SCHOOL OF ECONOMICS AND MANAGEMENT
FACULTY OF SOCIAL SCIENCES
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2009-48

Unstable volatility functions: the break preserving local linear estimator

Isabel Casas and Irene Gijbels

School of Economics and Management
Aarhus University
Bartholins Allé 10, Building 1322, DK-8000 Aarhus C
Denmark

Unstable volatility functions: the break preserving local linear estimator*

ISABEL CASAS†

Aarhus University

IRENE GIJBELS

Katholieke Universiteit Leuven

Abstract

The objective of this paper is to introduce the break preserving local linear (BPLL) estimator for the estimation of unstable volatility functions. Breaks in the structure of the conditional mean and/or the volatility functions are common in Finance. Markov switching models (Hamilton, 1989) and threshold models (Lin and Teräsvirta, 1994) are amongst the most popular models to describe the behaviour of data with structural breaks. The local linear (LL) estimator is not consistent at points where the volatility function has a break and it may even report negative values for finite samples. The estimator presented in this paper generalises the classical LL. The BPLL maintains the desirable properties of the LL with regard to the bias and the boundary estimation, it estimates the breaks consistently and it ensures that the volatility estimates are always positive.

Keywords: Breaks estimation; Heteroscedasticity; Local linear regression; Nonlinear time series; Volatility estimation.

JEL Classification: C13; C14; C22.

*The first author is very grateful to Robinson Kruse, Christos Ntantamis and Almut Veraart for the long discussions and valuable suggestions. I. Casas is also grateful for financial support to the Spanish Ministry of Education and Science (Jose Castillejo program) and the Danish Social Sciences Research Council (grant no. FSE 275-08-0078). The second author gratefully acknowledges support from the Research Fund of the KU Leuven (GOA/07/04-project) and from the Federal Science Policy, Belgium (IAP research network nr P6/03).

†Please address correspondence to: Isabel Casas, CREATES, Aarhus University, Universitetsparken building 1322, 8000 Aarhus, Denmark; e-mail: icasas@creates.au.dk.

1 Introduction

The authors are interested in estimating the volatility function at a given point x, which is denoted by $\sigma^2(x)$ hereafter. This volatility function is non–stochastic, in the sense that the dynamics in model (1) below come from the process and not from the behaviour of the volatility over the dimension t. In addition, the conditional mean and volatility functions may be discontinuous i.e. they may present a finite number of breaks which may be of different magnitudes. Breaks are abrupt changes in the structure of the function due to sudden events. These breaks may be caused by a financial crisis (see Cerra and Saxena, 2005 and Hamilton, 2005), or an abrupt change in the government policy (Hamilton, 1988 amongst others).

Let $\{(Y_t, X_t)\}$ be a two–dimensional strictly stationary process, distributed as (Y, X) and defined by the model:

$$Y_t = m(X_t) + \sigma(X_t)\epsilon_t, \tag{1}$$

where the conditional mean function $m(x)$ is also non–stochastic and it possibly has breaks. The innovations ϵ_t, $t = 1, \ldots, n$, are distributed as ϵ and satisfy $E(\epsilon|X) = 0$ and $E(\epsilon^2|X) = 1$. Clearly, $E(Y_t|X_t = x) = m(x)$ and $E(r_t^2|X_t = x) = \sigma^2(x)$ for $r_t = Y_t - m(X_t)$. Thus, it is natural to estimate the conditional mean and the volatility with a regression of Y_t and r_t^2 respectively. The conditional mean function is estimated using the technique in Gijbels *et al.* (2007) which combines the smooth preserving properties of the classical LL (Fan and Gijbels, 1996) and the break preserving properties of the estimator in Qiu (2003). Their conditional mean estimator is uniformly consistent in the points of continuity and pointwise consistent at the breaks, with bias of $O_p(h_1^2)$ for the bandwidth $h_1 > 0$. In their study, three estimators of the conditional mean are found at each point x: i) the LL estimator, $\hat{a}_c(x)$; ii) the right estimator, $\hat{a}_r(x)$; and iii) the left estimator $\hat{a}_l(x)$ whose mathematical expressions are given in the general formula,

$$\hat{a}_k(x) = \sum_{t=1}^{n} Y_t K_k\left(\frac{X_t - x}{h_1}\right)\frac{s_{k,2} - s_{k,1}(X_t - x)}{s_{k,0}s_{k,2} - s_{k,1}^2} \quad \text{for } k = c, l, r \tag{2}$$

where $s_{k,j} = \sum(X_t - x)^j K_k\left(\dfrac{X_t - x}{h_1}\right)$ and $K_c(\cdot)$ is a symmetric kernel centred around zero with support on $[-\frac{1}{2}, \frac{1}{2}]$. The other two kernels are the right and left kernels defined as $K_l(x) = K_r(-x) = K_c(x)$ when $x \in [-\frac{1}{2}, 0]$ and zero otherwise. The estimator $\hat{m}(x)$

is the appropriate value chosen amongst those three estimators.

Fan and Yao (1998) show that $\sigma^2(\cdot)$ is consistently estimated with the LL estimator over the series of estimated squared residuals $\hat{r}_t^2 = (Y_t - \hat{m}(X_t))^2$. In fact, although the bias of $\hat{m}(x)$ is of order $O_p(h_1^2)$, "its contribution to $\hat{\sigma}^2(x)$ is only of $o_p(h_1^2)$". However, the LL estimator is not consistent at break points. Intuitively, any estimator which uses a centred kernel is expected to lie in the middle of the two values at the break, independently of the data size n, and therefore it is inconsistent. The BPLL estimator uses the LL for the continuous parts, and chooses between the left or right estimators at the breaks neighbourhoods. The LL is asymptotically positive but this is not always true for finite samples. This paper suggests using the exponential local linear (ELL) estimator (see Ziegelmann, 2002) when negative values occur.

The contribution of the BPLL to the existing literature is at different levels. The main contribution of this paper is that the BPLL estimates volatility functions with breaks. Second, the parametric forms of the conditional mean and volatility functions are unknown, although they satisfy certain regularity conditions which are detailed in Appendix A. Third, the BPLL estimates the breaks consistently, in contrast to other existing kernel smoothing estimators of the volatility, and it ensures that the volatility estimates are always positive. Fourth, contrary to parametric structural break models, the BPLL is independent of the nature of the breaks. Finally, the BPLL is a one step estimator while popular structural break models find the location of the break first and then proceed to the estimation.

The paper is organised as follows. In Section 2 the left, right and centred estimators for the volatility and their statistical inference are introduced. Section 3 introduces the BPLL as a combination of the three estimators and the asymptotic theory. The results of Monte Carlo experiments are shown in Section 4 where the BPLL performance is compared to the LL performance. The conclusions are laid out in Section 5 and the proofs are left to the Appendixes.

2 The left, right and centred estimators

First, let the variable X have support in $[a, b]$ and $\{x_q : q = 1, \ldots, m\}$ be the finite set of breaks which, for simplicity, do not correspond with boundary points. Defining $h_2 > 0$ as the bandwidth, then two regions are differentiated: i) D_1 is the region where the volatility function is continuous, and ii) D_2 contains the breaks and their neighbourhoods:

$$D_1 = \left[a + \frac{h_2}{2}, b - \frac{h_2}{2}\right] \setminus D_2$$
$$D_2 = \bigcup_{q=1}^{m} \left[x_q - \frac{h_2}{2}, x_q + \frac{h_2}{2}\right] \tag{3}$$

The LL volatility estimator, also named centred estimator, is $\hat{\sigma}_c^2(x) = \hat{a}$ for:

$$(\hat{a}, \hat{b}) = \arg\min_{(a,b)} \sum_{t=1}^{n} \left[\hat{r}_t^2 - a - b(X_t - x)\right]^2 K_c\left(\frac{X_t - x}{h_2}\right) \tag{4}$$

The centred estimator is smooth because it uses the information obtained from data points on the right and left sides of x. For this reason though, it is inconsistent where $\sigma^2(x)$ is discontinuous. Two other estimators of the volatility (left and right) may be found using the asymmetric kernels $K_l(\cdot)$ and $K_r(\cdot)$ defined in the introduction. The technique of choosing between the left and right estimator has been previously reported, for instance: i) Hamrouni (1999) uses it to locate the breaks of the conditional mean function, and ii) Qiu (2003) and Gijbels *et al.* (2007) use it to estimate the conditional mean with breaks. A general expression of the three estimators is given by,

$$\hat{\sigma}_k^2(x) = \sum_{t=1}^{n} \hat{r}_t^2 K_k\left(\frac{X_t - x}{h_2}\right) \frac{s_{k,2} - s_{k,1}(X_t - x)}{s_{k,0}s_{k,2} - s_{k,1}^2} \quad \text{for} \quad k = c, l, r. \tag{5}$$

using also h_2 in calculating the quantities $s_{k,j}$.

