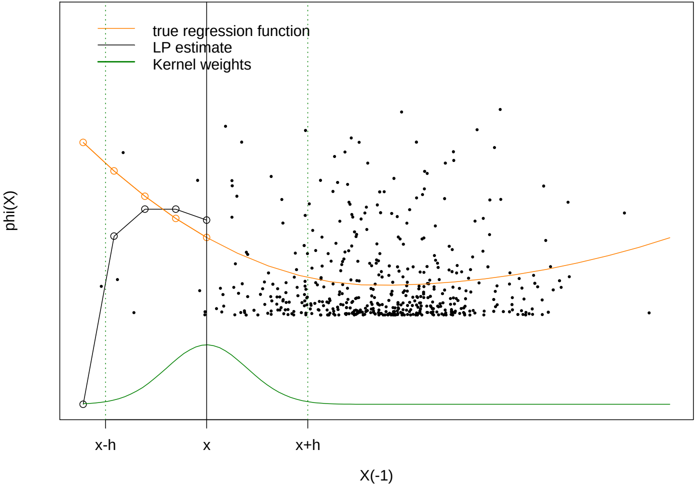
An illustrative example: global (fixed) bandwidth



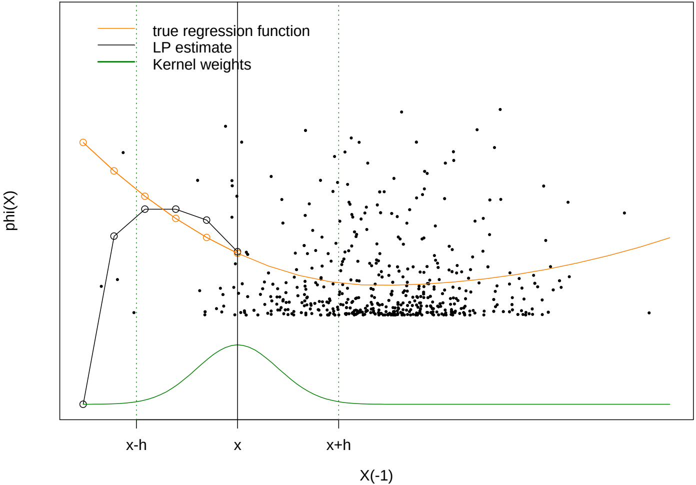
An illustrative example: global (fixed) bandwidth



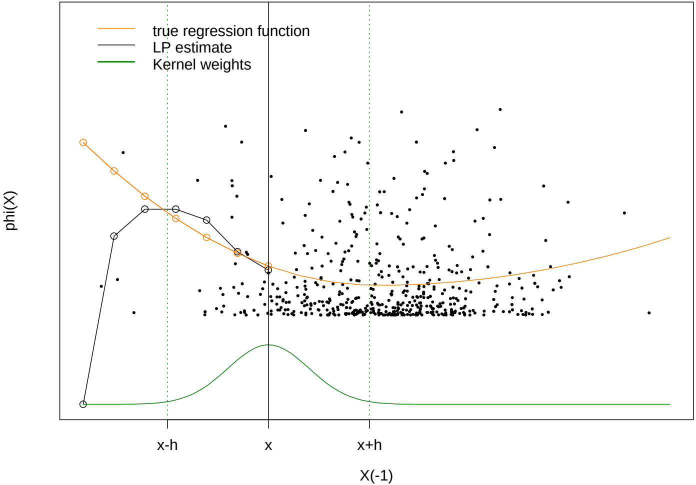
An illustrative example: global (fixed) bandwidth



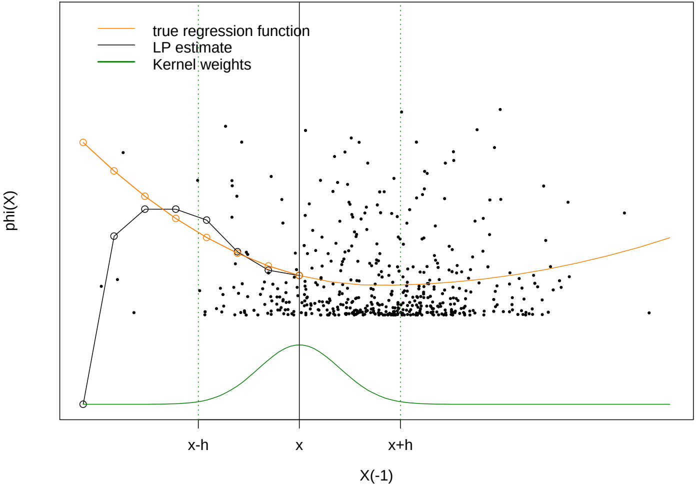
An illustrative example: global (fixed) bandwidth



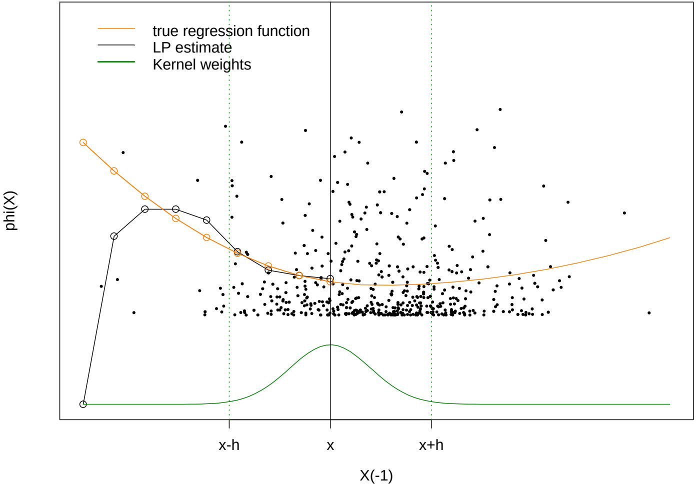
An illustrative example: global (fixed) bandwidth



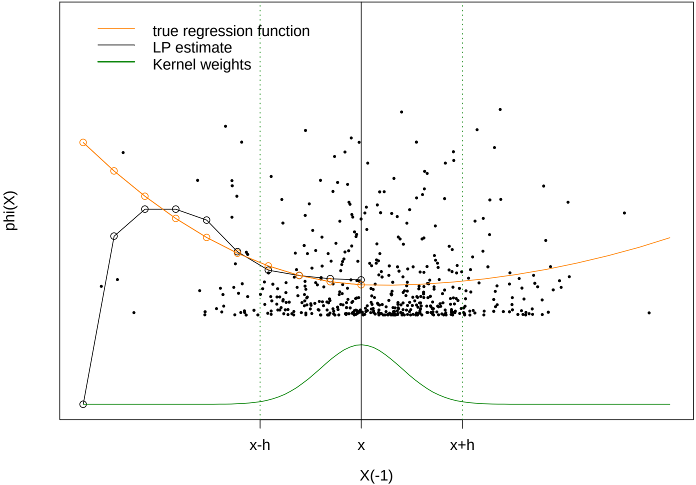
An illustrative example: global (fixed) bandwidth



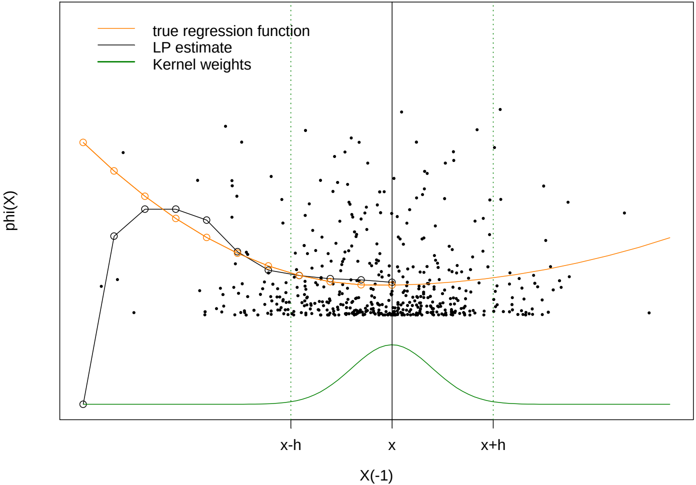
An illustrative example: global (fixed) bandwidth



