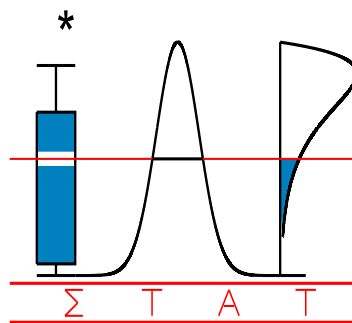


T E C H N I C A L  
R E P O R T

0634

**NONPARAMETRIC DENSITY ESTIMATION  
FOR POSITIVE TIME SERIES**

BOUEZMARNI, T. and J.V.K. ROMBOUTS



I A P S T A T I S T I C S  
N E T W O R K

**INTERUNIVERSITY ATTRACTION POLE**

# Nonparametric Density Estimation for Positive Time Series

Taoufik Bouezmarni\*

Jeroen V.K. Rombouts<sup>†</sup>

September 21, 2006

## Abstract

The Gaussian kernel density estimator is known to have substantial problems for bounded random variables with high density at the boundaries. For i.i.d. data several solutions have been put forward to solve this boundary problem. In this paper we propose the gamma kernel estimator as density estimator for positive data from a stationary  $\alpha$ -mixing process. We derive the mean integrated squared error, almost sure convergence and asymptotic normality. In a Monte Carlo study, where we generate data from an autoregressive conditional duration model and a stochastic volatility model, we find that the gamma kernel outperforms the local linear density estimator. An application to data from financial transaction durations, realized volatility and electricity price data is provided.

*Key words:* Gamma kernel, Nonparametric density estimation, Mixing process, Transaction durations, Realised volatility.

We would like to thank Luc Bauwens, Song Chen, Luc Devroye, Bruno Remillard, Jean-Marie Rolin and Ingrid Van Keilegom for useful comments.

---

\*HEC Montréal, CREF, Institute of Statistics (Université catholique de Louvain).

<sup>†</sup>Institute of Applied Economics at HEC Montréal, CIRANO, CIRPEE, CORE (Université catholique de Louvain), CREF, 3000 Cote Sainte Catherine, H3T 2A7 Montréal (QC), Canada.

# 1 Introduction

An important first step in applied econometric modelling is looking at the characteristics of the data at hand without making yet strong assumptions. Apart from the usual descriptive statistics, we also estimate univariate marginal densities of the variables of interest to examine for example the tails or to check if there is some multimodality in the density. This density estimation is usually done using a Gaussian kernel estimator of which the asymptotic properties are well established for i.i.d. data and for time series data, see for example Silverman (1986), Pagan and Ullah (1999) and Fan and Yao (2003) for textbook details. Another advantage of the Gaussian kernel estimator is the direct availability in most statistical software packages. A drawback is that it may give positive weight to points outside the support when we estimate the density of a bounded random variable. This causes a bias problem at the boundary and therefore inconsistency of the estimator. This problem, well known as the boundary bias problem, has given rise to new methods and new kernel estimators for i.i.d. data.

First solutions to this boundary problem are the reflection method, see Schuster (1985) and the local renormalisation method, see Diggle (1985) and Härdle (1990). The idea behind the reflection method is to recapture the missing weight by reflecting the estimate in the boundary. The renormalisation method uses the Gaussian estimator with renormalized kernels. Although both methods lower the bias, there is still more bias at the boundary compared to the interior of the support. To obtain the same bias order over the whole support Devroye and Györfi (1985) and Marron and Ruppert (1994), apply first a data transformation such that the first derivative of the density of the transformed data is equal to zero at the boundary points. Second, they use the reflection method to estimate the density function of the transformed data which is thereafter transformed to the density of the original data. In order to get the same order of bias without transforming the data, several authors propose to use an adapted kernel at the boundary region and the standard kernel in the interior region. See Jones (1993) for the boundary kernel estimator, Müller (1991) for optimal boundary kernel estimators and Lejeune and Sarda (1992) for the local linear estimator. The drawback of these estimators is that they can take negative values near the boundary. To solve this problem, Jones and Foster (1996) propose a simple nonnegative boundary corrected estimator. A recent solution is to use asymmetric and adapted kernels which do not assign any weight outside the support. Chen (1999) and Chen (2000) propose respectively beta kernel

densities for variables with compact support and gamma densities as kernels for variables with positive support. Chen (2002) also uses the gamma kernel in local linear smoothing to estimate a regression curve with a positive support. Fernandez and Monteiro (2005) establish a central limit theorem for gamma kernel functionals.

For economics data, it happens often that the range of the variable of interest is not the whole real line. This does not necessarily poses a problem for the Gaussian kernel estimator because the boundary can sometimes be ignored in practice. For example in the i.i.d. framework, income data from the computer industry are positive but there are no values near the zero boundary and therefore the boundary problem does not exist practically. However, income data for a country can have most of the density mass near zero because of high unemployment and/or high income inequality. Bouezmarni and Scaillet (2003), who derive weak convergence of the integrated absolute error for the gamma kernel estimator, illustrate this for an income density using data from Brazil.

For positive dependent data, there are also many illustrations in economics where there is an important concentration close to zero. Financial transaction data are typically highly dependent and often close or approximately equal to zero for frequently traded stocks. In the autoregressive conditional duration literature Engle and Russell (1998) analyse IBM trade duration data which a substantial concentration near zero. See Bauwens, Giot, Grammig, and Veredas (2004) for similar characteristics on other US stock data and Ghysels, Gouriéroux, and Jasiak (2004) for Alcatel traded on the Paris Stock Exchange. Another important example for this type of data is daily volatility series constructed using intraday data using realized volatility techniques. Andersen, Bollerslev, Diebold, and Labys (2001) investigate the shape of the unconditional distribution of realized exchange rate volatility and find that the distribution is extremely skewed. See Andersen, Bollerslev, Diebold, and Ebens (2001) for similar results on stock returns. A non finance example is efficient frontier analysis where inefficiency scores for a company related to its peer companies are computed over time, see Park and Simar (1994) for an application using railway company data.

The objective of this paper is to propose the gamma kernel estimator for positive time series. To do this, we build on Chen (2000) who introduces the gamma kernel estimator for i.i.d. data . We derive the convergence properties for data series from a stationary  $\alpha$ -mixing process. We compute the mean integrated squared error and obtain the optimal rate of convergence. We also prove the almost sure convergence and state the asymptotic normality for the gamma kernel estimator. The asymptotic normal distribution can be used for confidence intervals as we show in some applications.

In a detailed Monte Carlo study, where we generate data from an autoregressive conditional duration model and a stochastic volatility model, we show that, in terms of mean integrated squared error, the gamma kernel outperforms its closest competitor which is the local linear density estimator of Jones (1993).

The paper is organised as follows. The gamma kernel estimator for positive time series is introduced in Section 2. Section 3 provides convergence properties. In Section 4 we investigate the finite sample properties of the gamma kernel estimator for data from a conditional duration model and a stochastic volatility model. Both models are frequently used in financial econometrics. Section 5 first provides an application to transaction duration data for four US stocks traded on the New York Stock Exchange, the second application uses realised volatility series from the DM/dollar exchange rate and IBM, and the last application estimates the density of hourly electricity prices. Section 6 concludes. The proofs of the theorems are gathered in the Appendix.