The expression of the volatility first–order derivative is given by,

$$\hat{\dot{\sigma}}_k^2(x) = \sum_{t=1}^{n} \hat{r}_t^2 K_k\left(\frac{X_t - x}{h_2}\right) \frac{s_{k,0}(X_t - x) - s_{k,1}}{s_{k,0}s_{k,2} - s_{k,1}^2} \quad \text{for} \quad k = c, l, r. \tag{6}$$

Proposition 2.1 *Under regularity conditions (C1) to (C4) the expression of the mean squared error (MSE) for each estimator is as follows:*
For a given point $x \in D_1$,

$$MSE(\hat{\sigma}_k^2(x)) = \left[\frac{h_2^2 \ddot{\sigma}^2(x)}{2} \frac{\mu_{k,2}^2 - \mu_{k,1}\mu_{k,3}}{\mu_{k,2}\mu_{k,0} - \mu_{k,1}^2}\right]^2 + \frac{(E(\epsilon^4|X = x) - 1)\sigma^4(x)}{nh_2 f_X(x)} V_k + o_p\left(h_1^2 + h_2 + \frac{1}{nh_2}\right)$$

For a given point $x \in D_2$ such that $x = x_q + \tau h_2$ with $\tau \in [0, \frac{1}{2}]$, where the break is of *magnitude d_q, the left and centred estimators have an $O_p(1)$ bias:*

$$MSE(\hat{\sigma}_l^2(x)) = \left[d_q \int_{\tau}^{\frac{1}{2}} K_l(u)\frac{\mu_{l,2}-\mu_{l,1}u}{\mu_{l,0}\mu_{l,2}-\mu_{l,1}^2}du\right]^2 + \frac{(E(\epsilon^4|X=x)-1)\sigma^4(x)}{nh_2 f_X(x)}V_l + o_p\left(\frac{1}{nh_2}\right)$$

$$MSE(\hat{\sigma}_c^2(x)) = \left[d_q \int_{\tau}^{\frac{1}{2}} K_c(u)du\right]^2 + \frac{(E(\epsilon^4|X=x)-1)\sigma^4(x)}{nh_2 f_X(x)}V_c + o_p\left(\frac{1}{nh_2}\right)$$

For a given point $x \in D_2$ such that $x = x_q + \tau h_2$ with $\tau \in [-\frac{1}{2}, 0]$ where the break is of magnitude d_q, the right and centred estimators have an $O_p(1)$ bias:

$$MSE(\hat{\sigma}_r^2(x)) = \left[d_q \int_{-\frac{1}{2}}^{\tau} K_r(u)\frac{\mu_{r,2}-\mu_{r,1}u}{\mu_{r,0}\mu_{r,2}-\mu_{r,1}^2}du\right]^2 + \frac{(E(\epsilon^4|X=x)-1)\sigma^4(x)}{nh_2 f_X(x)}V_r + o_p\left(\frac{1}{nh_2}\right)$$

$$MSE(\hat{\sigma}_c^2(x)) = \left[d_q \int_{-\frac{1}{2}}^{\tau} K_c(u)du\right]^2 + \frac{(E(\epsilon^4|X=x)-1)\sigma^4(x)}{nh_2 f_X(x)}V_c + o_p\left(\frac{1}{nh_2}\right)$$

for $\mu_{k,j} = \int u^j K_k(u)du$ and $V_k = \int K_k^2(u)\left[\frac{\mu_{k,2}-\mu_{k,1}u}{\mu_{k,0}\mu_{k,2}-\mu_{k,1}^2}\right]^2 du.$

Proposition 2.1 shows that under regularity conditions imposed on the conditional mean and volatility functions, and if $h_1, h_2 \to 0$ and $nh_2 \to \infty$ as $n \to \infty$ is satisfied, then the centred, left and right estimators of the volatility function converge in probability and they are consistent for points $x \in D_1$. At points $x \in D_2$, the right estimator is consistent when x is in a neighbourhood at the right of the break x_q; and the left estimator is consistent in a neighbourhood at the left of x_q.

The weighted residual mean squares (WRMS) measures how well each estimator is fitted to the data:

$$WRMS_k(x) = \frac{\sum_{t=1}^n [\hat{r}_t^2 - \hat{\sigma}_k^2(x) - \hat{\dot{\sigma}}_k^2(x)(X_t - x)]^2 K_k\left(\frac{X_t - x}{h_2}\right)}{\sum_{t=1}^n K_k\left(\frac{X_t - x}{h_2}\right)}.$$

Proposition 2.2 *Under regularity conditions (C1) to (C4), asymptotical expressions*

of the WRMS are as follows:
For a given point $x \in D_1$,

$$WRMS_k(x) = \sigma^4(x)(E(\epsilon^4|X=x)-1) + R_{k,1}(x)$$

For a given point $x \in D_2$ such that $x = x_q + \tau h_2$ with $\tau \in [0, \frac{1}{2}]$, where the break is of magnitude d_q,

$$\begin{aligned}
WRMS_l(x) &= \sigma^4(x)(E(\epsilon^4|X=x)-1) + d_q^2 C_{l,\tau}^2 + R_{l,2}(x) \\
WRMS_r(x) &= \sigma^4(x)(E(\epsilon^4|X=x)-1) + R_{r,2}(x) \\
WRMS_c(x) &= \sigma^4(x)(E(\epsilon^4|X=x)-1) + d_q^2 C_{c,\tau}^2 + R_{c,2}(x)
\end{aligned}$$

For a given point $x \in D_2$ such that $x = x_q + \tau h_2$ with $\tau \in [-\frac{1}{2}, 0]$ where the break is of magnitude d_q,

$$\begin{aligned}
WRMS_l(x) &= \sigma^4(x)(E(\epsilon^4|X=x)-1) + R_{l,3}(x) \\
WRMS_r(x) &= \sigma^4(x)(E(\epsilon^4|X=x)-1) + d_q^2 C_{r,\tau}^2 + R_{r,3}(x) \\
WRMS_c(x) &= \sigma^4(x)(E(\epsilon^4|X=x)-1) + d_q^2 C_{c,\tau}^2 + R_{c,3}(x)
\end{aligned}$$

where

$$\begin{aligned}
C_{k,\tau}^2 = M_k \int_{-\tau}^{\frac{1}{2}} &\left[\int_{-\frac{1}{2}}^{-\tau} K_k(z)(\mu_{k,2} - \mu_{k,1} z) dz - u \int_{-\tau}^{\frac{1}{2}} K_k(z)(\mu_{k,0} z - \mu_{k,1}) dz \right]^2 K(u) du \\
+ M_k \int_{-\frac{1}{2}}^{-\tau} &\left[\int_{-\tau}^{\frac{1}{2}} K_k(z)(\mu_{k,2} - \mu_{k,1} z) dz + u \int_{-\tau}^{\frac{1}{2}} K_k(z)(\mu_{k,0} z - \mu_{k,1}) dz \right]^2 K(u) du,
\end{aligned}$$

and $M_k = \{\mu_{k,0}(\mu_{k,2}\mu_{k,0} - \mu_{k,1})^2\}^{-1}$.

The residuals $R_{k,j}$ for $k = c, l, r$ and $j = 1, 2, 3$ are asymptotically zero with probability

1 and uniformly in x.

3 The break preserving local linear estimator

Although the three estimators have the same asymptotical bias for $x \in D_1$, in practice the centred estimator is best because it is smoother. The left estimator should be chosen to estimate points in a small interval on the left side of the break and the right estimator is the only one consistent in a small interval on the right side of the break. Thus, this paper proposes an estimator that combines all appropriately:

$$\hat{\sigma}^2_{BPLL}(x) = \begin{cases} \hat{\sigma}^2_c(x) & \text{diff}(x) < u \\ \hat{\sigma}^2_l(x) & \text{diff}(x) \geq u \text{ and } WRMS_l(x) < WRMS_r(x) \\ \hat{\sigma}^2_r(x) & \text{diff}(x) \geq u \text{ and } WRMS_l(x) > WRMS_r(x) \\ (\hat{\sigma}^2_l(x) + \hat{\sigma}^2_r(x))/2 & \text{diff(x)} \geq u \text{ and } WRMS_l(x) = WRMS_r(x) \end{cases} \tag{7}$$

where $\text{diff}(x) = \max(WRMS_c(x) - WRMS_r(x), WRMS_c(x) - WRMS_l(x))$. Proposition 2.2 shows that for a given point $x \in D_1$, $\text{diff}(x)$ is very close to zero. However for a given point x in an interval of the break x_q, $\text{diff}(x) = d_q^2 C^2_{c,\tau} + o(1)$ $a.s.$ Therefore, $0 < u < \max_q(d_q^2)C^2_{c,\tau}$.