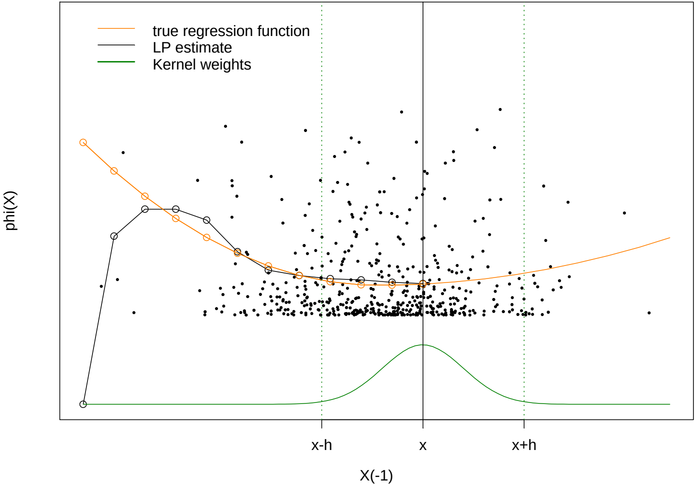
An illustrative example: global (fixed) bandwidth



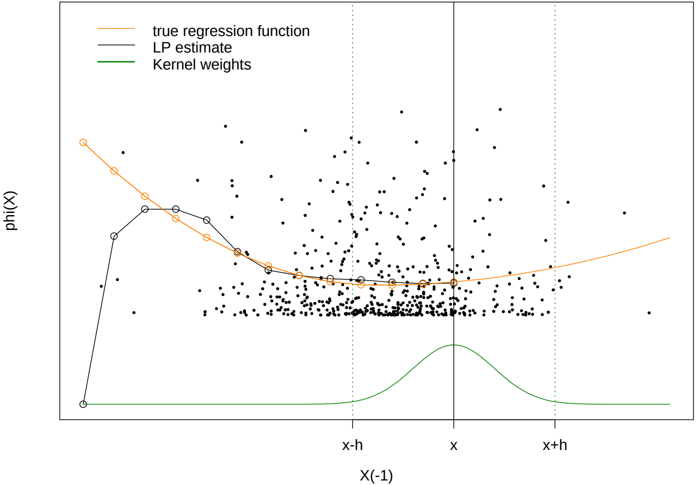
An illustrative example: global (fixed) bandwidth



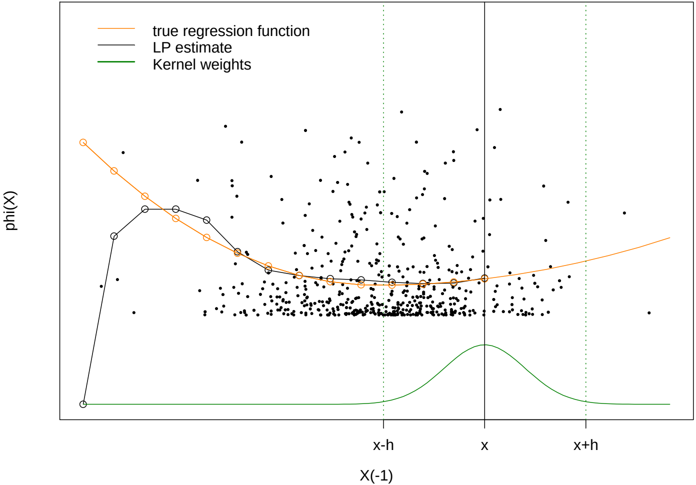
An illustrative example: global (fixed) bandwidth



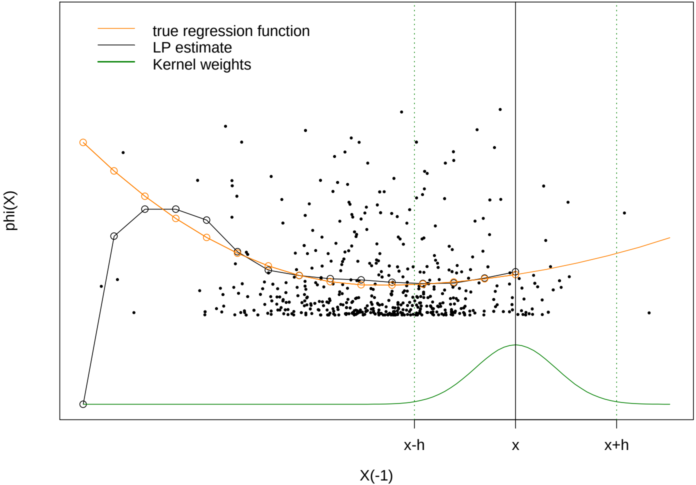
An illustrative example: global (fixed) bandwidth



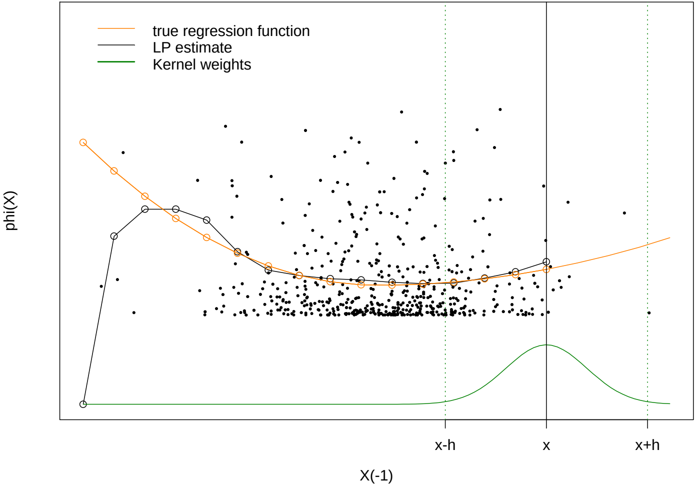
An illustrative example: global (fixed) bandwidth



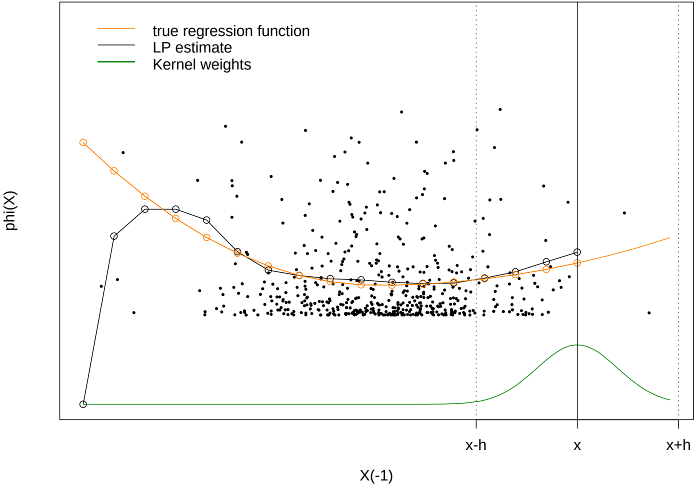
An illustrative example: global (fixed) bandwidth



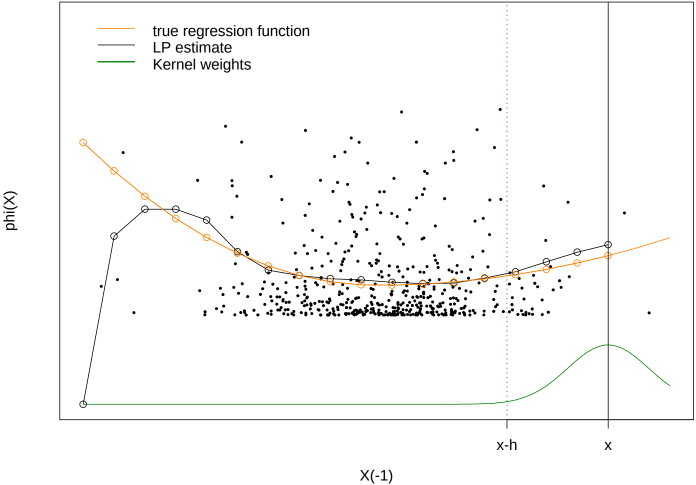
An illustrative example: global (fixed) bandwidth



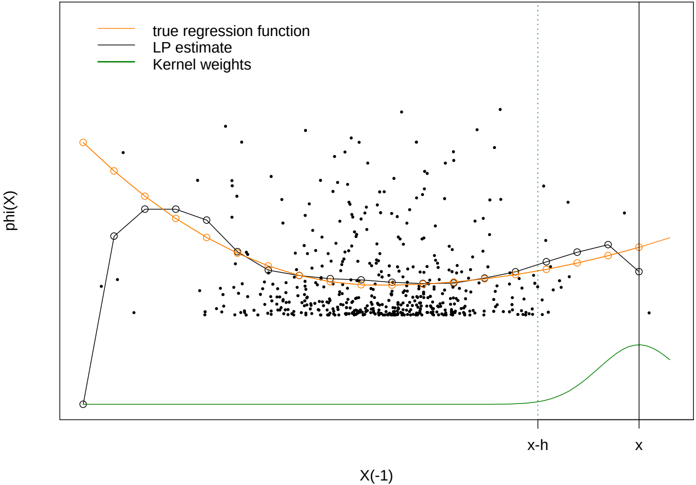
An illustrative example: global (fixed) bandwidth