## 2 Gamma kernel estimator

We observe  $X_1, \dots, X_n$  from a stationary  $\alpha$ -mixing process with density function  $f$ . Our goal is to estimate the function  $f(t)$  nonparametrically for  $t \in [0, \infty)$ . The gamma kernel estimator is defined next.

$$\hat{f}(x) = \frac{1}{n} \sum_{j=1}^n K_{\rho_b(x), b}(X_j) \quad (1)$$

where the kernel  $K_b$  is defined by

$$K_{\rho_b(x), b}(t) = \frac{t^{\rho_b(x)-1} \exp(-t/b)}{b^{\rho_b(x)} \Gamma(\rho_b(x))} \text{ and } \rho_b(x) = \begin{cases} x/b & \text{if } x \geq 2b \\ \frac{1}{4}(x/b)^2 + 1 & \text{if } x \in [0, 2b). \end{cases} \quad (2)$$

For notational simplicity we write  $K_{x,b}$  instead of  $K_{\rho_b(x), b}$ . The shape of the gamma kernel and the amount of smoothing vary according to the position where the density is estimated, which implies that the gamma kernel estimator is an adaptive density estimator. This is a crucial difference with the Gaussian or other fixed symmetric kernel estimators. The support of the gamma kernel matches the support of the probability density function to be estimated, and therefore no weight is

lost when the density is estimated at the boundary region. The gamma kernel density estimator is easy to implement, free of boundary bias and always non-negative. It achieves the optimal rate of convergence, in the i.i.d. case, for the mean integrated squared error within the class of non-negative kernel density estimators. Furthermore, their variance reduces as the position where the smoothing is made moves away from the boundary. We show the importance of this feature when we deal with the asymptotic normality of the gamma kernel estimator. For i.i.d. data, Chen (2000) shows by simulations that, in term of mean squared error, the gamma kernel estimator outperforms the local linear estimator of Jones (1993), the non-negative estimator of Jones and Foster (1996) and the boundary kernel estimator of Müller (1991). Note that the last estimators are proposed to remove the boundary bias problem of the standard kernel estimator. Bouezmarni and Scaillet (2003) state the uniform weak consistency for the gamma kernel estimator on each compact set in the positive real line when  $f$  is continuous on its support and the weak convergence in terms of mean integrated absolute error. For unbounded densities at the origin, they investigate the performance of the gamma kernel density estimator through simulations, and prove that it converges in probability to infinity at the origin where the density is unbounded.

### 3 Convergence properties

As already mentioned, the boundary bias problem for the standard kernel density estimator of density functions defined on  $[0, +\infty)$  is still present when the data are dependent. In this section we prove that the gamma kernel estimator is free of boundary bias in the case of dependant data. We find that, under mild conditions, it is possible to obtain the same convergence rates as in the i.i.d. case. In fact, we prove that the gamma kernel estimator achieves the optimal rate of convergence for the mean integrated squared error within the class of non-negative kernel density estimators. We also state the asymptotic normality and almost sure convergence.

For the remainder we denote by  $f_i$  the joint density of  $(X_1, X_{1+i})$  and  $g_i(x, y) = f_i(x, y) - f(x)f(y)$ . We suppose that the process satisfies the following conditions

**Condition 1. (condition on  $g$ )**

*Suppose that  $g_i$  satisfies the Lipschitz's condition*

$$|g_i(\xi') - g_i(\xi)| \leq l \|\xi' - \xi\|; \quad \xi', \xi \in R^+ \times R^+ \text{ for some constant } l \quad (3)$$

We require that the process  $\{X_i\}_{i \geq 1}$  is  $\alpha$ -mixing. The sequence is  $\alpha$ -mixing if the mixing coefficient  $\alpha(i) \rightarrow 0$  as  $i \rightarrow \infty$ , where

$$\alpha(i) = \sup_k |P(A \cap B) - P(A)P(B)|,$$

$A \in \mathcal{F}_1^k(X)$  and  $B \in \mathcal{F}_{k+i}^k(X)$  where  $\mathcal{F}_i^k(X)$  is the  $\sigma$ -field of events generated by  $\{X_j, i \leq j \leq k\}$ . The concept of  $\alpha$ -mixing is omnipresent in time series analysis and is less restrictive than  $\beta$  and  $\rho$ -mixing. The following condition requires that the  $\alpha$ -mixing coefficient decreases geometrically.

**Condition 2. (conditions on mixing processes)**

$(X_t)$  is supposed to be  $\alpha$ -mixing such that

$$\alpha(i) \leq \gamma i^{-\beta}, i \geq 1 \tag{4}$$

for some positive constants  $\gamma$  and  $\beta$ .

The next theorem states the mean integrated square error of the gamma kernel estimator using conditions 1 and 2, and the same conditions on the bandwidth parameter as in the i.i.d. case. In the sequel we continue to denote the bandwidth parameter as  $b$  without stressing that it depends on the sample size  $n$ .

**Theorem 1. (mean integrated squared error)**

Suppose that  $f$  is twice continuously differentiable. Under conditions 1 and 2 with  $\beta > \frac{3}{2}$ , and further assume that

$$\lim_{n \rightarrow \infty} b = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} n b^{1/2} = +\infty. \tag{5}$$

Then the mean integrated squared error of the gamma kernel estimator is given by

$$E \left( \int_0^\infty (\hat{f}(x) - f(x))^2 dx \right) = b^2 \beta + n^{-1} b^{-1/2} V + o(b) + o(n^{-1} b^{-1/2}) \tag{6}$$

where

$$\beta = \frac{1}{4} \int_0^\infty x^2 f''^2(x) dx \quad \text{and} \quad V = \frac{1}{2\sqrt{\pi}} \int_0^\infty x^{-1/2} f(x) dx. \tag{7}$$

The optimal bandwidth which minimizes the asymptotic mean integrated squared error is

$$b^* = \left( \frac{V}{4\beta} \right)^{2/5} n^{-2/5}. \tag{8}$$

This leads to the optimal mean integrated squared error:

$$E \left( \int_0^\infty (\hat{f}(x) - f(x))^2 dx \right) \sim \frac{5}{4^{4/5}} V^{4/5} \beta^{1/5} n^{-4/5}. \quad (9)$$

■

As in the i.i.d. case, the gamma kernel estimator achieves the same optimal rate of convergence in terms of mean integrated squared error. Note that relative to the Gaussian kernel estimator, the gamma kernel estimator has a different optimal mean integrated squared error but has the same optimal mean squared error. Theorem 1 also provides the theoretical formula of the optimal bandwidth parameter which can not be used in practice since it depends on the unknown density function  $f$ . Several techniques, such as cross-validation and block bootstrap, are available to obtain a bandwidth in practice. In Section 5, we suppose as if the density function is gamma with parameters estimated by maximum likelihood, and then apply (8). We will refer to this as the gamma rule.

The asymptotic distribution, usually Gaussian, of density estimators is useful for goodness of fit and confidence intervals. The following theorem deals with the asymptotic normality of gamma kernel estimator.

**Theorem 2. (asymptotic normality)**

Let  $f$  be a density function twice continuously differentiable. Under condition 2 with  $\beta > 2$  and  $b = O\left(\frac{1}{\text{LogLog}n} n^{-2/5}\right)$  we have for all  $x$  such that  $f(x) > 0$

$$\left( n^{1/2} b^{1/4} \frac{\hat{f}(x) - f(x)}{\sqrt{V^*(x)}} \right) \rightarrow N(0, 1) \quad (10)$$

where

$$V^*(x) = \begin{cases} \frac{f(x)}{2\sqrt{\pi x}} & \text{if } x > b \rightarrow \infty \\ \frac{\Gamma(2\kappa+1)b^{-1/2}}{2^{1+2\kappa}\Gamma^2(\kappa+1)} f(x) & \text{if } x/b \rightarrow \kappa \end{cases} \quad (11)$$

■

For the standard symmetric kernel, the asymptotic normality is proved for the same conditions when the boundedness of the support does not cause a problem. If we are interested in confidence



intervals for the density function, a good feature of the gamma kernel estimator is that the variance vanishes when the position is farther away from the origin, so that the confidence intervals narrow. The unknown density function  $f(x)$  in the expression of  $V^*$  can be replaced by  $\hat{f}(x)$ . The result of Theorem 2 remains valid thanks to the almost sure convergence of  $\hat{f}(x)$ . Indeed, Theorem 3 establishes the almost sure convergence of  $\hat{f}(x)$ .