Remark. As the $\hat{\sigma}^2_{BPLL}(x)$ chooses the appropriate estimator at every point, its bias at a given point x is:

$$\beta_k(x) = \frac{h_2^2 \ddot{\sigma}^2(x)}{2} \frac{\mu^2_{k,2} - \mu_{k,1}\mu_{k,3}}{\mu_{k,2}\mu_{k,0} - \mu^2_{k,1}}. \tag{8}$$

Theorem 3.1 *Under conditions (C1) to (C6),* $\sqrt{nh_2}(\sigma^2(x) - \hat{\sigma}^2_{BPLL}(x) - \beta_k(x))$ *is asymptotically normal with mean 0 and variance*

$$\frac{(E(\epsilon^4|X) - 1)\sigma^4(x)}{f_X(x)} \int K_k^2(u) \left[\frac{\mu_{k,2} - \mu_{k,1}u}{\mu_{k,0}\mu_{k,2} - \mu^2_{k,1}}\right]^2 du + o_p\left(\frac{1}{nh_2}\right),$$

where β_k *is as in equation (8).*

3.1 Ensuring positiveness

The fact that the LL estimator is sometimes negative for finite samples is widely known. Asymptotically, this is not the case and the number of negative values decreases as n increases. A solution to this problem must be found for this estimator to be useful. The simplest solution is to discard those negative values. A much better solution is the re–weighted Nadaraya–Watson estimator (see Hall *et al.*, 1999; Cai, 2001; and Phillips and Xu, 2007). However, the re–weighted Nadaraya–Watson estimator cannot be extended to estimate volatility functions with breaks because $\sum_{t=1}^{n}(X_t - x)w_t(x)K_k\left(\frac{X_t - x}{h_2}\right) = 0$ for $k = l, r$ cannot be satisfied for $w_t \geq 0$ weights adding up to one. The exponential local linear (ELL) estimator has been proposed by Ziegelmann (2002) for continuous volatility functions. It is outside the scope of this paper to study a break preserving ELL estimator. However, the ELL left, right and centred estimators would have similar properties to the LL left, right and centred estimators because they are both a nonparametric local linear estimator. In fact, the LL and the ELL have the same variance. Regarding the bias, the ELL may have a smaller bias than the LL if certain conditions are satisfied. Therefore, the occasional use of the ELL does not distort the results of the BPLL.

The solution proposed here is to substitute any negative values of $\hat{\sigma}_k^2(x)$ by $\hat{\sigma}_{k,ELL}^2(x)$ for $k = c, l, r$ where $\hat{\sigma}_{k,ELL}^2(x) = \exp(\hat{c})$ such that

$$(\hat{c}, \hat{d}) = \arg\min_{b,c} \sum [\hat{r}_t^2 - \exp(c + d(X_t - x))]^2 K_k \left(\frac{X_t - x}{h_2}\right).$$

3.2 Bandwidth selection

The expressions of the optimal global and local bandwidths, as the values that minimise the mean integrated squared error (MISE) and the MSE, depend on unknown functions such as $f_X(x)$ and $\sigma^2(x)$, for which plug–in estimators are needed. Instead, the authors propose to use the leave-one-out cross validation to find the bandwidth and the threshold value u when $\{X\}$ is an iid process. A two dimensional minimisation must be performed in the following way,

$$(h_2^{cv}, u^{cv}) = \arg\min_{h_2,u} \sum_{t=1}^{n} \left[\hat{r}_t^2 - \hat{\sigma}_{-t}^2(X_t)\right]^2$$

where $\hat{\sigma}_{-t}^2(X_t)$ is the estimator obtained when the pair (Y_t, X_t) is left out.

In case that the lag of dependency in $\{Y, X\}$ is of size l, the authors propose the leave an l–block–out cross validation,

$$(h_2^l, u^l) = \arg\min_{h2,u} \sum_{t=1}^{n} \left[\hat{r}_t^2 - \hat{\sigma}^2_{-l_t}\right]^2$$

where $\hat{\sigma}^2_{-l_t}$ is calculated without using the $2l{+}1$ pairs $(X_{t-l}, Y^2_{t-l}), \ldots, (X_t, Y_t^2), \ldots, (X_{t+l}, Y^2_{t+l})$.

4 Simulation experiments

The theoretical results for finite samples are tested in this section. The first experiment tests the performance of the BPLL when (Y_t, X_t) is and iid process. Experiment 2 assumes that there is an AR(1) dependency within X_t.

4.1 Experiment 1: iid variables

The model $Y_t = m(X_t) + \sigma(X_t)\epsilon_t$ considered is based on Example 2 in Fan and Yao (1998), with $X_t \overset{iid}{\sim} U[-2,2]$ and innovations $\epsilon_t \overset{iid}{\sim} N(0,1)$. The points of estimation x are $T = 250$ equidistant values in [-1.8,1.8]. Processes of size $n = 500, 1000$ and 2000 where simulated $N = 200$ number of times. The centred kernel used is the Epanechnikov kernel: $K_c(u) = 12/11(1-u^2)I[-\frac{1}{2} \leq u \leq \frac{1}{2}]$.

The volatility function $\sigma^2 : \Re \to \Re^+$ is bounded and continuous everywhere except for $x_1 = -1$ and $x_2 = 1$. Therefore the two differentiated regions are $D_1 = [-2,2] \setminus D_2$ and $D_2 = [-1 - \frac{h_2}{2}, -1 + \frac{h_2}{2}] \bigcup [1 - \frac{h_2}{2}, 1 + \frac{h_2}{2}]$. The volatility function second–order derivative exists and is bounded in $x \in D_1$. The left and right second derivatives at x_1 and x_2 exist and are bounded. The volatility expression is:

$$\sigma(x) = \begin{cases} 0.4\exp(-2x^2) + 0.2 & x \leq -1 \\ 0.4\exp(-2x^2) & -1 < x \leq 1 \\ 0.4\exp(-2x^2) + 0.1 & x > 1 \end{cases} \quad (9)$$

The estimator $\hat{m}(X_t)$ is obtained with the methodology in Gijbels *et al.* (2007). The authors distinguish amongst four scenarios:

Scenario I: the conditional mean function is known, this is equivalent to: $m \equiv 0$.

Scenario II: the conditional mean is unknown and continuous in the support of x: $m(x) = 0.5(x + 2\exp(-16x^2))$.

Scenario III: the conditional mean is unknown and has one break at $x_3 = 0$. Its second derivative exists and is bounded for the continuous part and the left and right second derivatives exist and are bounded in $x_3 = 0$:

$$m(x) = \begin{cases} 0.5(x + 2\exp(-16x^2)) & x \leq 0 \\ 0.5(x + 2\exp(-16x^2)) - 0.7 & x > 0 \end{cases} \tag{10}$$

Scenario IV: the conditional mean is unknown and it has breaks at the same points as the volatility function. Its second derivative exists for the continuous part and is bounded in the continuous points and the left and right second derivatives exist and are bounded in x_1 and x_2.

$$m(x) = \begin{cases} 0.5(x + 2\exp(-16x^2)) - 0.5 & x \leq -1 \\ 0.5(x + 2\exp(-16x^2)) & -1 < x \leq 1 \\ 0.5(x + 2\exp(-16x^2)) + 0.9 & x > 1 \end{cases} \tag{11}$$

Comparison of the two models MISE is a way to compare their global performance. A numerical approximation of the estimator MISE may be obtained by

$$\widehat{MISE} = \frac{1}{N}\sum_{k=1}^{N} \widehat{ISE}_k \qquad \widehat{ISE}_k = \sum_{t=1}^{T-1} \frac{SE_k(x_t) + SE_k(x_{t+1})}{2T} \tag{12}$$

where $SE_k(x_t)$ is the squared error of the kth–simulated sample. Around the points of break x_q: $\widehat{ISE}_{q,k} = \sum_{t=1}^{T-1} \frac{SE_k(x_t)+SE_k(x_{t+1})}{2T} I[x_q - 0.1 < x_t < x_{t+1} < x_q + 0.1]$. The local MISE estimator is:

$$\widehat{MISE}_q = \frac{1}{N}\sum_{q=1}^{m}\sum_{k=1}^{N} \widehat{ISE}_{k,q}. \tag{13}$$

The MISE comparison of the two estimators for each scenario is shown in Table 1. The first thing to notice is that the MISE decreases as the sample size increases which demonstrates the convergence results. The values of the MISE are statistically the same amongst the four different scenarios because the error committed in estimating the conditional mean does not affect the volatility estimation. The global MISE for the LL is better than the global MISE of the BPLL when the sample size is $n = 500$ but this advantage disappears as the sample size increases. The BPLL performs strikingly better than the LL in the points of break for all the cases and sizes. In fact, the global

MISE for the LL estimator mainly comes from the error committed at the breaks when $n = 2000$.

Method	LL		BPLL	
	$\widehat{\text{MISE}}$	$\widehat{\text{MISE}}_q$	$\widehat{\text{MISE}}$	$\widehat{\text{MISE}}_q$
$n = 500$				
Scenario I	0.0089	0.0060	0.0098	0.0039
Scenario II	0.0098	0.0062	0.0120	0.0039
Scenario III	0.0093	0.0061	0.0109	0.0040
Scenario IV	0.0107	0.0067	0.0123	0.0042
$n = 1000$				
Scenario I	0.0047	0.0034	0.0037	0.0015
Scenario II	0.0044	0.0032	0.0043	0.0018
Scenario III	0.0048	0.0034	0.0045	0.0017
Scenario IV	0.0044	0.0032	0.0040	0.0016
$n = 2000$				
Scenario I	0.0021	0.0016	0.0012	0.0005
Scenario II	0.0020	0.0016	0.0012	0.0006
Scenario III	0.0020	0.0016	0.0012	0.0006
Scenario IV	0.0022	0.0016	0.0013	0.0005

Table 1: MISE of LL and BPLL comparison for Experiment 1.