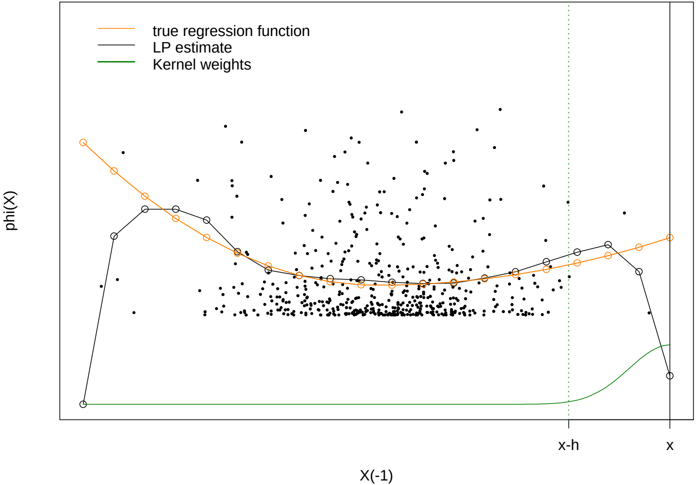
An illustrative example: global (fixed) bandwidth



An illustrative example: global (fixed) bandwidth



An illustrative example: global (fixed) bandwidth



Local vs global bandwidths

- The following relation holds:

$$AMSE\left\{x; h_L^{opt}(x)\right\} \leq AMSE\left\{x; h_G^{opt}\right\}.$$

Local vs global bandwidths

- The following relation holds:

$$AMSE\left\{x; h_L^{opt}(x)\right\} \leq AMSE\left\{x; h_G^{opt}\right\}.$$

- Anyway, the estimation of $h_L^{opt}(x)$ (=function) is less efficient than the estimation of h_G^{opt} (=mean value)

Local vs global bandwidths

- The following relation holds:

$$AMSE\left\{x; h_L^{opt}(x)\right\} \le AMSE\left\{x; h_G^{opt}\right\}.$$

- Anyway, the estimation of $h_L^{opt}(x)$ (=function) is less efficient than the estimation of h_G^{opt} (=mean value)
- **Some questions follow**:
  - what is the effective gain in using the local bandwidth?
  - Is it always convenient to perform a local smoothing instead of a global smoothing?
  - How to deal with pilot bandwidths?

Our proposal: a two stage procedure

Suppose we want to estimate the function $\sigma^2(x)$ on a support $I_X \subset \mathbb{R}$.

Our proposal: a two stage procedure

Suppose we want to estimate the function $\sigma^2(x)$ on a support $I_X \subset \mathbb{R}$.

1. **First stage: evaluating the "homogeneity" on the support**
   Given the estimates of $h_L^{opt}(x)$ and h_G^{opt}, use some relative indicator in order to evaluate the potential gain in using the local bandwidth instead of the global bandwidth on I_X.

Our proposal: a two stage procedure

Suppose we want to estimate the function $\sigma^2(x)$ on a support $I_X \subset \mathbb{R}$.

1. **First stage: evaluating the "homogeneity" on the support**
   Given the estimates of $h_L^{opt}(x)$ and h_G^{opt}, use some relative indicator in order to evaluate the potential gain in using the local bandwidth instead of the global bandwidth on I_X.

2. **Second stage: deriving the locally global bandwidth**
   Derive the optimal global bandwidths h_G^{opt} on "homogeneous" subsets of I_X.

Our proposal: a two stage procedure

Suppose we want to estimate the function $\sigma^2(x)$ on a support $I_X \subset \mathbb{R}$.

1. **First stage: evaluating the "homogeneity" on the support**
   Given the estimates of $h_L^{opt}(x)$ and h_G^{opt}, use some relative indicator in order to evaluate the potential gain in using the local bandwidth instead of the global bandwidth on I_X.

2. **Second stage: deriving the locally global bandwidth**
   Derive the optimal global bandwidths h_G^{opt} on "homogeneous" subsets of I_X.

- How to estimate h_G^{opt} and $h_L^{opt}(x)$ on I_X?
- Which relative indicator to use?
- How to smooth such bandwidths on I_X?

First stage: deriving the relative indicator

- Consider the relative increment of the AMSE, $\forall x \in I_X$,

$$\Delta_{AMSE}(x) = \frac{AMSE\{\hat{\sigma}^2(x; h_G^{opt})\} - AMSE\{\hat{\sigma}^2(x; h_L^{opt}(x))\}}{AMSE\{\hat{\sigma}^2(x; h_L^{opt}(x))\}}.$$

First stage: deriving the relative indicator

- Consider the relative increment of the AMSE, $\forall x \in I_X$,

$$\Delta_{AMSE}(x) = \frac{AMSE\{\hat{\sigma}^2(x; h_G^{opt})\} - AMSE\{\hat{\sigma}^2(x; h_L^{opt}(x))\}}{AMSE\{\hat{\sigma}^2(x; h_L^{opt}(x))\}}.$$

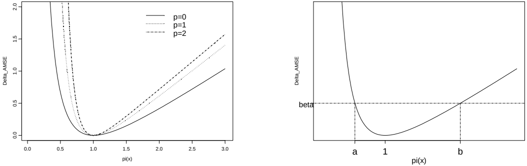
- We managed to express the $\Delta_{AMSE}(x)$ of the estimator in a **model free form:**

$$\Delta_{AMSE}(\pi_h(x)) = \frac{1}{2p+3}\left[\pi_h(x)\right]^{-2(p+1)} + \frac{2p+2}{2p+3}\left[\pi_h(x)\right] - 1$$

where $\pi_h(x) = \frac{h_L^{opt}(x)}{h_G^{opt}} \geq 0$.

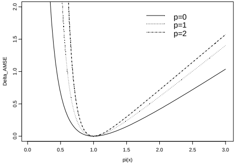
First stage: deriving the relative indicator

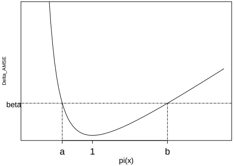
$$\Delta_{AMSE}(\pi_h(x)) = \frac{1}{2p+3}\left[\pi_h(x)\right]^{-2(p+1)} + \frac{2p+2}{2p+3}\left[\pi_h(x)\right] - 1$$



First stage: deriving the relative indicator

$$\Delta_{AMSE}(\pi_h(x)) = \frac{1}{2p+3}[\pi_h(x)]^{-2(p+1)} + \frac{2p+2}{2p+3}[\pi_h(x)] - 1$$



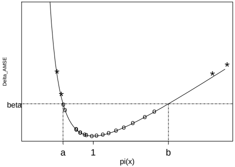


- **Remark 1**: as a function of $\pi_h(x)$, the relative indicator $\Delta_{AMSE}(x)$ is completely model free.
- **Remark 2** : $\Delta_{AMSE} < \beta$ *iif* $\pi_h(x) \in [a_\beta, b_\beta], \forall \beta > 0$.

A global measure of Δ_{AMSE} on the interval I_X

- A global measure of Δ_{AMSE} would indicate, eventually, the need to use on the interval I_X a local bandwidth.
- For example, **the mean value of Δ_{AMSE} on the interval** I_X could be considered...
$$\int_{I_X} \Delta_{AMSE}(x) f_X(x) dx.$$
- ... but we get some advantages if we consider **the median value of Δ_{AMSE} on the interval** I_X, or some other quantile.

The median of Δ_{AMSE} on the interval I_X



- For a fixed threshold β, consider the two set of points

$$\mathbb{S}_\beta = \{x : x \in I_X, \pi(x) \in [a_\beta, b_\beta]\}; \ \bar{\mathbb{S}}_\beta = \{x : x \in I_X, \pi(x) \notin [a_\beta, b_\beta]$$

- Given the measure of the process μ_X, the median of Δ_{AMSE} is the threshold value β^* for which $\mu_X(\mathbb{S}_{\beta^*}) = \mu_X(\bar{\mathbb{S}}_{\beta^*})$.

The algorithm of the first stage

1. Consider an estimation of both the global bandwidth h_G and the local bandwidth function $h_L(X_t)$, such as
   - the neural network method proposed in the second stage;
   - some other local method proposed in the literature.
2. Estimate $\pi_h(X_t)$, for all $X_t \in I_X$, then search for β^* such that

$$\sum_{X_t \in I_X} \mathbb{I}\left\{\hat{\pi}_h(X_t) \in [a_{\beta^*}, b_{\beta^*}]\right\} \approx \sum_{X_t \in I_X} \mathbb{I}\left\{\hat{\pi}_h(X_t) \notin [a_{\beta^*}, b_{\beta^*}]\right\}.$$

   where $\mathbb{I}(\cdot)$ is the indicator function.
3. Split the subset I_X if $\beta^* > \tau$, where τ represents the max relative increment tolerated for Δ_{AMSE} when using a global bandwidth.