**Theorem 3. (almost sure convergence)**

Let  $f$  be a continuous density. Under the condition 2 with  $\beta > 2$  and following conditions on the bandwidth parameter:

$$\lim_{n \rightarrow \infty} b = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{n b^{1/2}}{(\log n)^2} = +\infty$$

we have

$$\hat{f}(x) \longrightarrow f(x), \quad \text{almost surely.}$$

■

## 4 Finite sample properties

We investigate the performance of the gamma kernel estimator for data generated from an autoregressive conditional duration (ACD) model and a stochastic volatility (SV) model. Both models are frequently used in financial econometrics and typically exhibit relatively high dependence when fitted to high frequency financial data. The performance of the gamma kernel estimator is compared to the local linear density estimator of Lejeune and Sarda (1992) and Jones (1993) which is defined next. For non-negative integer  $s$  and any symmetric kernel  $K$  with a compact support  $[-1, 1]$ , define

$$a_s(x, h) = \int_{-1}^{x/h} t^s K(t) dt. \tag{12}$$

The local linear estimator is defined as follows

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^T K_l \left( x, h, \frac{x - X_i}{h} \right) \tag{13}$$

where

$$K_l(x, h, t) = \frac{a_2(x, h) - a_1(x, h)t}{a_0(x, h)a_2(x, h) - a_1^2(x, h)} K(t) \tag{14}$$

is a local linear kernel used at  $x$ . Note that  $K_l(x, h, t) = K(t)$  for  $x > h$ . Simulation results of Jones (1993) and Chen (2000) show that the local linear estimator outperforms the boundary kernel estimator of Müller (1991) and the non-negative estimator Jones and Foster (1996). The latter estimators are therefore not considered here.

For each replication in the simulation, the bandwidth parameter is selected by minimizing the  $L_2 = \int (\hat{f}(t) - f(t))^2 dt$  norm. Since the underlying density  $f$  is unknown for the ACD and the SV models, we consider the empirical density function based on 500000 observations. For each model we perform 500 replications for different degrees of dependence and for sample sizes 125, 250, 500, 1000 and 2000.

#### 4.1 Autoregressive conditional durations

The analysis of durations between financial transactions is of central importance in the understanding of financial markets microstructure. Engle and Russell (1998) propose a new class of point processes with dependent arrival rates. They model the expected duration between events by an autoregressive structure and call it therefore the autoregressive conditional duration (ACD) model. Let the duration be defined as  $x_i = t_i - t_{i-1}$ , that is the interval between two arrival times. Their ACD model is defined as

$$\begin{aligned} x_i &= \psi_i \epsilon_i \\ \psi_i &= \omega + \alpha x_{i-1} + \beta \psi_{i-1} \end{aligned} \tag{15}$$

where the innovations  $\epsilon_i, i = 1, \dots, n$  are i.i.d random variables. Given a density assumption on  $\epsilon_i$ , for example the generalized gamma or the Burr distributions, the model is estimated easily by maximum likelihood. Inference on a semiparametric model, where the distribution of  $\epsilon_i$  is determined only by the data sample, is provided by Drost and Werker (2004). The unconditional distribution of  $x_i$  is unknown but expressions for the unconditional moments are available, see Bauwens and Giot (2001), Chapter 3, and Fernandez and Grammig (2006). For extensions and applications of the ACD model, see for example, Engle (2000), Hamilton and Jorda (2002), Bauwens and Veredas (2004) and Bauwens, Giot, Grammig, and Veredas (2004).

We consider three innovation densities, the standard exponential, the Weibull with shape and scale parameters equal to 0.91 and 1 respectively, and the Gamma with shape and scale parameters equal to 2 and .05 respectively. As can be seen in Figure 1, the exponential and Weibull densities

have a peak at zero, while the Weibull density is unbounded at zero. We consider three levels of

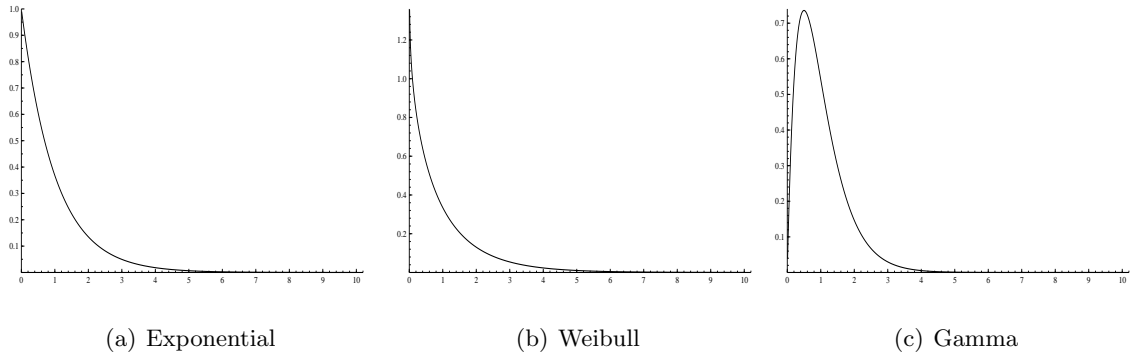


Figure 1: Innovation densities.

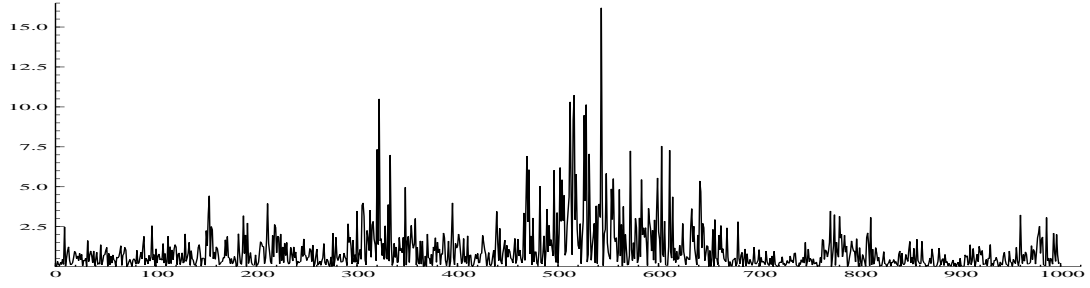
dependence: high, low and independence. The parameter values for these levels are given in Table 1. The parameters for the high dependence case are rounded values from Fernandez and Grammig

Table 1: Parameter values for the ACD model

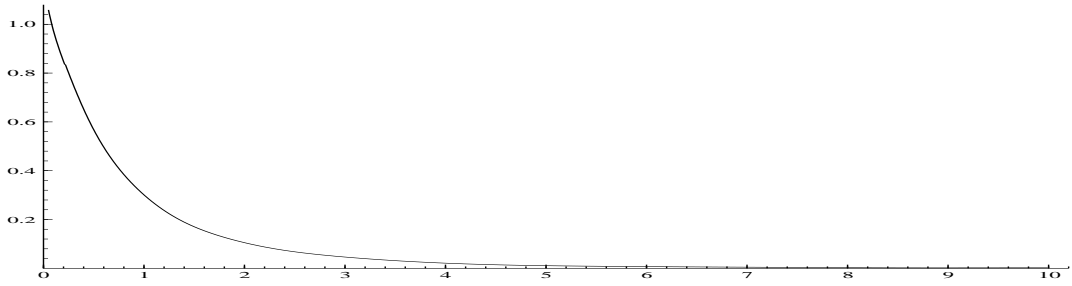
		$\omega$	$\alpha$	$\beta$
I	high dependence	0.017	0.11	0.88
II	low dependence	0.153	0.11	0.80
III	independence	1.700	0	0

(2005) who estimate the ACD model on NYSE price duration data. In general, the choice of  $\alpha$ ,  $\beta$  and the parameter of the density function of  $\epsilon_i$  obey  $|\beta| < 1$  and  $E(\beta + \alpha\epsilon_i)^2 < 1$ , such that the process is  $\beta$ -mixing with exponential decay. See also Carasco and Chen (2002) for more mixing properties of GARCH and ACD models.