For a further comparison of the two estimators, the authors also calculate the mean absolute deviation error (MADE) which is more robust to outliers than the MISE. Its expression is given by:

$$MADE_k = \frac{1}{T}\sum_{t=1}^{T} |\sigma(x_t) - \hat{\sigma}(x_t)|$$

where k refers to the kth–simulated sample. A local MADE is also obtained for points in the neighbourhood of the breaks,

$$MADE_{q,k} = \frac{1}{T}\sum_{q=1}^{m}\left\{\sum_{t=1}^{T} |\sigma(x_t) - \hat{\sigma}(x_t)| I[x_q - 0.1 < x_t < x_q + 0.1]\right\}.$$

The boxplot of the MADE for scenario IV is shown in Figure 1. A similar interpretation than with the MADE results is drawn here. Figures 1(a), 1(c) and 1(e) display the results with the global MADE. The LL estimator has a lower MADE when n is small than the BPLL. Although, as it can be seen in Figures 1(b), 1(d) and 1(f), the local MADE referring to the point–wise performance at the breaks is better estimated by the BPLL. Moreover, the mean of the LL local MADE is around 0.035 and does not decrease as n increases, which shows its inconsistency.

Figure 2 compares the LL with the BPLL estimator. Figures 2(a) and 2(b) correspond to the LL estimator for sets of length $n = 500$ and $n = 2000$ respectively. Figures 2(c) and 2(d) correspond to the BPLL estimator counterparts. The solid line draws $\sigma(x)$. The the dashed line corresponds to the estimate whose MADE is equal to its median amongst the 200 different samples. The dotted lines correspond to the 5%, 95% confidence interval. The LL estimator is smoother than the BPLL estimator in the points of continuity, as expected. One also appreciates that the LL estimator is inconsistent in x_1 and x_2, its confidence intervals do not include these points. On the other hand, the BPLL estimator of the volatility at x_1 and x_2 improves as n increases.

4.2 Experiment 2: a square-root diffusion

The first to assume that the interest rates behave like a square–root diffusion process was Cox *et al.* (1985) with the CIR model. In this model, the X is an AR(1) and therefore it is a good example to show that the BPLL also works when there is dependency in the data.

The process is of the form:

$$dX_t = \kappa(\theta - X_t)dt + \sigma\sqrt{X_t}dB_t$$

where κ is the persistence of X_t, θ is the long–run mean of X_t, $\sigma > 0$ is the instantaneous standard deviation and B_t is a Brownian process. The process was generated following the algorithm in Section 3.4 of Glasserman (2004).

The conditional mean and volatility functions have breaks at $x_1 = 0.07$ and $x_2 = 0.10$. The values of the model parameters are taken from Aït–Sahalia (1996) and Chapman and Pearson (2000). These are: $\kappa_1 = 0.21459$, $\sigma_1 = 0.07830$, $\kappa_2 = 0.85837$, $\sigma_2 = 0.15660$ and $\theta = 0.085711$ and were estimated from the U.S. interest rates and the Eurodollar data sets. The expressions of the conditional mean and volatility functions used for the simulation are as follows:

$$m(x) = \begin{cases} \kappa_1(\theta - x) & 0.07 < x \leq 0.1 \\ \kappa_2(\theta - x) & \text{otherwise} \end{cases} \quad (14)$$

$$\sigma(x) = \begin{cases} \sigma_1\sqrt{x} & 0.07 < x \leq 0.1 \\ \sigma_2\sqrt{x} & \text{otherwise} \end{cases} \quad (15)$$

The bandwidths h_1 and h_2 are obtained using the l–block–out cross validation for the lag $l = 1$. A set of N=400 simulations were performed for series of length $n = 500, 1000$ and 2000. The points of estimation x are $T = 250$ equidistant values in [0.03, 0.12]. Figures 3(a) and 3(b) correspond to the LL estimator for sets of length $n = 500$ and $n = 2000$ respectively. Figures 3(c) and 3(d) correspond to the BPLL estimator counterparts. This is clearly a more difficult problem than Experiment 1 and needs more points to ensure a good estimation. In fact, Chapman and Pearson (2000) simulate sets of length 7500 and 15000. Therefore the estimator for $n = 500$ is not very good. The LL estimator is too smooth, staying in the middle of the different values of the function. However the BPLL, although it is more variable, estimates the function better, reaching the points of break. The LL also performs worse in the boundaries. There are not many points in the neighbourhood of 0.03 and the performance of the centred estimator is worse than the performance of the right estimator in those points. The LL needs a larger bandwidth to compensate this problem in the boundaries. In fact, the average bandwidth for the LL when $n = 2000$ is 0.024, while its average value is 0.016 for the BPLL. Graphically this issue is shown in Figures 2 (b) and 2 (d).

The global performance of the two estimators is compared in Table 2 and Figure 4. Both the global and local MISE are better for the BPLL estimator. The same behaviour pointed out in Experiment 1 is found here. The inconsistency of the LL estimator manifests on the local MADE results whose mean stays fixed independently of the process size.

Method	LL		BPLL	
	$\widehat{\text{MISE}}$	$\widehat{\text{MISE}}_q$	$\widehat{\text{MISE}}$	$\widehat{\text{MISE}}_q$
$n = 500$	0.0241	0.0083	0.0114	0.0057
$n = 1000$	0.0120	0.0041	0.0039	0.0022
$n = 2000$	0.0059	0.0021	0.0013	0.0008

Table 2: MISE of LL and BPLL comparison for Experiment 2.

5 Conclusions

This paper introduces a novel one step estimator, the BPLL, for a volatility function with breaks. Although the number of breaks and their location are unknown, this procedure is able to estimate the volatility function consistently at any point. The simulation experiments illustrate the asymptotic results for both, when (Y, X) is an independent two–dimensional process and when (Y, X) satisfies the β–mixing condition. In fact, it is important to point out that the BPLL performs strikingly better than the LL in the latter scenario.

Since the availability of intra–daily data, much work has gone into the estimation and forecasting of the stochastic spot volatility. Certainly, kernel smoothing techniques have developed further in this area as well, for example Kristensen (2007) and Bandi and Renò (2008) amongst others. The extension of the BPLL to the estimation of the spot volatility with breaks is not straightforward. A priori, i) the process (Y, X) must be independent, ii) jumps in the sense of extreme shocks must be considered, and iii) microstructure noise may cause a heavy bias on the BPLL estimator. Therefore, research into the estimation of the spot volatility with breaks is left for future work.

References

Bandi, F. M. and Renò, R. (2008) Nonparametric stochastic volatility. *Working paper.*

Bradley, R. C. (2005) Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys* **2**, 107–144.

Cai, Z. (2001) Weighted Nadaraya–Watson regression estimation. *Statistics and Probability Letters* **51**, 307–318.

Chapman, D. A. and Pearson, N. (2000) Is the Short Rate Drift Actually Nonlinear? *Journal of Finance* **55**, 355–388.

Cerra, V. and Sweta, S. C. (2005) Did output recover from the Asian crisis? *IMF Staff Papers* **52**, 1–23.

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985) A theory of the term structure of interest rates. *Econometrica* **53**, 385–407.

Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications.* Chapman and Hall/CRC.

Fan, J., and Yao, Q. (1998) Efficiente estimation of conditional variance functions in stochastic regression. *Biometrika* **85**, 645–660.

Gijbels, I., Lambert, A., and Qiu, P. (2007) Jump-preserving regression and smoothing using local linear fitting: a compromise. *The Annals of the Institute of Statistical Mathematics* **59**, 235–272.

Glasserman, P. (2004), *Monte Carlo Methods in Financial Engineering.* Springer.

Hall, P., Wolff, R. C. L., and Yao, Q. (1999) Methods for estimating a conditional distribution function. *Journal of the American Statistical Association* **94**, 154–163.

Hamilton, J. D. (1988) Rational–expectations econometric analysis of changes in regime: an investigation of the term structure of interest rates. *Journal of Economic Dynamics and Control* **12**, 385–423.

Hamilton, J. D. (1989) A new approach to the economic analysis of time series. *Econometrica* **57**, 357–384.

Hamilton, J. D. (2005) What's real about the business cycle? *Federal Reserve Bank of St. Louis Review* **87**, 435–452.

Hamrouni, Z. (1999) *Inférence statistique par lissage linéaire local pour une fonction de régression présentant des discontinuités.* PhD. Thesis from the University of Joseph Fourier, Grenoble I.

Kristensen, D. (2007) Nonparametric filtering of the realized spot volatility: a kernel–based approach. *CREATES Research Paper 2007–2.*

Lambert, A. (2005), *Nonparametric estimations of discontinuous curves and surfaces.* PhD thesis, Institut de Statistique, Université Catholique de Louvain, Belgium.

Lin, C. J. and Teräsvirta, T. (1994) Testing the constancy of regression parameters against continuous structural change. *Journal of Econometrics* **62**, 211–228.

Phillips, P. C. B., and Xu, K. L. (2007) Tilted nonparametric estimation of volatility functions. *Cowles foundation discussion paper no. 1612.*

Qiu, P. (2003) A jump–preserving curve fitting procedure based on local piecewise–linear kernel estimation. *Nonparametric Statistics* **15**, 437–453.

Ziegelmann, F. A. (2002). Nonparametric estimation of volatility functions: the local exponential estimator. *Econometric Theory* **18**, 985–991.