Second stage: the locally global bandwidth

- Given the “homogeneous” subset $I_X \subset \mathbb{R}$, we want to estimate

$$h_G^{opt} = \left\{ \frac{C_{p,K} \times R_{\text{var}}^{I_X}}{n \times R_{f,\text{bias}}^{I_X}} \right\}^{1/(2p+3)}$$

  where:

  - $R_{\text{var}}^{I_X} = \int_{I_X} v(x) d\omega_{I_X}$;
  - $R_{f,\text{bias}}^{I_X} = \int_{I_X} d^2(x) f_X(x) d\omega_{I_X}$.

1. **How to derive the measure ω_{I_X}?** (we used $d\omega_{I_X} = \frac{dx}{\mu(I_X)}$)
2. **How to estimate $v(x)$ and $d^2(x)$ on I_X?**

Estimating the functions $v(x)$ and $d^2(x)$

- Note that:

$$
\begin{aligned}
v(x) &= Var\{X_t^2 | X_{t-1} = x\} \equiv m_4(x) - m_2^2(x) \\
d(x) &= \frac{\partial^{p+1}\sigma^2(x)}{\partial x^{p+1}} \equiv m_2^{(p+1)}(x)
\end{aligned}
$$

Estimating the functions $v(x)$ and $d^2(x)$

- Note that:

$$\begin{aligned} v(x) &= Var\{X_t^2 | X_{t-1} = x\} \equiv m_4(x) - m_2^2(x) \\ d(x) &= \frac{\partial^{p+1} \sigma^2(x)}{\partial x^{p+1}} \equiv m_2^{(p+1)}(x) \end{aligned}$$

- As usual, we should have to estimate separately the three functions

$$m_4(x), \quad m_2(x), \quad m_2^{(p+1)}(x)$$

(Fan & Yao, 1998; Franke & Diagne, 2006)

Estimating the functions $v(x)$ and $d^2(x)$

- Note that:

$$
\begin{aligned}
v(x) &= Var\{X_t^2 | X_{t-1} = x\} \equiv m_4(x) - m_2^2(x) \\
d(x) &= \frac{\partial^{p+1}\sigma^2(x)}{\partial x^{p+1}} \equiv m_2^{(p+1)}(x)
\end{aligned}
$$

- As usual, we should have to estimate separately the three functions

$$
m_4(x), \quad m_2(x), \quad m_2^{(p+1)}(x)
$$

(Fan & Yao, 1998; Franke & Diagne, 2006)

- **An interesting result here is that we estimate the three functions by only one estimator!**

Estimating the functions $v(x)$ and $d^2(x)$

- We use only the following neural network estimator

$$\hat{\boldsymbol{\eta}} = \arg\min_{\boldsymbol{\eta}} \sum_{t=2}^{n} \left[X_t^2 - q(X_{t-1}; \boldsymbol{\eta}) \right]^2$$

Estimating the functions $v(x)$ and $d^2(x)$

- We use only the following neural network estimator

$$\hat{\eta} = \arg\min_{\boldsymbol{\eta}} \sum_{t=2}^{n} \left[X_t^2 - q(X_{t-1};\boldsymbol{\eta})\right]^2$$

where

  - $q(X_{t-1};\boldsymbol{\eta}) = \sum_{k=1}^{d} c_k \Gamma\left(a_k X_{t-1} + b_k\right) + c_0$ is the neural network function;
  - $\boldsymbol{\eta} = (c_0, c_1, \ldots, c_d, a_1, \ldots, a_d, b_1 \ldots b_d)$ is the vector of parameters to be estimated;
  - d is the number of nodes in the hidden layer;
  - $\Gamma(\cdot)$ is the *logistic activation function*.

Estimating the functions $v(x)$ and $d^2(x)$

- By defining $m_{4\varepsilon} = E(\epsilon_t^4)$, we reparameterize as follows:

$$\begin{aligned} v(x) &= m_4(x) - m_2^2(x) = m_2^2(x)\,[m_{4\varepsilon} - 1] \\ d(x) &= m_2^{(p+1)}(x) \end{aligned}$$

- then we propose the following estimators

$$\begin{aligned} \hat{m}_2(x) &= q(x, \hat{\eta}) \\ \hat{m}_{4\varepsilon} &= \frac{\sum_{t=2}^n X_t^4}{\sum_{t=2}^n [q(X_t, \hat{\eta})]^2} \\ \hat{m}_2^{(p+1)}(x) &= q^{(p+1)}(x; \hat{\eta}) \end{aligned}$$

Estimating the locally global bandwidth

- Finally, given the ergodicity of the process, the locally global bandwidth is estimated by:

$$\hat{R}_{\text{var}}^{I_X} = \frac{\sum_{i=1}^{n^*} \hat{v}(x_i)}{n}. \qquad \hat{R}_{f,\text{bias}}^{I_X} = \frac{\sum_{X_t \in I_X} \left[\hat{m}_2^{(p+1)}(X_t)\right]^2}{\sum_{t=1}^{n} \mathbb{I}(X_t \in I_X)},$$

$$\hat{h}_{I_X} = \left\{ \frac{C_{p,K} \times \hat{R}_{\text{var}}^{I_X}}{n \times \hat{R}_{f,\text{bias}}^{I_X}} \right\}^{1/(2p+3)}$$

- The points $\{x_1, x_2, \ldots, x_{n^*}\}$ in $\hat{R}_{\text{var}}^{I_X}$ are uniformly spaced in the interval I_X, and $\mathbb{I}(\cdot)$ is the indicator function.

Assumptions

(a1) the errors ε_t have continuous and positive density function and, for some $\delta > 4$,

$$E(\varepsilon_t^2) = 1, \quad E(\varepsilon_t) = E(\varepsilon_t^3) = 0, \quad E|\varepsilon_t|^\delta < \infty;$$

(a2) the functions $m(\cdot)$ and $\sigma(\cdot)$ have continuous second derivative. Moreover, the function $\sigma(\cdot)$ is positive;

(a3) there exist the constants $M_1 > 0$ and $M_2 > 0$ such that, for all $y \in \mathbb{R}$,

$$|m(y)| \leq M_1(1 + |y|), \ |\sigma(y)| \leq M_2(1 + |y|), \ M_1 + M_2 \left[E|\varepsilon_t|^\delta\right]^{1/\delta} < 1;$$

(a4) the density function $f_X(\cdot)$ of the (stationary) measure of the process μ_X exists, it is bounded, continuous and positive on every compact set in $\mathbb{R}$.

(b1) $\sum_{k=1}^{d} |c_k| \leq \Delta_n$.

(b2) $d \equiv d_n$, $d_n \to \infty$, $\Delta_n \to \infty$ as $n \to \infty$.

(b3) Let $K_1(n) := \frac{\Delta_n^2 d_n \log\left(\Delta_n^2 d_n\right)}{\sqrt{n}}$ and $K_2(n) := \frac{\Delta_n^4}{n^{1-\delta}}$.

The consistency of the bandwidth selector

Using the framework in Franke and Diagne (2006), we can state the following results.

Lemma 1 Under assumptions (a1)-(a4) and (b1)-(b3), the volatility function estimator is consistent in the sense that:

- If $K_1(n) \to 0$ as $n \to \infty$, then

$$E \int \left(q\left(x; \hat{\eta}\right) - \sigma^2(x) \right)^2 d\mu_X(x) \to 0 \qquad n \to \infty$$

- if, additionally, $K_2(n) \to 0$ for some $\delta > 0$, then

$$\int \left(q\left(x; \hat{\eta}\right) - \sigma^2(x) \right)^2 d\mu_X(x) \xrightarrow{a.s.} 0 \qquad n \to \infty$$

The consistency of the bandwidth selector

Lemma 2 Under assumptions (a1)-(a4) and (b1)-(b3), the estimator of the second derivative of $\sigma^2(x)$ is consistent in the sense that:

- If $K_1(n) \to 0$ as $n \to \infty$, then

$$E \int \left(q''(x; \hat{\eta}) - [\sigma^2(x)]'' \right)^2 d\mu_X(x) \to 0 \qquad n \to \infty$$

- if, additionally, $K_2(n) \to 0$ for some $\delta > 0$, then

$$\int \left(q''(x; \hat{\eta}) - [\sigma^2(x)]'' \right)^2 d\mu_X(x) \xrightarrow{a.s.} 0 \qquad n \to \infty$$

The consistency of the bandwidth selector

Theorem 1 Under the conditions (a1)-(a4) and (b1)-(b3), $\hat{R}^{I_X}_{f,\text{bias}}$, with $I_X \subseteq \mathbb{R}$, is consistent in the sense that:

- If $K_1(n) \to 0$ as $n \to \infty$, then

$$\hat{R}^{I_X}_{f,\text{bias}} \xrightarrow{p} R^{I_X}_{f,\text{bias}} \quad n \to \infty$$

- if, additionally, $K_2(n) \to 0$ for some $\delta > 0$, then

$$\hat{R}^{I_X}_{f,\text{bias}} \xrightarrow{a.s.} R^{I_X}_{f,\text{bias}} \quad n \to \infty$$

Corollary Using the same conditions as in theorem 1, then $\hat{m}_{4\varepsilon}$ is consistent in the sense that:

- If $K_1(n) \to 0$ as $n \to \infty$, then $\hat{m}_{4\varepsilon} \xrightarrow{p} m_{4\varepsilon} \quad n \to \infty$
- if, additionally, $K_2(n) \to 0$ for some $\delta > 0$, then $\hat{m}_{4\varepsilon} \xrightarrow{a.s.} m_{4\varepsilon} \quad n \to \infty$

The consistency of the bandwidth selector

Theorem 2 Using the same conditions as in theorem 1, then $\hat{R}_{\mathrm{var}}^{I_X}$ with $I_X \subset \mathbb{R}$ and $n^* = O(n)$, is consistent in the sense that:

- If $K_1(n) \to 0$ as $n \to \infty$, then

$$\hat{R}_{\mathrm{var}}^{I_X} \xrightarrow{p} R_{\mathrm{var}}^{I_X} \quad n \to \infty$$

- if, additionally, $K_2(n) \to 0$ for some $\delta > 0$, then

$$\hat{R}_{\mathrm{var}}^{I_X} \xrightarrow{a.s.} R_{\mathrm{var}}^{I_X} \quad n \to \infty$$

The consistency of the bandwidth selector

- For a fixed $z \in \mathbb{R}$, let I_z be a non null measure set which contains the point z.
- Let n_z be the number of observed values in I_z such that $n_z \to \infty$ when $n \to \infty$.

Theorem 3 Using the same conditions as in Lemma 2, if $n_z = o(n)$, for a set $I_X \subset \mathbb{R}$, then $\hat{h}_{I_z}$ is consistent in the sense that:

- If $K_1(n) \to 0$ as $n \to \infty$, then

$$\sup_{z \in I_X} \left| \hat{h}_{I_z} - h_L^{opt}(z) \right| \xrightarrow{p} 0 \quad n \to \infty$$

- if, additionally, $K_2(n) \to 0$ for some $\delta > 0$, then

$$\sup_{z \in I_X} \left| \hat{h}_{I_z} - h_L^{opt}(z) \right| \xrightarrow{a.s.} 0 \quad n \to \infty$$

The consistency of the bandwidth selector

The previous results imply that:

1. $$\hat{h}_G \xrightarrow{p} h_G^{opt} \quad \text{or} \quad \hat{h}_G \xrightarrow{a.s.} h_G^{opt}.$$

2. For all $z \in I_X$

$$\hat{h}_{I_z} \xrightarrow{p} h_L^{opt}(z) \quad \text{or} \quad \hat{h}_{I_z} \xrightarrow{a.s.} h_L^{opt}(z),$$

3. Uniformly for each $z \in I_X$

$$\hat{\pi}_h(z) \xrightarrow{p} \pi_h(z) \quad \text{or} \quad \hat{\pi}_h(z) \xrightarrow{a.s.} \pi_h(z)$$

The simulation study: setting the models

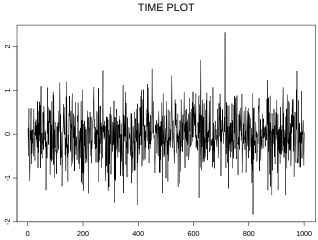
Model 1: $X_t = [\psi(X_{t-1} + 1.2) + 1.5\psi(X_{t-1} - 1.2)]\,\epsilon_t$ $\epsilon_t \sim N(0,1)$
$\psi(z)$= d.f. of $N(0,1)$

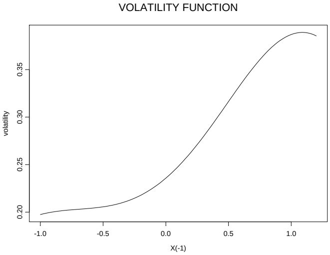
Model 2: $X_t = \sqrt{0.1 + 0.3X_{t-1}^2}\epsilon_t$ $\epsilon_t \sim N(0,1)$

Model 3: $X_t = \sqrt{0.01 + 0.1X_{t-1}^2 + 0.2X_{t-1}^2 I_{X_{t-1}<0}}\epsilon_t$ $\epsilon_t \sim N(0,1)$

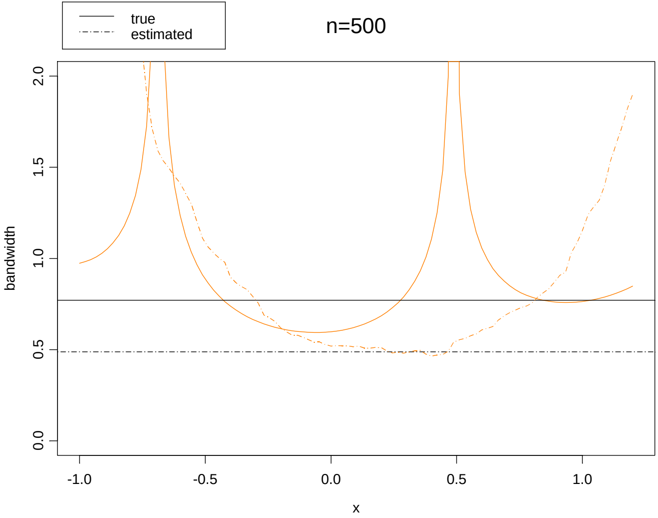
Model 1: from Härdle and Tsybakov (1997)

$$X_t = [\psi(X_{t-1} + 1.2) + 1.5\psi(X_{t-1} - 1.2)]\,\epsilon_t; \qquad \begin{array}{rcl} \epsilon_t & \sim & N(0,1); \\ \psi(z) & = & \text{d.f. of } N(0,1) \end{array}$$

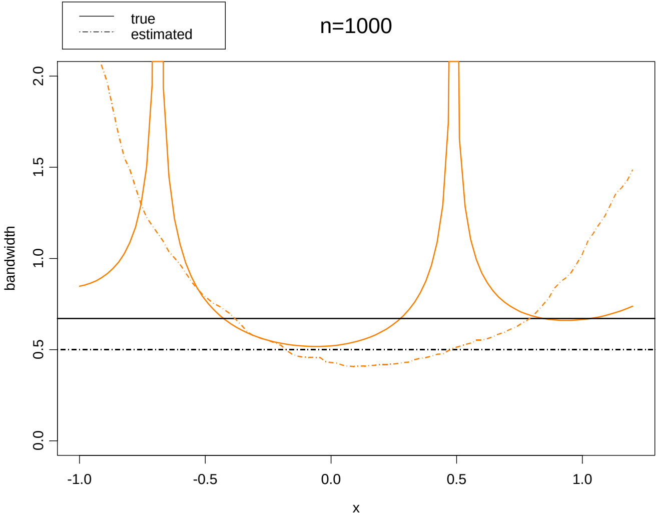




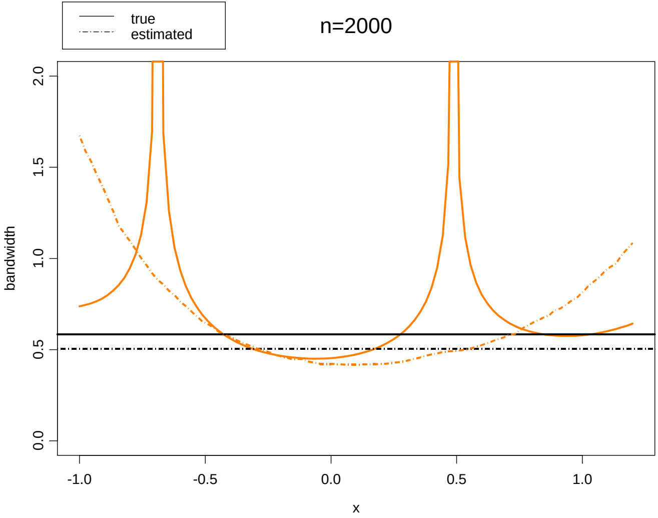
Model 1: Estimated local and global bandwidths