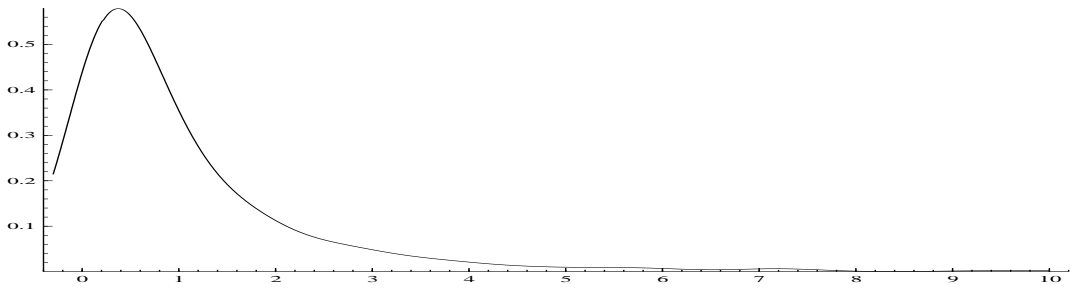
By including the limiting independence case we can appreciate how the performance of the gamma kernel changes when dependence is introduced. Figure 2 displays in the first panel a simulated sample path of the high dependence model with exponential innovation density for 1000 observations. The second panel shows the gamma kernel estimator which, as it should be, has a peak at zero. The bandwidth parameter which minimizes the integrated squared error is equal to 0.1064. As a comparison, the last panel displays the normal kernel estimator, using the rule of thumb to obtain the bandwidth, for which we immediately observe the boundary problem.



(a) Sample path from ACD model



(b) Gamma kernel estimator



(c) Gaussian kernel estimator

Figure 2: ACD model with high dependence and exponential innovation density.

Table 2 shows the mean of the  $L_2$  norm for the gamma (G) and the local linear (LL) estimators for ACD Models. For each replication, the bandwidth parameter is chosen such that the integrated squared error is minimized. As expected, the mean norm decreases when the sample size rises, and this holds for any dependence level. For any sample size, there is a positive relation between the mean norm and the dependence level. Therefore, minimizing the  $L_2$  norm is more difficult in the dependent data case. In most situations, the gamma kernel estimator dominates the local linear estimator. The difference is more striking when the sample size is small and when the dependence becomes high. For example, for the sample size of 125, the mean norm in the high dependence case for the gamma kernel estimator is 0.0374 and 0.0515 for the local linear estimator, compared to only 0.0077 and 0.0089 in the medium dependence case. This difference is less pronounced for the exponential innovation density in the independent case.

Table 2: Mean of  $L_2$  norm for gamma (G) and local linear (LL) estimators for ACD Models

		$n = 125$		$n = 250$		$n = 500$		$n = 1000$		$n = 2000$	
		G	LL	G	LL	G	LL	G	LL	G	LL
A	I	0.0374	0.0515	0.0267	0.0339	0.0175	0.0207	0.0114	0.0148	0.0072	0.0078
	II	0.0077	0.0089	0.0048	0.0050	0.0030	0.0029	0.0020	0.0020	0.0013	0.0013
	III	0.0048	0.0046	0.0032	0.0030	0.0021	0.0019	0.0014	0.0014	0.0010	0.0009
B	I	0.0342	0.0437	0.0278	0.0349	0.0197	0.0244	0.0137	0.0164	0.0095	0.0113
	II	0.0093	0.0103	0.0057	0.0066	0.0039	0.0046	0.0027	0.0033	0.0021	0.0026
	III	0.0061	0.0064	0.0040	0.0045	0.0028	0.0033	0.0021	0.0025	0.0017	0.0021
C	I	0.0345	0.0439	0.0277	0.0314	0.0171	0.0181	0.0115	0.0115	0.0060	0.0067
	II	0.0092	0.0102	0.0058	0.0065	0.0037	0.0043	0.0026	0.0031	0.0019	0.0023
	III	0.0063	0.0070	0.0042	0.0049	0.0030	0.0036	0.0022	0.0027	0.0017	0.0021

I: high dependence, II: low dependence, III: independent case. The innovations are drawn from exponential (A), Weibull (B) and gamma (C) densities. For each of the 500 replications, the bandwidth parameter is chosen such that the integrated squared error is minimized.

Table 3 shows the variances of the  $L_2$  norm for the gamma and the local linear estimators for ACD Models. As expected, there is a negative relation between the size of the variance of the two estimators and the sample size. Also, the variance increases for higher dependence levels. Except

for the high dependence case and sample size 125, the variance of the two estimators are similar for the Gamma innovation density. For the exponential and the Weibull innovation density the gamma estimator has a smaller variance than the local linear estimator. In particular, in the high dependence case the variance of local linear estimator is about twice the variance of gamma kernel estimator. As an example, for the Weibull innovation density and a sample size 500, the variance is respectively 4.41 and 2.40.

Table 3: Variance ( $\times 10^{-4}$ ) of  $L_2$  norm for gamma (G) and local linear (LL) estimators for ACD Models

		$n = 125$		$n = 250$		$n = 500$		$n = 1000$		$n = 2000$	
		G	LL	G	LL	G	LL	G	LL	G	LL
A	I	10.73	24.89	5.52	10.9	2.52	4.75	1.29	2.19	0.462	0.81
	II	0.324	0.818	0.108	0.18	0.0361	0.0576	0.0144	0.0225	0.0049	0.0081
	III	0.102	0.127	0.0361	0.04	0.0121	0.0144	0.0036	0.0064	0.0016	0.0016
B	I	4.959	10.70	4.41	8.06	2.40	4.41	1.53	2.65	0.7744	1.36
	II	0.442	0.775	0.152	0.324	0.0576	0.115	0.0361	0.0625	0.0169	0.029
	III	0.144	0.175	0.044	0.073	0.0169	0.0289	0.0081	0.0169	0.0049	0.01
C	I	12.11	21.72	8.82	11.4	3.24	3.88	1.51	1.66	0.640	0.624
	II	0.9133	0.956	0.25	0.28	0.096	0.108	0.036	0.04	0.0144	0.017
	III	0.222	0.225	0.078	0.0841	0.0361	0.040	0.0144	0.0169	0.0049	0.0064

I: high dependence, II: low dependence, III: independent case. The innovations are drawn from exponential (A), Weibull (B) and gamma (C) densities. For each of the 500 replications, the bandwidth parameter is chosen such that the integrated squared error is minimized.

## 4.2 Stochastic volatility

The stochastic volatility (SV) model, see for example Taylor (1986), is a popular alternative to the GARCH type volatility model. Estimation of the SV model is less straightforward than the GARCH model, the likelihood function involves a high dimensional integral, but it has direct links with finance theory such as option and bond pricing for example. As discussed by Harvey, Ruiz, and Shephard (1994) the SV model can be interpreted as a discrete-time approximation to the

continuous-time Ornstein-Uhlenbeck process which is often considered as the generating process for realised volatility. See Barndorff-Nielsen and Shephard (2001) or Barndorff-Nielsen and Shephard (2002) for more details on this. The basic SV model is given by

$$\begin{aligned} y_i &= \exp\left(\frac{x_i}{2}\right) \mu_i \\ x_i &= \gamma + \delta x_{i-1} + \nu \epsilon_i \end{aligned} \tag{16}$$

where  $\mu_i$  and  $\epsilon_i$  are i.i.d normal random variables and independent from each other. Various extensions and other formulations exist of the basic SV model, see Shephard (2005) for a collection of selected readings. Our interest lies in the marginal density of the variance.