Appendix A

Define $f_X(\cdot)$ as the marginal density function of X with support in $[a, b]$, $f_{\epsilon|X}(\cdot)$ as the conditional density function of ϵ given X, $m(\cdot)$ as the conditional mean function, $\sigma^2(\cdot) > 0$ as the volatility function, $K(\cdot)$ is a symmetric kernel with support in $[-\frac{1}{2}, \frac{1}{2}]$, $h_1 > 0$ is the bandwidth used in the estimation of the conditional mean and $h_2 > 0$ is the bandwidth in the estimation of the volatility. The following regularity conditions are assumed.

(C1) The kernel $K_c(\cdot)$, $f_X(\cdot)$ and $f_{\epsilon|X}(\cdot)$ are Lipschitz continuous.

(C2) For some $\delta \in [0, 1)$, $E(Y^{4(1+\delta)}) < \infty$.

(C3) Functions $m(\cdot)$ and $\sigma^2(\cdot)$ have continuous second–order derivatives, $\ddot{m}(\cdot)$ and $\ddot{\sigma}^2(\cdot)$, in $x \in D_1$.

(C4) The function $m(\cdot)$ has left and right bounded second–order derivatives in its discontinuities x_q, $\ddot{m}_-(x_q)$ and $\ddot{m}_+(x_q)$. Similarly, $\sigma^2(\cdot)$ has left and right bounded second–order derivatives in its discontinuities x_p, $\ddot{\sigma}^2_-(x_p)$ and $\ddot{\sigma}^2_+(x_p)$.

(C5) $h_i \to 0$, $nh_i \to \infty$ and $\liminf nh_i^4 > 0$, for $i = 1, 2$ as $n \to \infty$.

(C6) The strictly stationary process $\{(X_i, Y_i)\}$ satisfies the β–mixing condition (see Bradley, 2005 for definition).

Remarks. Conditions (C2) and (C6) refer to the degree of dependency of the process $\{(X_i, Y_i)\}$. In the case of an independent process, $\delta = 0$. Condition (C6) implies that although there is dependency within $\{(X_i, Y_i)\}$, it does not last forever.

Appendix B: Proof Proposition 2.1

First, notice that using the Lipschitz property of the kernel, Lemma A.2 in Gijbels *et al.* (2007) is extended for the case of a random variable X,

$$s_{k,j} = nh_2^{j+1} f_X(x)\mu_{k,j} + O_p\left(\frac{1}{nh_2}\right). \tag{B.1}$$

Also,

$$\hat{r}_i^2 = r_i^2 + 2[m(X_i) - \hat{m}(X_i)]\sigma(X_i)\epsilon_i + [m(X_i) - \hat{m}(X_i)]^2 \tag{B.2}$$

For a given point $x \in D_1$, substituting (B.2) in expression (5),

$$\begin{aligned}
\hat{\sigma}_k^2(x) - \sigma^2(x) = & \textstyle\sum_{i=1}^n (r_i^2 - \sigma^2(x)) K_k \left(\frac{X_t - x}{h_2}\right) \frac{s_{k,2} - s_{k,1}(X_t - x)}{s_{k,0}s_{k,2} - s_{k,1}^2} \\
& + 2\textstyle\sum_{i=1}^n [m(X_i) - \hat{m}(X_i)]\sigma(X_i)\epsilon_i K_k \left(\frac{X_t - x}{h_2}\right) \frac{s_{k,2} - s_{k,1}(X_t - x)}{s_{k,0}s_{k,2} - s_{k,1}^2} \\
& + \textstyle\sum_{i=1}^n [m(X_i) - \hat{m}(X_i)]^2 K_k \left(\frac{X_t - x}{h_2}\right) \frac{s_{k,2} - s_{k,1}(X_t - x)}{s_{k,0}s_{k,2} - s_{k,1}^2} \\
= & \; I_1 + I_2 + I_3
\end{aligned} \tag{B.3}$$

Applying (B.1) and that the bias of the conditional mean estimator is $O_p\left(h_1^2\right)$, it follows:

$$I_3 = \frac{1}{nh_2 f_X(x)} \sum_{i=1}^n [m(X_i) - \hat{m}(X_i)]^2 K_l \left(\frac{X_t - x}{h_2}\right) \frac{\mu_{l,2} - \mu_{l,1}(\frac{X_i - x}{h_2})}{\mu_{l,0}\mu_{l,2} - \mu_{l,1}^2} + O_p\left(\frac{1}{nh_2}\right) \tag{B.4}$$

Therefore $E(I_3) = O_p\left(h_1^4 + \frac{1}{nh_2}\right)$. Similarly $E(I_2) = O_p\left(h_2^2 + \frac{1}{nh_2}\right)$. Applying that $r_i^2 = \sigma^2(X_i)(\epsilon_i^2 - 1) + \sigma^2(X_i)$, that $E(\epsilon_i^2 | X_i = x) = 1$ and the result in (B.1):

$$\begin{aligned}
E(I_1) = & \; E\left([\sigma^2(X_i)(\epsilon_i^2 - 1) + \sigma^2(X_i) - \sigma^2(x)] K_k \left(\frac{X_t - x}{h_2}\right) \frac{s_{k,2} - s_{k,1}(X_t - x)}{s_{k,0}s_{k,2} - s_{k,1}^2}\right) \\
= & \int (\omega^2 - 1) f_{\epsilon|X}(\omega) \int \left[\sigma^2(x) + h_2 \dot{\sigma}^2(x) + \frac{h_2^2 u^2 \ddot{\sigma}^2(x)}{2} + o_p(h_2^2)\right] K_k(u) \\
& \times \frac{\mu_{k,2} - \mu_{k,1}u}{\mu_{k,0}\mu_{k,2} - \mu_{k,1}^2} [f_X(x) + o_p(1)] du d\omega \\
& + \int \left[h_2 \dot{\sigma}^2(x) + \frac{h_2^2 u^2 \ddot{\sigma}^2(x)}{2} + o_p(h_2^2)\right] K_k(u) \frac{\mu_{k,2} - \mu_{k,1}u}{\mu_{k,0}\mu_{k,2} - \mu_{k,1}^2} [f_X(x) + o_p(1)] du
\end{aligned} \tag{B.5}$$

Adding all the terms in (B.3):

$$Bias(x) = \frac{h_2^2 \ddot{\sigma}^2(x)}{2} \frac{\mu_{k,2}^2 - \mu_{k,1}\mu_{k,3}}{\mu_{k,0}\mu_{k,2} - \mu_{k,1}^2} + O_p\left(h_1^2 + \frac{1}{nh_2}\right) \tag{B.6}$$

Performing the same type of calculations, the variance:

$$\begin{aligned}
Var(\hat{\sigma}_k^2(x)) = & \; E\left(\left[(r_i^2 - \sigma^2(x)) K_k \left(\frac{X_t - x}{h_2}\right) \frac{s_{k,2} - s_{k,1}(X_t - x)}{s_{k,0}s_{k,2} - s_{k,1}^2}\right]^2\right) + \text{lower order terms} \\
= & \; \frac{\sigma^4(x)(E(\epsilon^4 | X = x) - 1)}{nh_2 f_X(x)} \int K_k^2(u) \left[\frac{\mu_{k,2} - \mu_{k,1}u}{\mu_{k,0}\mu_{k,2} - \mu_{k,1}^2}\right]^2 du + o_p\left(\frac{1}{nh_2}\right)
\end{aligned} \tag{B.7}$$

For a given point $x \in D_2$ such that $x = x_q + \tau h_2$ with $\tau \in [-\frac{1}{2}, 0]$ where the jump is of magnitude d_q, and if conditions (C3) and (C4) are satisfied then the right estimator is given by,

$$
\begin{aligned}
\hat{\sigma}_r^2(x) = & \sum_{X_i<x_q} [\sigma_-^2(x_q) + o_p(1)] K_r\left(\frac{X_t - x}{h_2}\right) \frac{s_{r,2} - s_{r,1}(X_t - x)}{s_{r,0}s_{r,2} - s_{r,1}^2} \\
& + \sum_{X_i\geq x_q} [\sigma_-^2(x_q) + d_q + o_p(1)] K_r\left(\frac{X_t - x}{h_2}\right) \frac{s_{r,2} - s_{r,1}(X_t - x)}{s_{r,0}s_{r,2} - s_{r,1}^2} \\
= & \ \sigma_-^2(x_q) + \sum_{X_i\geq x_q} d_q K_r\left(\frac{X_t - x}{h_2}\right) \frac{s_{r,2} - s_{r,1}(X_t - x)}{s_{r,0}s_{r,2} - s_{r,1}^2} + o_p(1) \\
= & \ \sigma_-^2(x_q) + d_q \int_\tau^0 K_r(z) \frac{\mu_{r,2} - \mu_{r,1}u}{\mu_{r,0}\mu_{r,2} - \mu_{r,1}^2} dz + o_p(1).
\end{aligned}
\tag{B.8}
$$

The expression of the centred estimator may be obtained similarly applying that $\mu_{c,0} = 1$ and $\mu_{c,1} = 0$. A similar expression for the left and centred estimators can be obtained for $\tau \in [0, \frac{1}{2}]$. Thus, equations (B.6)–(B.8) prove Proposition 2.1.