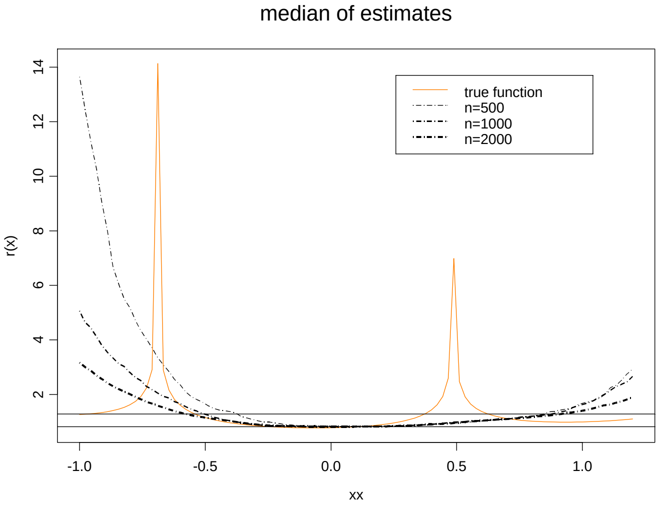
Model 1: Estimated local and global bandwidths



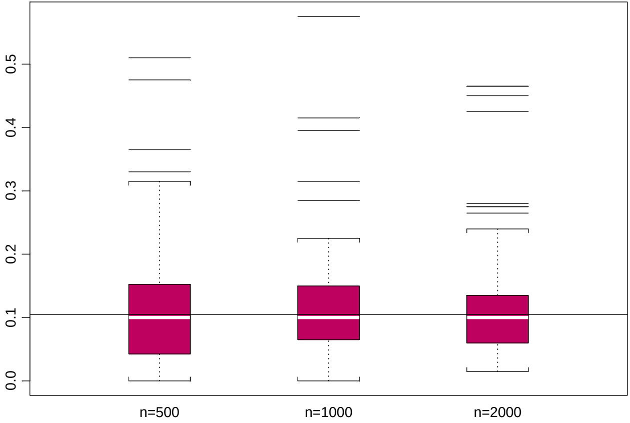
Model 1: Estimated local and global bandwidths



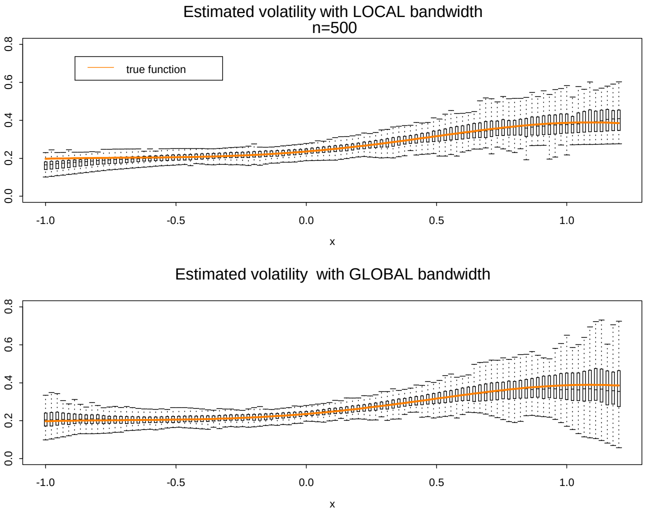
Model 1: $\pi_h(x)$ estimated functions



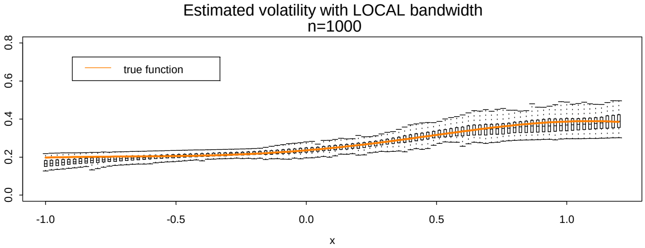
Model 1: median of $\Delta(AMSE)$ estimates

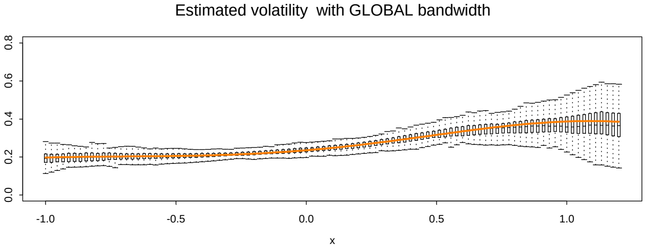


Model 1: LPE estimates of the volatility function

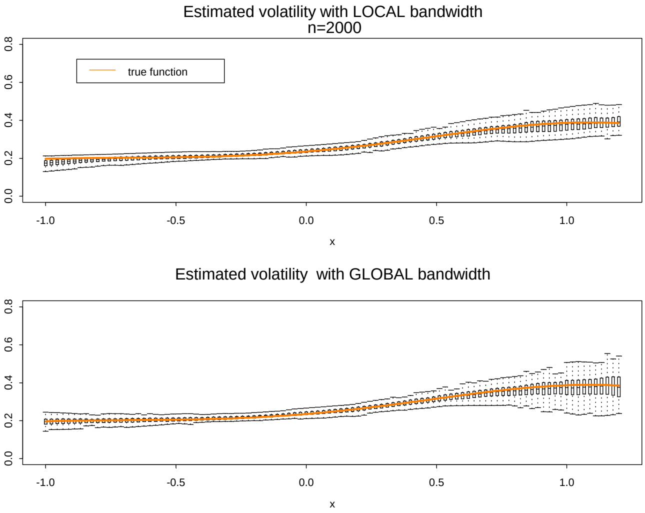


Model 1: LPE estimates of the volatility function





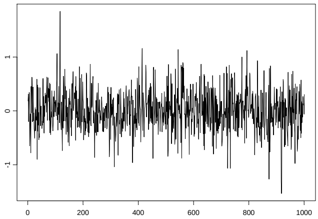
Model 1: LPE estimates of the volatility function



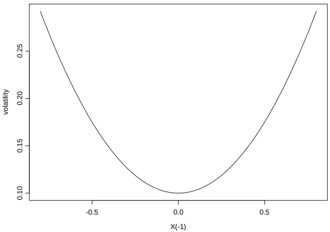
Model 2: ARCH(1)

$$X_t = \sqrt{0.1 + 0.3X_{t-1}^2}\,\epsilon_t; \qquad \epsilon_t \sim N(0,1)$$

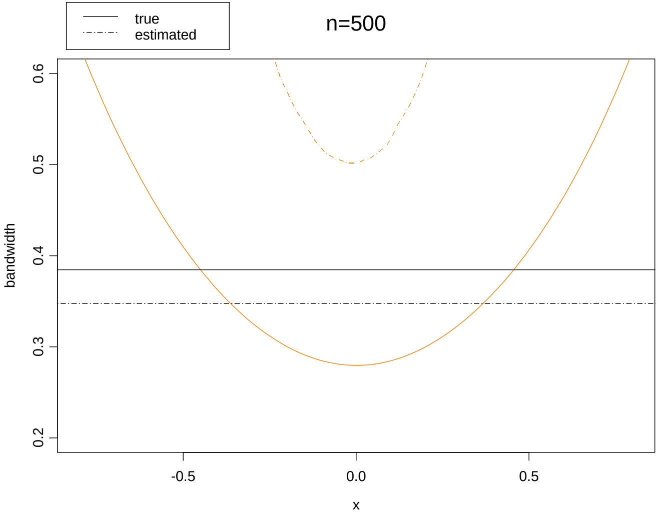
TIME PLOT



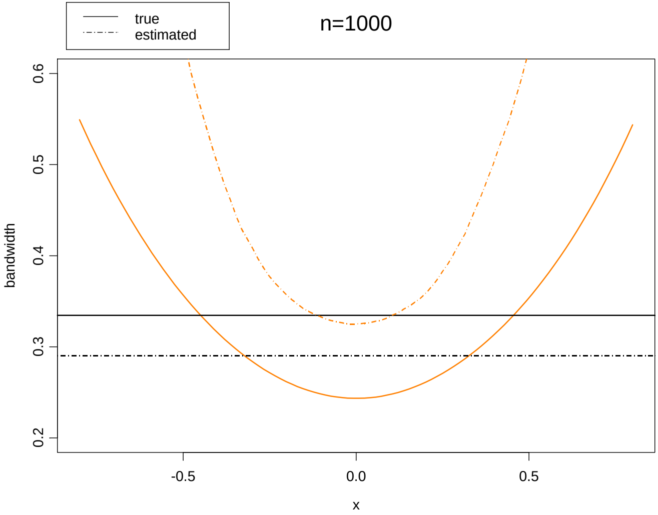
VOLATILITY FUNCTION



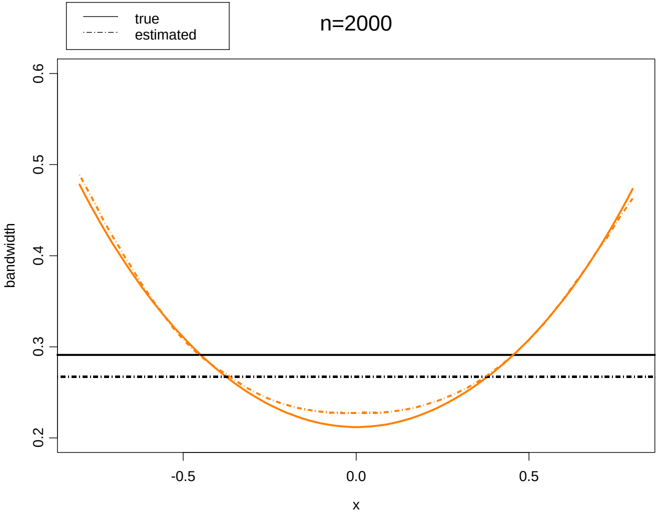
Model 2: Estimated local and global bandwidths