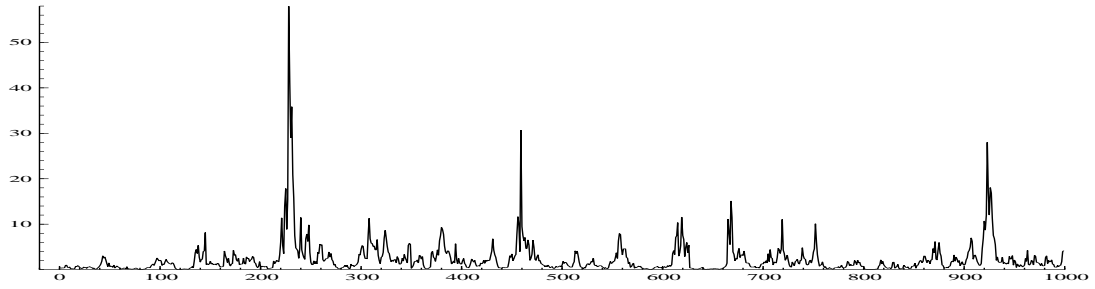
In the simulation study we consider two innovation densities for  $\mu_i$ , the normal and the student distribution with 5 degrees of freedom. As before, we consider three levels of dependence: high, low and independence, the parameters of which are given in Table 4. To illustrate the SV process,

Table 4: Parameter values for the SV model

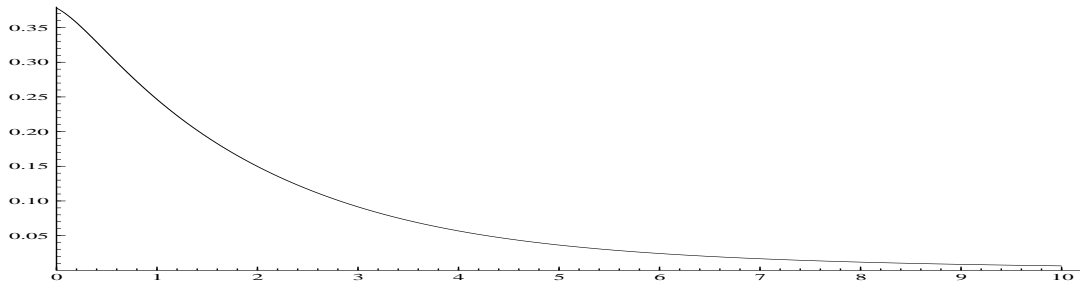
		$\gamma$	$\delta$	$\nu$
I	high dependence	0.01	0.98	0.3
II	low dependence	0.01	0.95	0.3
III	independence	1	0	0

Figure 3 displays in the first panel a simulated sample path of the high dependence SV model with normal innovation density for 1000 observations. The second panel shows the gamma kernel estimator which, as it should be, has a peak at zero. The bandwidth parameter which minimizes the integrated squared error is equal to 0.7324. The normal kernel estimator in the last panel suffers from the boundary problem as expected. Here, the bandwidth is obtained using the rule of thumb.

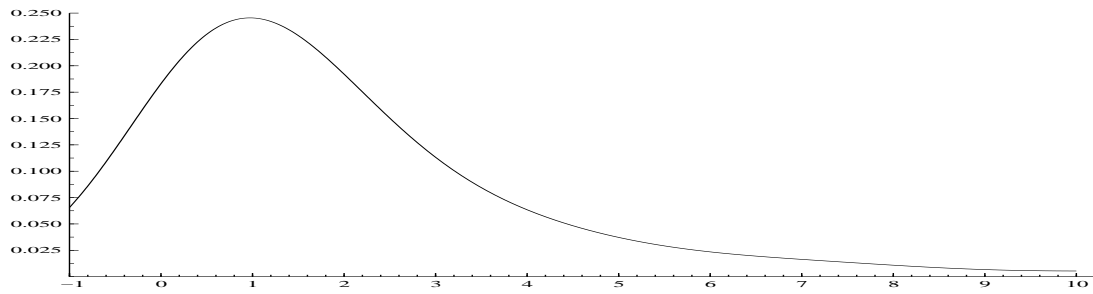
Table 5 contains the mean of the  $L_2$  error for the gamma kernel and local linear density estimators for the variance of the SV Model. Broadly speaking, the results are the same as the results of the ACD model in Section 4.1. The mean error decreases with the sample size at about the same rate as for the ACD models. However, the rate is higher for the local linear estimator. This is true for all dependence levels and for both innovation densities. The mean error increases for higher dependence levels no matter the innovation density and the sample size. The mean error of the



(a) Sample path from SV model



(b) Gamma kernel estimator



(c) Gaussian kernel estimator

Figure 3: SV model with high dependence and Gaussian innovation density.



gamma kernel estimator is remarkably lower than the local linear estimator in the high dependence case for sample sizes up to 500. For example, for 125 observations and a normal innovation density, the mean errors for the gamma kernel and the local linear estimator are respectively 0.0624 and 0.0926. For higher sample sizes, this difference becomes smaller. For 1000 observations the difference in the mean error is respectively 0.0130 and 0.0135.

Table 5: Mean of  $L_2$  error for gamma (G) and local linear (LL) estimators for the variance of the SV Model

		$n = 125$		$n = 250$		$n = 500$		$n = 1000$		$n = 2000$	
		G	LL	G	LL	G	LL	G	LL	G	LL
A	I	0.0624	0.0926	0.0453	0.0601	0.0313	0.0369	0.0205	0.0219	0.0130	0.0135
	II	0.0526	0.0562	0.0349	0.0361	0.0218	0.0214	0.0124	0.0118	0.0069	0.0065
	III	0.0089	0.0125	0.0055	0.0081	0.0036	0.0052	0.0023	0.0032	0.0015	0.0020
B	I	0.0748	0.1089	0.0578	0.0689	0.0405	0.0427	0.0251	0.0256	0.0150	0.0153
	II	0.0475	0.0539	0.0320	0.0346	0.0212	0.0227	0.0116	0.0128	0.0066	0.0069
	III	0.0076	0.0085	0.0045	0.0053	0.0028	0.0034	0.0016	0.0020	0.0010	0.0013

I: high dependence, II: low dependence, III: independent case. The data are drawn from A: normal and B: student densities. For each of the 500 replications, the bandwidth parameter is chosen such that the integrated squared error is minimized.

The variance of the  $L_2$  error for the gamma kernel and local linear density estimators for the variance of the SV Model are given in Table 6. The main conclusions are similar as for the ACD models of the previous section. The variance decreases with the sample size but at a rate that is slower than that for the ACD models. However, the rate of decrease in variance is higher for the local linear estimator. For example, in the high dependence case for the student innovation density the variance of the  $L_2$  error for the gamma kernel and local linear density estimators are respectively 13.875 and 64.717 for 125 observations and respectively 1.8521 and 1.9243 for 2000 observations. This observation holds true in general for all dependence levels and for both innovation densities. The variance increases for higher dependence levels, and for small sample sizes the difference between both density estimators becomes large. This difference is less strong for the ACD models.

Table 6: Variance ( $\times 10^{-4}$ ) of  $L_2$  error for gamma (G) and local linear (LL) estimators for the variance of the SV Model

		$n = 125$		$n = 250$		$n = 500$		$n = 1000$		$n = 2000$	
		G	LL	G	LL	G	LL	G	LL	G	LL
A	I	16.827	55.946	11.212	32.584	6.2972	14.434	2.8235	3.9522	1.1688	1.1901
	II	15.935	21.829	8.2979	9.2024	3.8092	3.9159	1.2030	1.0535	0.4139	0.3344
	III	0.2865	0.3892	0.1277	0.1562	0.0476	0.0635	0.0162	0.0192	0.0058	0.0059
B	I	13.875	64.717	12.702	30.508	9.3288	14.466	4.371	4.8185	1.8521	1.9243
	II	12.943	20.716	7.3572	9.0309	3.5174	3.5112	0.9473	0.9459	0.26535	0.2676
	III	0.2565	0.2245	0.0798	0.0897	0.0281	0.0345	0.0075	0.0094	0.00268	0.0028

I: high dependence, II: low dependence, III: independent case. The data are drawn from A: normal and B: student densities. For each of the 500 replications, the bandwidth parameter is chosen such that the integrated squared error is minimized.

## 5 Applications

In this section we use the gamma kernel to estimate some densities of positive financial time series data the type of which are frequently used in financial econometric modelling. We also compute pointwise confidence intervals based on the asymptotic results in Section 3. The first application is on financial durations. We take a subset of the duration data of Bauwens, Giot, Grammig, and Veredas (2004). They use the durations to compare several financial duration models via density forecasts. The second application is using realized volatility data series from an exchange rate and IBM. The last application estimates the density of hourly electricity prices.