Appendix C: Proof Proposition 2.2

$$
WRMS_k(x) = \frac{\sum_{i=1}^n [\hat{r}_i^2 - \hat{\sigma}_k^2(x) - \hat{\dot{\sigma}}_k^2(x)(X_i - x)]^2 K_k\left(\frac{X_t - x}{h_2}\right)}{\sum_{i=1}^n K_k\left(\frac{X_t - x}{h_2}\right)}.
$$

By the ergodic theorem, the denominator:

$$
D(x) = \sum_{i=1}^n K_k\left(\frac{X_t - x}{h_2}\right) \xrightarrow{p} nh_2\mu_{k,0}f_X(x). \tag{C.1}
$$

The numerator:

$$
\begin{aligned}
N(x) = & \ \textstyle\sum_{i=1}^n [\hat{r}_i^2 - \hat{\sigma}_k^2(x) - \hat{\dot{\sigma}}_k^2(x)(X_i - x)]^2 K_k\left(\frac{X_t - x}{h_2}\right) \\
= & \ \textstyle\sum_{i=1}^n (\epsilon_i^2 - 1)^2 \sigma^4(X_i) K_k\left(\frac{X_t - x}{h_2}\right) \\
& +2\textstyle\sum_{i=1}^n (\epsilon_i^2 - 1)\sigma^2(X_i)[\sigma^2(X_i) - \hat{\sigma}_k^2(x) - \hat{\dot{\sigma}}_k^2(x)(X_i - x)] K_k\left(\frac{X_t - x}{h_2}\right) \\
& +\textstyle\sum_{i=1}^n [\sigma^2(X_i) - \hat{\sigma}_k^2(x) - \hat{\dot{\sigma}}_k^2(x)(X_i - x)]^2 K_k\left(\frac{X_t - x}{h_2}\right) \\
= & \ N_1 + N_2 + N_3
\end{aligned}
\tag{C.2}
$$

Given $x \in D_1$:

$$
\hat{\sigma}_k^2(x) = \sum_{i=1}^n \hat{r}_i^2 K_k\left(\frac{X_t - x}{h_2}\right) \frac{s_{k,0}(X_t - x) - s_{k,1}}{s_{k,0}s_{k,2} - s_{k,1}^2} \tag{C.3}
$$

and $\dot{\sigma}^2(x) - \hat{\dot{\sigma}}_k^2(x) = o_p(1/h_2)$. Therefore,

$$
\begin{aligned}
N_3 = \ & \textstyle\sum_{i=1}^n [\sigma^2(X_i) - \hat{\sigma}_k^2(x)]^2 K_k\left(\dfrac{X_t - x}{h_2}\right) \\
& 2\textstyle\sum_{i=1}^n [\sigma^2(x) - \hat{\sigma}_k^2(x)][\dot{\sigma}^2(x) - \hat{\dot{\sigma}}_k^2(x)](X_i - x) K_k\left(\dfrac{X_t - x}{h_2}\right) \\
& \textstyle\sum_{i=1}^n [\dot{\sigma}^2(x) - \hat{\dot{\sigma}}_k^2(x)]^2 (X_i - x)^2 K_k\left(\dfrac{X_t - x}{h_2}\right) + o_p(nh_2) \\
& = o_p(nh_2).
\end{aligned}
\tag{C.4}
$$

It is easy to see that $N_2 = 0$ because $E(\epsilon_i^2|X_i = x) = 1$. On the other hand, using similar derivations than in the Appendix B,

$$
N_1 = nh_2(E(\epsilon^4|X_i = x) - 1)\sigma^4(x)\mu_{k,0} f_X(x) + o_p(nh_2) \tag{C.5}
$$

and therefore

$$
\frac{N(x)}{D(x)} = (E(\epsilon_i^4|X_i = x) - 1)\sigma^4(x) + o_p(1) \tag{C.6}
$$

For a given point $x \in D_2$ such that $x = x_q + \tau h_2$ with $\tau \in [-\frac{1}{2}, 0]$ where the jump is of magnitude d_q, and if conditions (C3) and (C4) are satisfied then the right estimator of the first–order derivative of $\sigma^2(\cdot)$ is given by,

$$
\hat{\dot{\sigma}}_r^2(x_q) = \frac{d_q}{h_2}\int_\tau^0 K_r(z)\frac{\mu_{r,0}u - \mu_{r,1}}{\mu_{r,0}\mu_{r,2} - \mu_{r,1}^2} dz + o_p(1/h_2). \tag{C.7}
$$

Using equations (B.8) and (C.7), the expression of N_3 from the left estimator is:

$$
\begin{aligned}
N_3 = & \sum_{X_i<x_q}\left[\sigma^2(X_i)-\sigma^2_-(x_q)-d_q\int_\tau^{1/2}K_r(z)\frac{\mu_{r,2}-\mu_{r,1}u}{\mu_{r,0}\mu_{r,2}-\mu^2_{r,1}}dz\right.\\
& \left.-\frac{d_q}{h_2}\int_\tau^{1/2}K_r(z)\frac{\mu_{r,0}u-\mu_{r,1}}{\mu_{r,0}\mu_{r,2}-\mu^2_{r,1}}dz(X_i-x)\right]^2 K_r\left(\frac{X_t-x}{h_2}\right)\\
& \sum_{X_i\geq x_q}\left[\sigma^2(X_i)-\sigma^2_-(x_q)-d_q\int_\tau^{1/2}K_r(z)\frac{\mu_{r,2}-\mu_{r,1}u}{\mu_{r,0}\mu_{r,2}-\mu^2_{r,1}}dz\right.\\
& \left.-\frac{d_q}{h_2}\int_\tau^{1/2}K_r(z)\frac{\mu_{r,0}u-\mu_{r,1}}{\mu_{r,0}\mu_{r,2}-\mu^2_{r,1}}dz(X_i-x)\right]^2 K_r\left(\frac{X_t-x}{h_2}\right)+o_p(nh_2)\\
= & \; nh_2f_X(x)\int_0^\tau\left[d_q\int_\tau^{1/2}K_r(z)\frac{\mu_{r,2}-\mu_{r,1}u}{\mu_{r,0}\mu_{r,2}-\mu^2_{r,1}}dz+ud_q\int_\tau^{1/2}K_r(z)\frac{\mu_{r,0}u-\mu_{r,1}}{\mu_{r,0}\mu_{r,2}-\mu^2_{r,1}}dz\right]^2K_r(u)du\\
& +nh_2f_X(x)\int_\tau^{1/2}\left[d_q\int_0^\tau K_r(z)\frac{\mu_{r,2}-\mu_{r,1}u}{\mu_{r,0}\mu_{r,2}-\mu^2_{r,1}}dz-ud_q\int_\tau^{1/2}K_r(z)\frac{\mu_{r,0}u-\mu_{r,1}}{\mu_{r,0}\mu_{r,2}-\mu^2_{r,1}}dz\right]^2K_r(z)du\\
& +o_p(nh_2)\\
= & \; d_q^2C^2_{r,\tau}nh_2f_X(x)+o_p(nh_2).
\end{aligned}
\tag{C.8}
$$

When $X_i < x_q$, the volatility is approximated in a small interval around x_q by $\sigma^2(X_i) = \sigma^2_-(x_q) + o_p(1)$ *a.s.* However when $X_i \geq x_q$, $\sigma^2(X_i) = \sigma^2_-(x_q) + d_q + o_p(1)$ *a.s.* Due to $WRMS_r(x) = N(x)/D(x)$, then Proposition 2.2 is proven.

Appendix D: Proof Theorem 3.1

Fan and Yao (1998) show that under the regularity conditions in Appendix A, the centred estimator is asymptotically normal. The right estimator is asymptotically normal in the points of continuity and in the points to the right of a discontinuity. Furthermore, the left estimator is asymptotically normal in the points of continuity and in the points to the left of a discontinuity. These two last results have been proven for the conditional mean estimator in Lambert (2005).

Following the idea in the proof of Theorem 3.2 of Gijbels *et al.* (2007), the estimator BPLL can be rewritten as:

$$
\hat{\sigma}^2_{BPLL}(x) = \hat{\sigma}^2_c(x)A_n(x) + \hat{\sigma}^2_l(x)B_n(x) + \hat{\sigma}^2_r(x)C_n(x) + \frac{\hat{\sigma}^2_l(x)+\hat{\sigma}^2_r(x)}{2}D_n(x) \tag{D.1}
$$

where $A_n(x), B_n(x), C_n(x)$ and $D_n(x)$ are the regions regarding each of the inequalities in equation (7). It is easy to see that these regions are mutually exclusive and that for any $x \in [a, b]$, $I(A_n(x)) + I(B_n(x)) + I(C_n(x)) + I(D_n(x)) = 1 \;\; a.s.$

For a given $x \in D_1$, by Proposition (2.2), the value of diff(x) is asymptotically zero and therefore the threshold value u also tends to zero which means that $I(An(x))$ tends to 1 $a.s.$ and $I(B_n(x)) + I(C_n(x)) + I(D_n(x))$ tends to zero $a.s.$ Therefore $\hat{\sigma}^2_{BPLL}(x) = \sigma^2_c(x)$ $a.s.$ in that region.