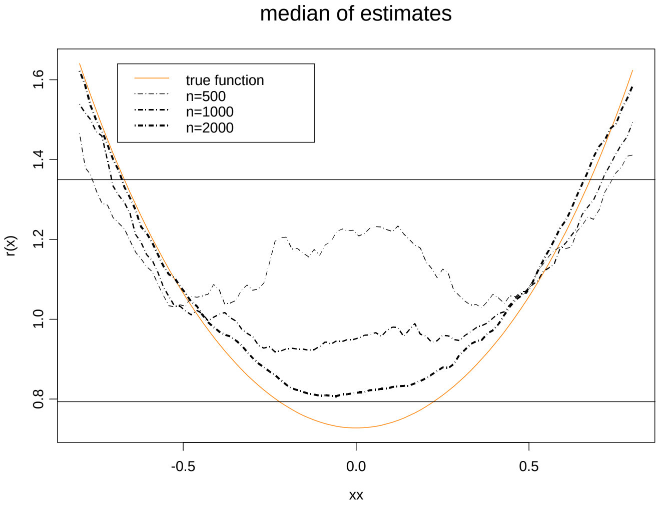
Model 2: Estimated local and global bandwidths



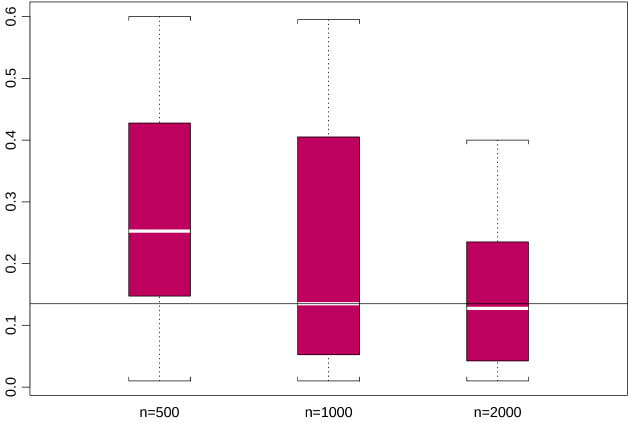
Model 2: Estimated local and global bandwidths



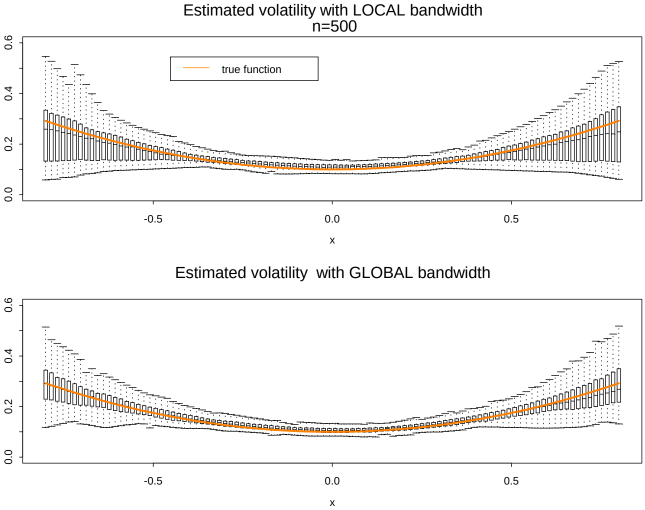
Model 2: $\pi_h(x)$ estimated functions



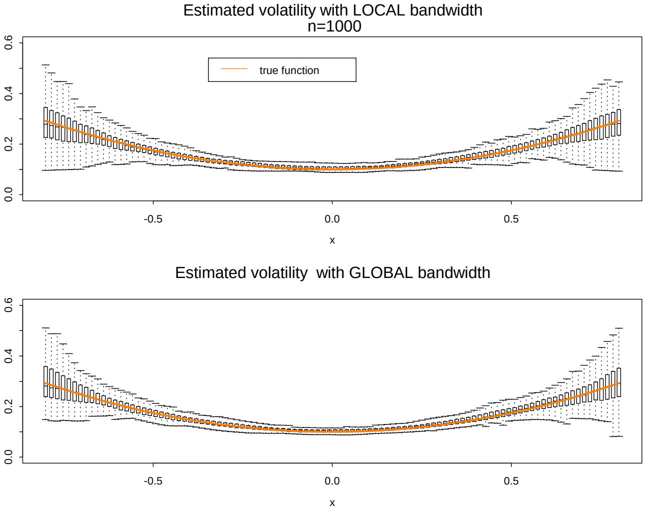
Model 2: median of $\Delta(AMSE)$ estimates



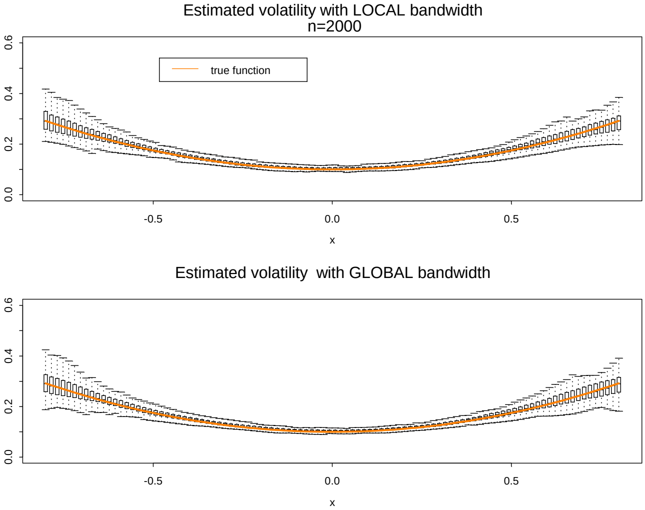
Model 2: LPE estimates of the volatility function



Model 2: LPE estimates of the volatility function

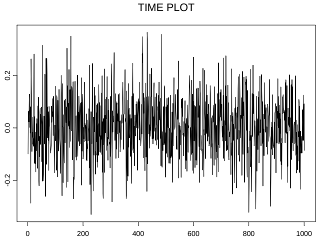


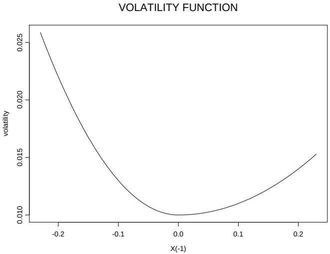
Model 2: LPE estimates of the volatility function



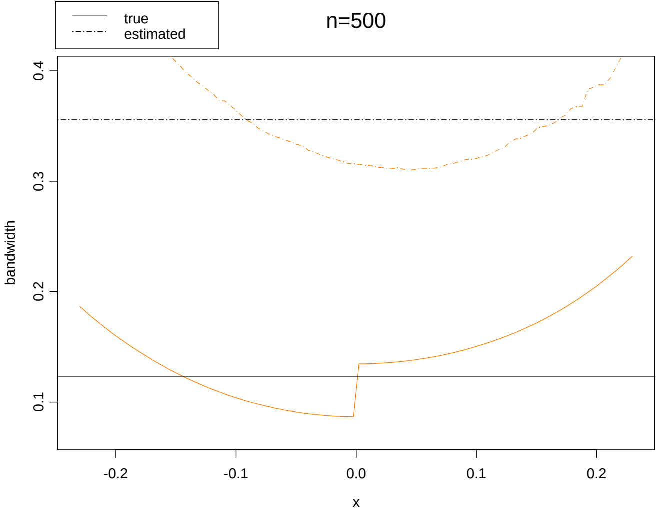
Model 3: from Franke and Diagne (2006)

$$X_t = \sqrt{0.01 + 0.1X_{t-1}^2 + 0.2X_{t-1}^2 I_{X_{t-1}<0}}\epsilon_t; \qquad \epsilon_t \sim N(0,1)$$