### 5.1 Price durations

We use transaction data from New York Stock Exchange which trading process combines a market maker system and an order book system. Apart from an opening auction, trading is continuous from 9h30 to 16h. Data were extracted from the Trade and Quote (TAQ) database supplied by the NYSE. The data cover the months September, October and November 1996 and is available for

Boeing, Coca-Cola, Disney and Exxon which are actively traded Dow Jones stocks. In particular, we use price durations which are defined as the time interval needed to observe a cumulative change in the mid-price not less than a threshold that is set at 0.125. Price durations provide a measure of the instantaneous volatility of the quoted mid-quoted price process as explained in Engle and Russell (1998). For a more detailed description of the data and other types of durations, we refer to Bauwens, Giot, Grammig, and Veredas (2004).

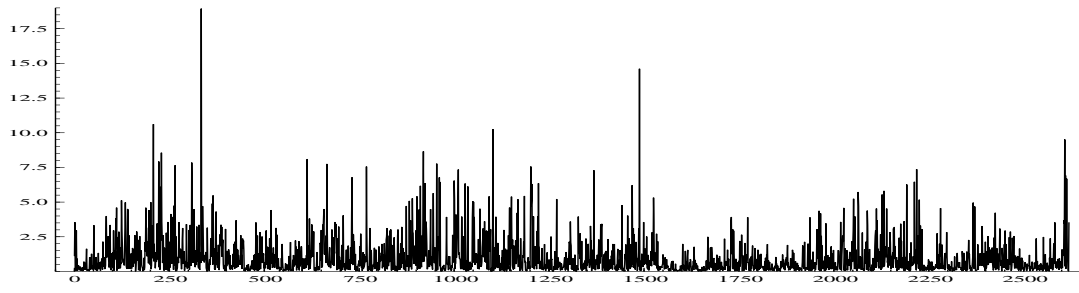
Figure 4 displays the durations series for the four stocks. The sample sizes differ for each stock and range between 1609 and 2717 observations many of which are small because the stocks are actively traded. We observe the typical characteristics of high frequency financial duration data for all the series. Fitting ACD models to the durations typically implies a high persistence,  $\alpha + \beta$  is close to one as explained in Section 4, because of the high dependence between the durations. See Bauwens, Giot, Grammig, and Veredas (2004) for a model comparison using density forecasts between several ACD models (different conditional mean and innovation density specifications) using the same data as in this paper. Figure 5 displays the gamma kernel density estimates for the four stocks. The bandwidth parameter is obtained using the gamma rule, as explained in Section 3. There is an important concentration of observations near the origin. This makes the Gaussian kernel estimator inadequate as already explained in Section 4. The shape of the four densities is quite similar, although we remark that the Boeing and Coke price durations are more concentrated at the origin than Disney and Exxon. The confidence intervals are close to the density estimates because of the large sample sizes and are inverse trumpet shaped according to the results in Theorem 2.

## 5.2 Realized volatility

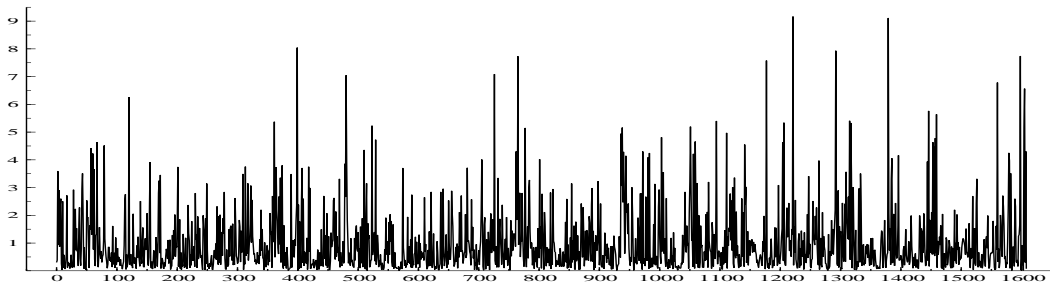
A trading day can be divided in  $N$  equidistant periods, if  $r_{i,t}$  denotes the intradaily return of the  $i$ th interval of day  $t$ , the realized volatility for day  $t$  is defined as

$$V_t = \sum_{i=1}^N r_{i,t}^2. \quad (17)$$

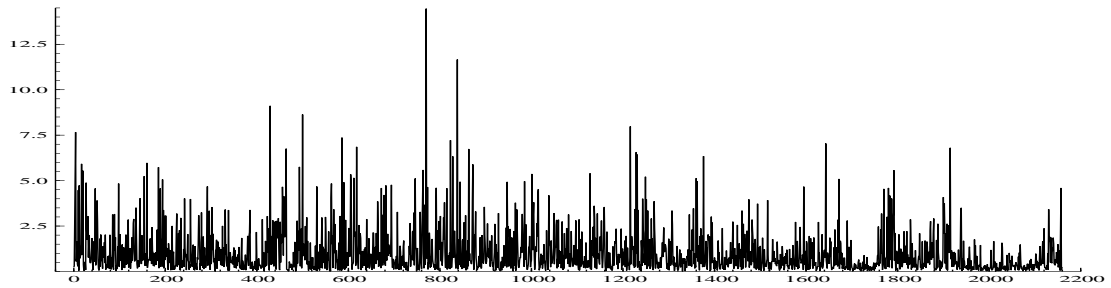
If the intraday returns have zero mean and are uncorrelated, then (17) is an unbiased estimator of volatility. Andersen, Bollerslev, Diebold, and Ebens (2001) show for the 30 stocks in the Dow Jones Industrial Average index that the unconditional distributions of the realized variances are extremely right-skewed and leptokurtic. See Andersen, Bollerslev, Diebold, and Labys (2003) and



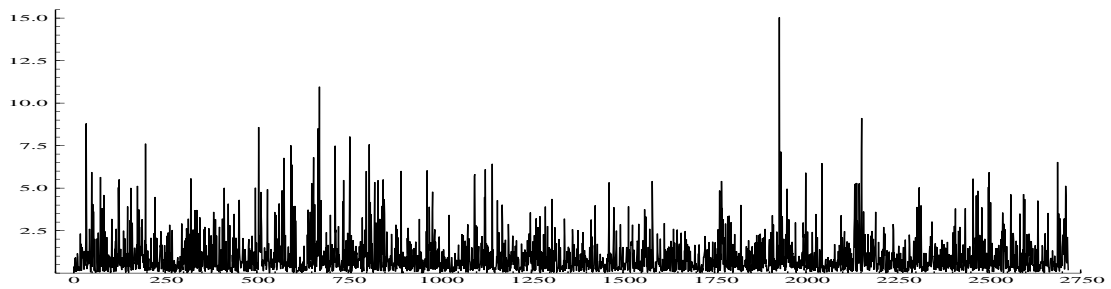
(a) Boeing Price (2620 obs)



(b) Coke Price (1609 obs)

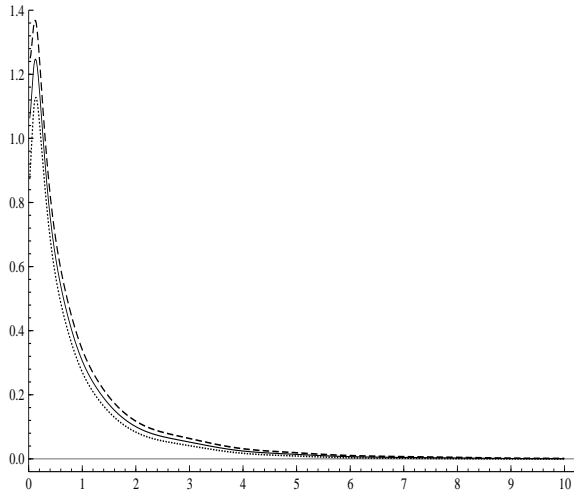


(c) Disney Price (2610 obs)