For $x = x_q + \tau h_2 \in D_2$ where $-\frac{1}{2} \leq \tau \leq 0$, with a magnitude of jump d_q,

$$\text{diff}(x) = \max(d_q^2 C_{c,\tau}^2 + R_{c,3}(x) - R_{l,3}(x), d_q^2(C_{c,\tau}^2 - C_{r,\tau}^2) + R_{c,3}(x) - R_{r,3}(x)). \quad \text{(D.2)}$$

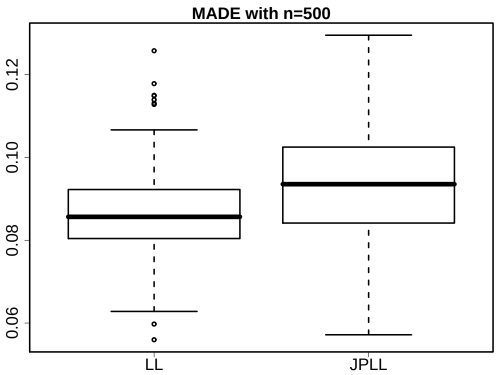
We have,

$$\lim_{n\to\infty} \text{diff}(x) = d_q^2 \max(C_{c,\tau}^2, C_{c,\tau}^2 - C_{r,\tau}^2) = d_q^2 C_{c,\tau}^2 \quad \text{(D.3)}$$

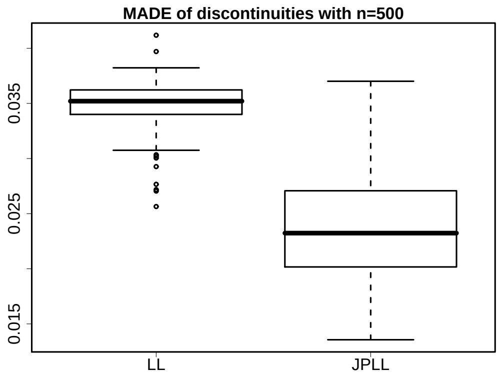
Since, $0 < u < d_q^2 C_{c,\tau}^2$ then diff$(x) > u$ and $I(A_n) = 0 \;\; a.s.$ Now, is it $WRMS_r(x) > WRMS_l(x)$?

$$WRMS_r(x) - WRMS_l(x) = d_q^2 C_{r,\tau}^2 + R_{l,3} - R_{r,3} \to d_q^2 C_{r,\tau}^2 > 0 \;\; as \;\; n \to \infty, \quad \text{(D.4)}$$

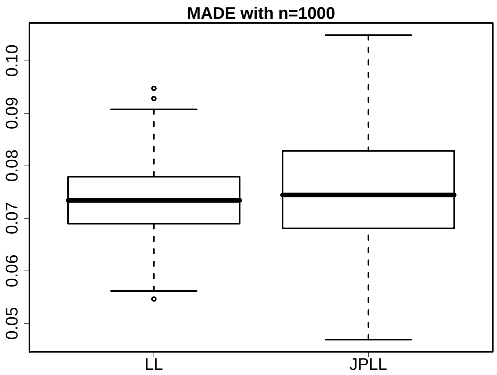
therefore $I(B_n(x)) = 1$ $a.s.$ and $\hat{\sigma}^2_{BPLL}(x) = \hat{\sigma}^2_l(x)$. Equivalently, $I(C_n(x)) = 1$ $a.s.$ for points on a small interval to the right of a discontinuity and $I(D_n(x)) = 1$ $a.s$ for points on a small interval of a discontinuity where $WRMS_l(x) = WRMS_r(x)$. As each of the estimators is asymptotically normal in the areas where they are chosen then σ^2_{BPLL} is asymptotically normal and Theorem 3.1 is proven.



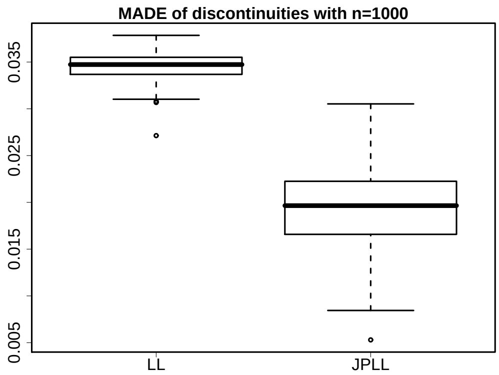
(a) $MADE$ with $n = 500$



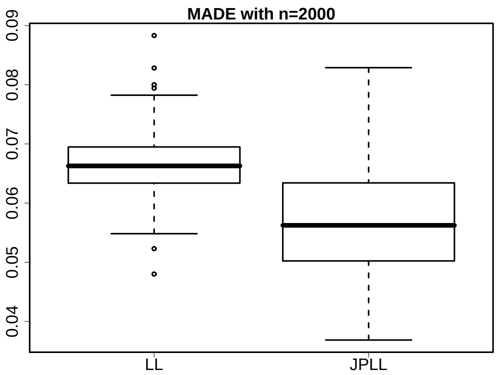
(b) $MADE_{q,k}$ with $n = 500$



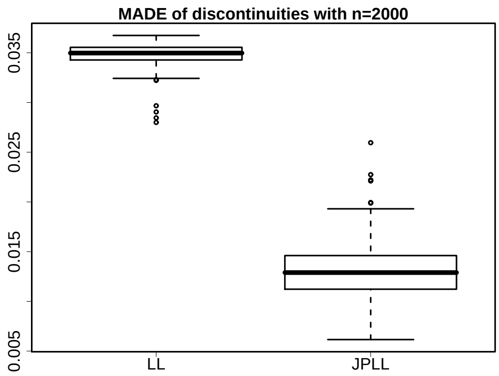
(c) $MADE$ with $n = 1000$



(d) $MADE_{q,k}$ with $n = 1000$

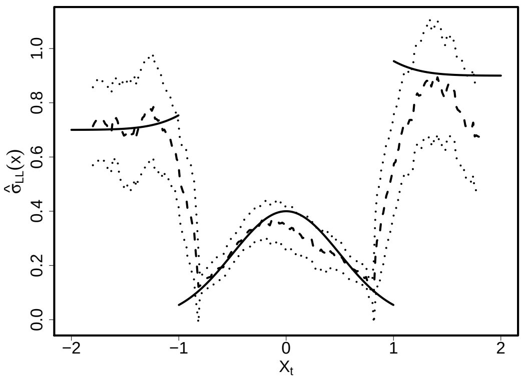


(e) $MADE$ with $n = 2000$

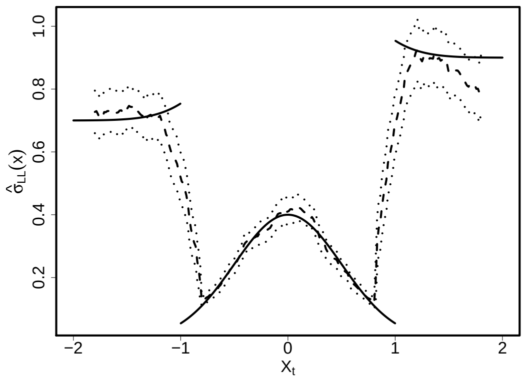


(f) $MADE_{q,k}$ with $n = 2000$

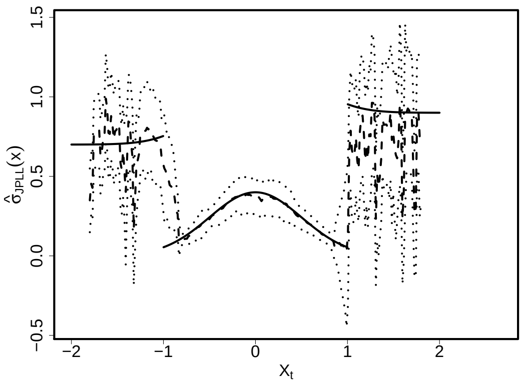
Figure 1: Boxplot of the MADE for scenario IV of Experiment 1.