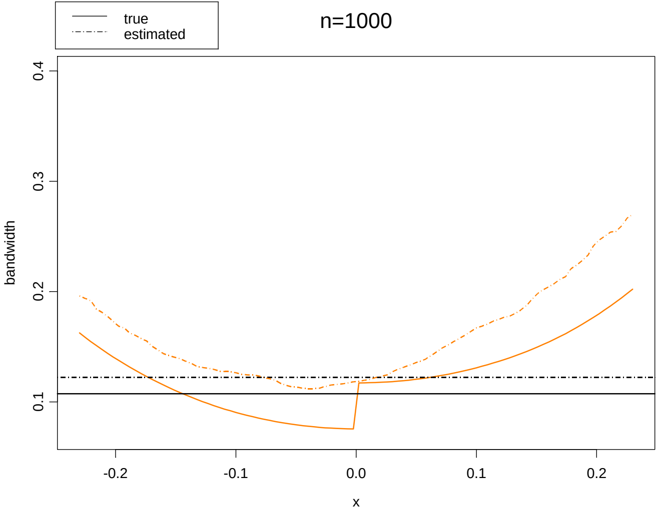




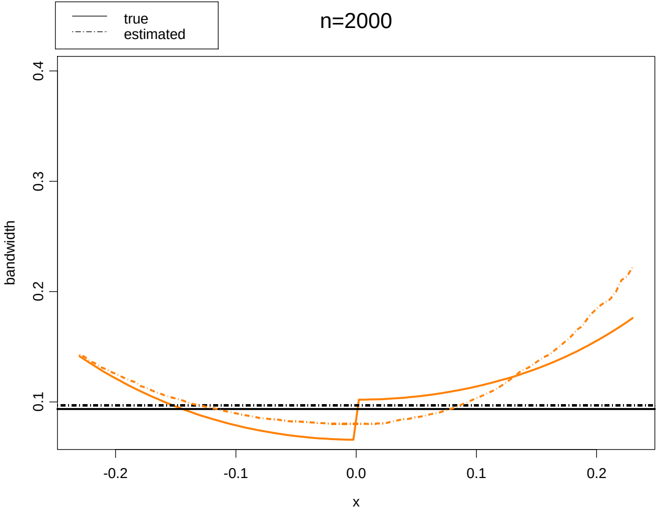
Model 3: Estimated local and global bandwidths



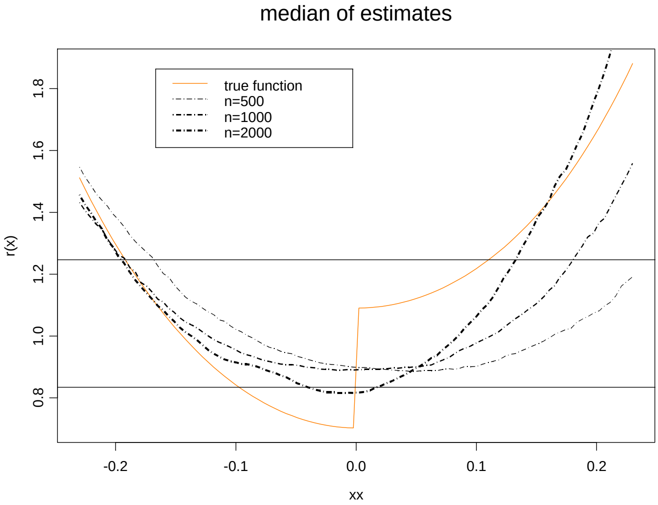
Model 3: Estimated local and global bandwidths



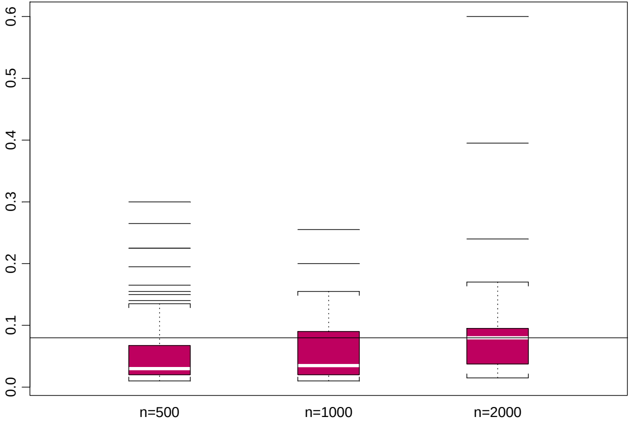
Model 3: Estimated local and global bandwidths



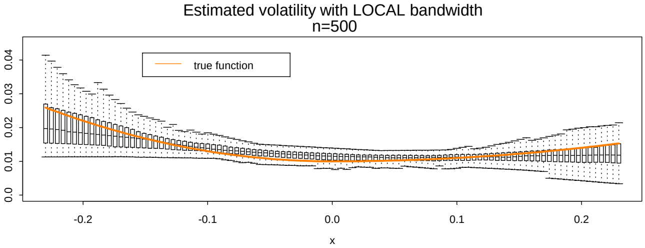
Model 3: $\pi_h(x)$ estimated function

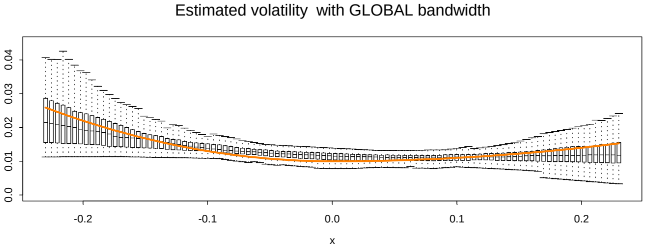


Model 3: median of $\Delta(AMSE)$ estimates

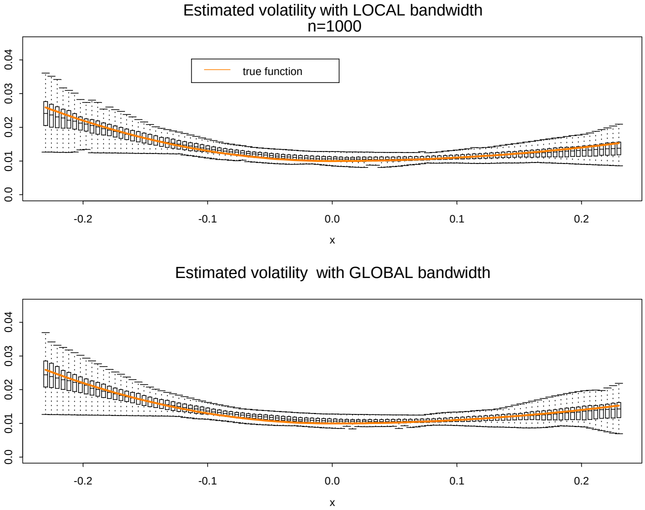


Model 3: LPE estimates of the volatility function

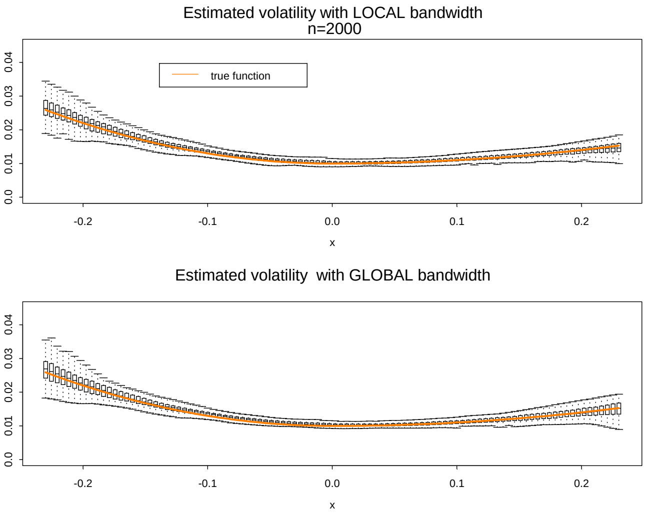




Model 3: LPE estimates of the volatility function



Model 3: LPE estimates of the volatility function



Conclusions

- From the simulation study, it seems that the use of the local bandwidth sometimes does not produce better results.
- A global bandwidth derived on a suitable subset perform as well as the local bandwidth.
- Given a compact subset I_z, we derived a consistent estimator of the *local bandwidth*.

- **Further research (under development)**:
  1. improving the estimation of the derivative function $m_2^{(p+1)}$;
  2. analysing the different orders of the two bandwidth estimators (local and global);
  3. identifying some diagnostic tools useful for the choice of the most suitable type of smoothing;
  4. extending the results to the multivariate framework.