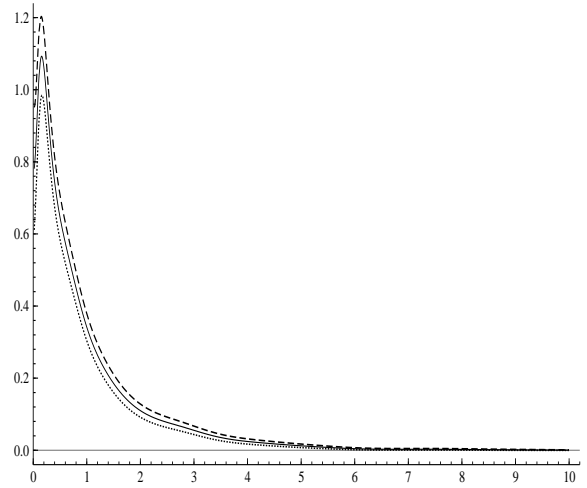


(d) Exxon Price (2717 obs)

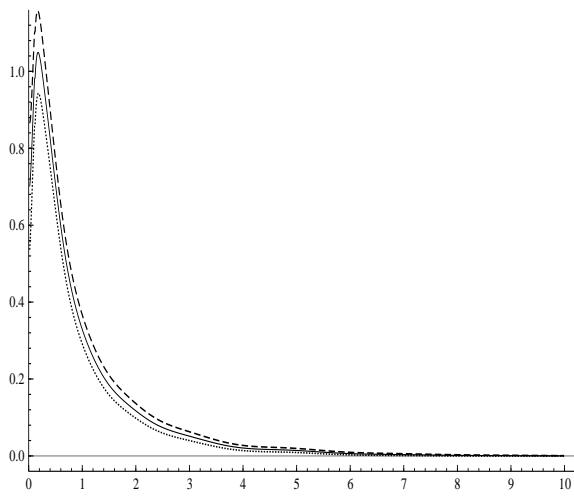
Figure 4: Price durations of four US stocks.



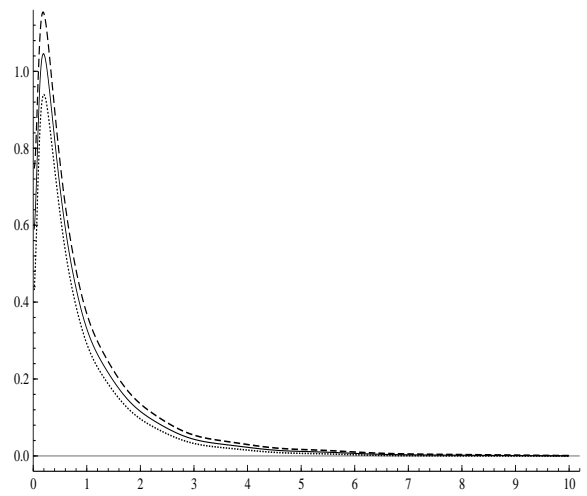
(a) Boeing Price



(b) Coke Price



(c) Disney Price

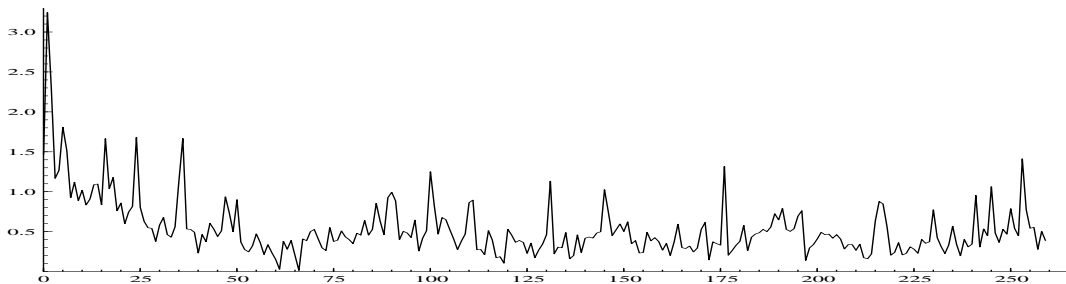


(d) Exxon Price

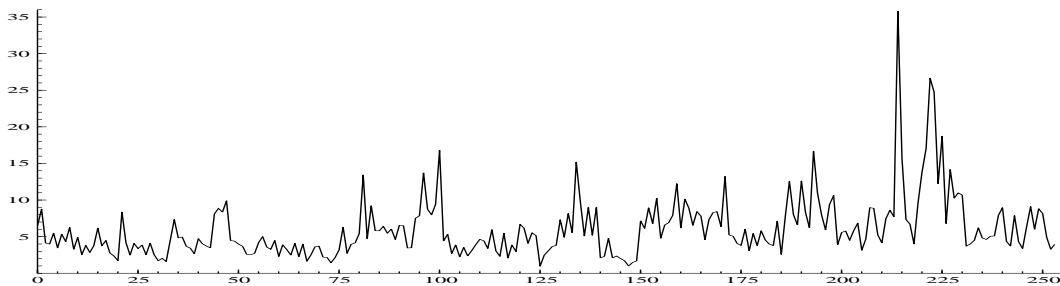
Figure 5: Density estimates (solid line) and 95% pointwise confidence intervals.

Andersen, Bollerslev, Diebold, and Labys (2001) for similar features on exchange rate data.

We use the DM/\$ exchange rate and IBM data from Hansen and Lunde (2005) which are available online from the Journal of Applied Econometrics Data Archive. The DM/\$ exchange rate data cover the period from October 1, 1992 to September 30, 1993 with  $N = 288$  or five minute intervals, which implies 260 daily volatility estimates. Similarly, the IBM dataset contains 254 daily volatility estimates from June 1, 1999 to May 31, 2000. Both realized volatility series are displayed in Figure 6. Figure 7 displays the kernel density estimates for the realized volatility



(a) Exchange rate data



(b) IBM data

Figure 6: Realized volatility series.

series. The bandwidth parameter is obtained using the gamma rule. Although the mode for the exchange rate data is closer to the origin than for the IBM data, the boundary problem for the latter is more apparent because of its relatively high dispersion. Since there are only about 250 realized volatilities, the lengths of the confidence intervals are larger than in the price durations application.

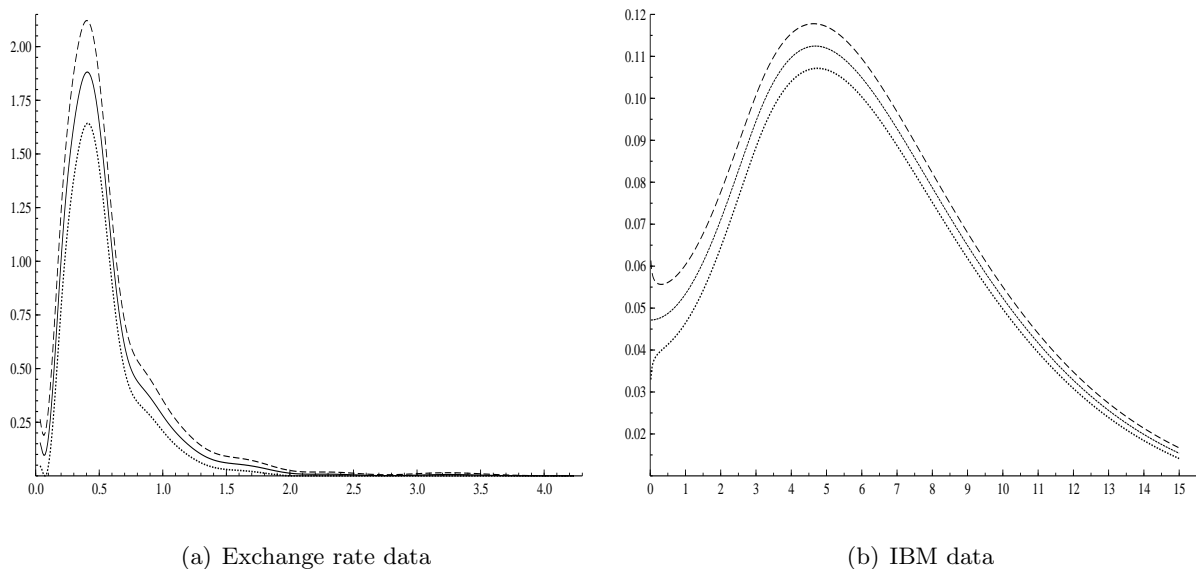


Figure 7: Density estimates (solid line) and 95% pointwise confidence intervals.