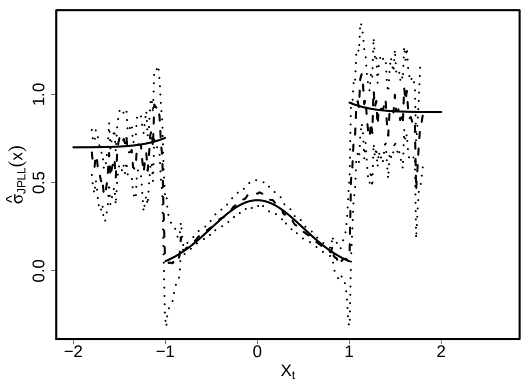
(a) LL estimator with $n = 500$



(b) LL estimator with $n = 2000$

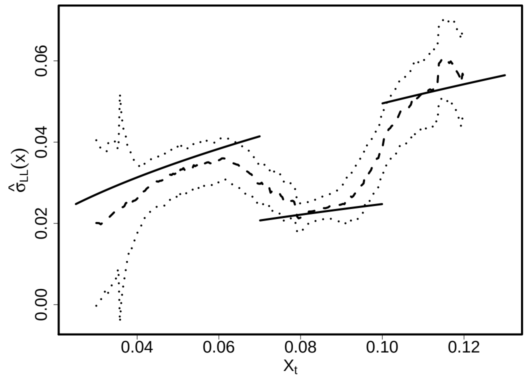


(c) JPLL estimator with $n = 500$

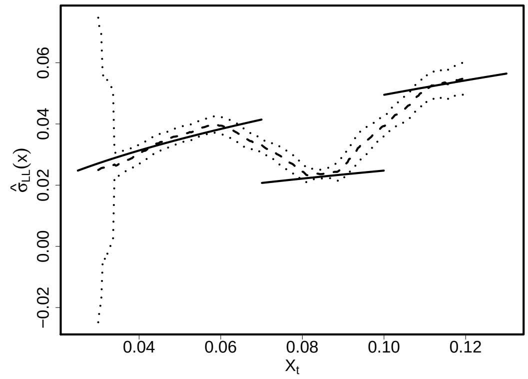


(d) JPLL estimator with $n = 2000$

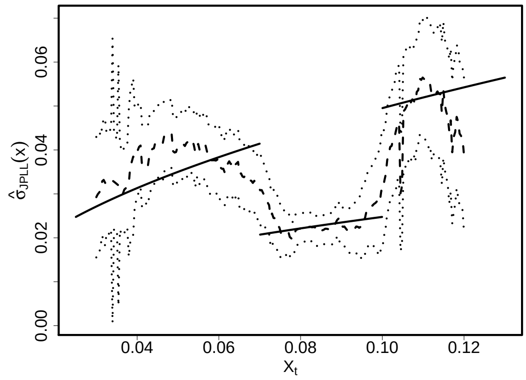
Figure 2: Comparison between the LL and the BPLL estimators from Scenario II of Experiment 1. The real function $\sigma(x)$ is plotted with a solid line, the estimate whose MADE correspond to the 50% quantile is plotted with a dashed line, and the 5% and the 95% confidence interval is plotted with a dotted line.



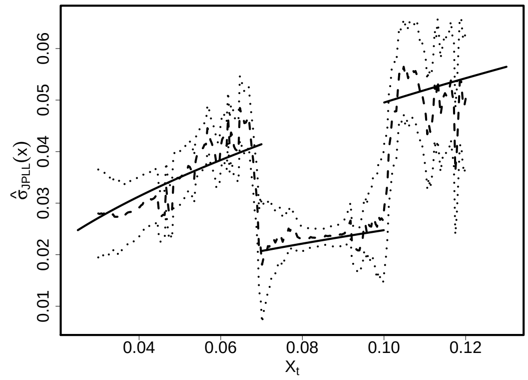
(a) LL estimator with $n = 500$



(b) LL estimator with $n = 2000$

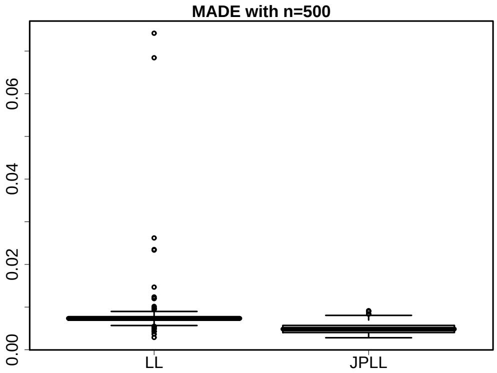


(c) JPLL estimator with $n = 500$

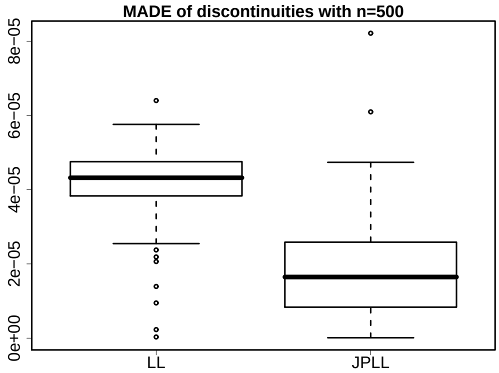


(d) JPLL estimator with $n = 2000$

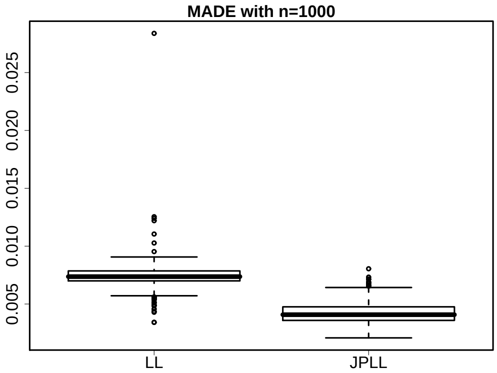
Figure 3: Comparison between the LL and the BPLL estimators of the CIR model (Experiment 2). The real function $\sigma(x)$ is plotted with a solid line, the estimate whose MADE correspond to the 50% quantile is plotted with a dashed line, and the 5% and the 95% confidence interval is plotted with a dotted line.



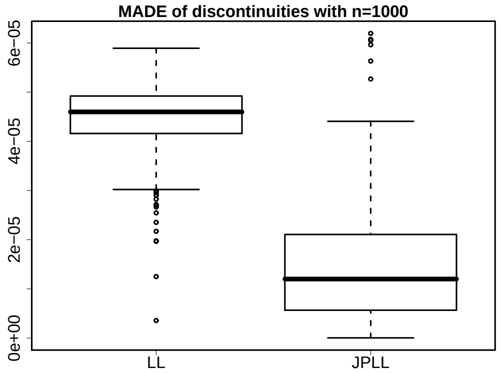
(a) $MADE$ with $n = 500$



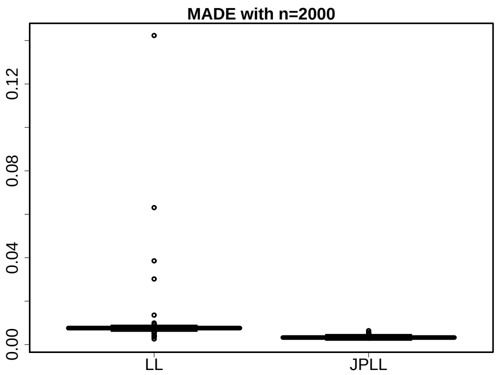
(b) $MADE_{q,k}$ with $n = 500$



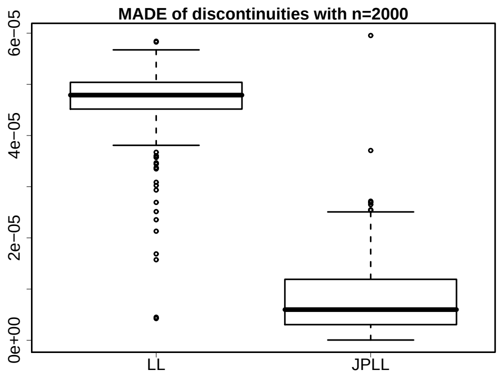
(c) $MADE$ with $n = 1000$



(d) $MADE_{q,k}$ with $n = 1000$



(e) $MADE$ with $n = 2000$



(f) $MADE_{q,k}$ with $n = 2000$

Figure 4: Boxplot of the MADE for Experiment 2.

Research Papers 2009

CREATES

Center for Research in Econometric Analysis of Time Series

2009-35: Peter Christoffersen, Jeremy Berkowitz and Denis Pelletier: Evaluating Value-at-Risk Models with Desk-Level Data

2009-36: Tom Engsted and Thomas Q. Pedersen: The dividend-price ratio does predict dividend growth: International evidence

2009-37: Michael Jansson and Morten Ørregaard Nielsen: Nearly Efficient Likelihood Ratio Tests of the Unit Root Hypothesis

2009-38: Frank S. Nielsen: Local Whittle estimation of multivariate fractionally integrated processes

2009-39: Borus Jungbacker, Siem Jan Koopman and Michel van der Wel: Dynamic Factor Models with Smooth Loadings for Analyzing the Term Structure of Interest Rates

2009-40: Niels Haldrup, Antonio Montañés and Andreu Sansó: Detection of additive outliers in seasonal time series

2009-41: Dennis Kristensen: Pseudo-Maximum Likelihood Estimation in Two Classes of Semiparametric Diffusion Models

2009-42: Ole Eiler Barndorff-Nielsen and Robert Stelzer: The multivariate supOU stochastic volatility model

2009-43: Lasse Bork, Hans Dewachter and Romain Houssa: Identification of Macroeconomic Factors in Large Panels

2009-44: Dennis Kristensen: Semiparametric Modelling and Estimation: A Selective Overview

2009-45: Kim Christensen, Silja Kinnebrock and Mark Podolskij: Pre-averaging estimators of the ex-post covariance matrix

2009-46: Matias D. Cattaneo, Richard K. Crump and Michael Jansson: Robust Data-Driven Inference for Density-Weighted Average Derivatives

2009-47: Mark Podolskij and Mathias Vetter: Understanding limit theorems for semimartingales: a short survey

2009-48: Isabel Casas and Irene Gijbels: Unstable volatility functions: the break preserving local linear estimator