### 5.3 Electricity prices

The econometrics of electricity prices is becoming an active research field, see for example Fare, Grosskopf, Noh, and Weber (2005) and Haldrup and Nielsen (2006). The electricity data characteristics are often different from the well known stylized facts of financial return data. We have data available from ISO New England which is a regional transmission organization, serving Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island and Vermont. The company is an independent, not-for-profit corporation and is responsible for the constant availability of electricity for New England’s residents and businesses, The dataset consists of hourly reserve prices measured in \$/MWh from January 1, 2003 to January 14, 2003 ( $n=336$ ). As one can see in Figure 8, the price can become very close to zero, especially during the night, and peaks occur frequently many of which are around seven in the morning. The density estimate and pointwise confidence intervals are also given in Figure 8. The bandwidth parameter is obtained using the gamma rule.

## 6 Conclusion

There are many examples of economic time series data that are bounded and have a substantial density mass at the boundary. We propose the gamma kernel estimator as density estimator for positive data from a stationary  $\alpha$ -mixing process. The gamma kernel is flexible enough to vary

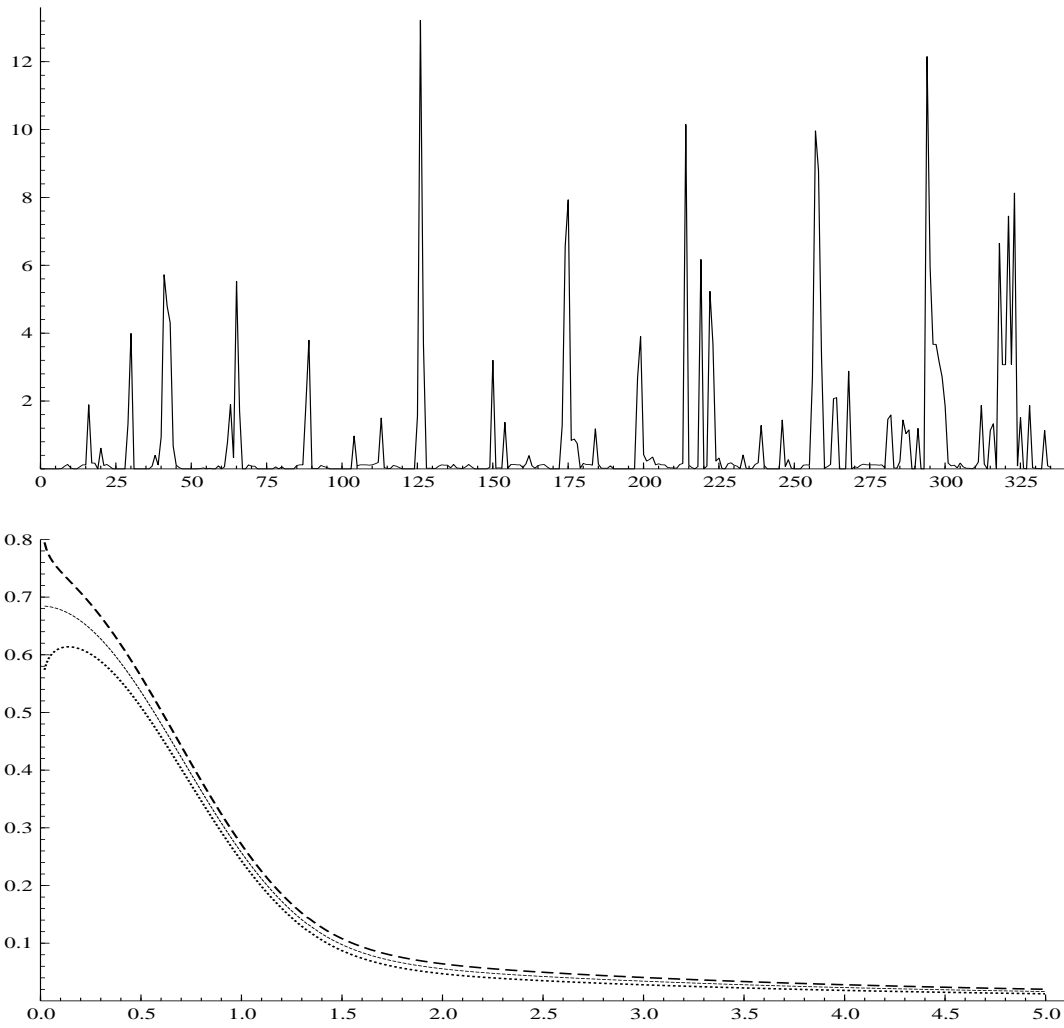


Figure 8: Electricity price data (upper panel) and density estimates (solid line) and 95% pointwise confidence intervals.



its shape and the amount of smoothing according to the location within the support. The gamma kernel estimator does not allocate weight outside the support of the underlying density and is therefore free of boundary bias. Furthermore, for positive time series not concentrated at zero the gamma kernel estimator has the same performance as the Gaussian kernel estimator. We derive the mean integrated squared error, almost sure convergence and asymptotic normality. In a Monte Carlo study, we find that the gamma kernel estimator outperforms the local linear density estimator. We provide an illustration using data from transaction durations, realized volatility and hourly electricity prices.

We think of several extensions of this paper. Firstly, the gamma kernel can be used to smooth the Kaplan Meier estimator in order to estimate the density and the hazard rate functions for  $\alpha$ -mixing and right censored data. Secondly, an extension to the bivariate case using either a copula or a bivariate gamma density is desirable, for example to estimate the density of realized multivariate volatility. Finally, it would be interesting to develop confidence bands for the gamma kernel estimator.

## Appendix

### Proof of Theorem 1

We start with the usual decomposition

$$\begin{aligned}
 E((\hat{f}(x) - f(x))^2) &= (E(\hat{f}(x)) - f(x))^2 + \frac{1}{n} \text{var}(K_{x,b}(X_j)) \\
 &+ \frac{2}{n} \sum_{i=1}^{n-1} (1 - n^{-1}i) \text{cov}((K_{x,b}(X_1), K_{x,b}(X_{1+i})) \\
 &= B^2(x) + V(x) + C(x).
 \end{aligned}$$

The expression of  $B(x)$  and  $V(x)$  can be derived from Chen (2000)

$$B(x) = b \left\{ \frac{1}{2} x f''(x) \right\} + o(b)$$

and

$$V(x) = \begin{cases} n^{-1} b^{-1/2} \frac{f(x)}{2\sqrt{\pi x}} + o(n^{-1} b^{-1/2}) & \text{if } x/b \rightarrow \infty \\ \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa} \Gamma^2(\kappa+1)} n^{-1} b^{-1} \{f(x) + O(n^{-1})\} & \text{if } x/b \rightarrow \kappa. \end{cases} \quad (18)$$

The terms  $B$  and  $V$  are the same as in the i.i.d case. Note that the variance increases at the boundary region but the integral of the variance is not affected by this. In fact, for  $\delta = b^{1-\epsilon}$ , where  $0 < \epsilon < 1/2$

$$\int_0^\infty V(x) dx = \int_0^\delta V(x) dx + \int_\delta^\infty V(x) dx = \frac{n^{-1} b^{-1/2}}{2\sqrt{\pi x}} \int_0^\infty \frac{f(x)}{x^{1/2}} dx + o(n^{-1} b^{-1/2}).$$

Now it remains to calculate the covariance term. From (3) and lemma (1.3) in Bosq (1996)

$$\text{cov}((K_b(t, X_1), K_b(t, X_{1+i}))) \leq \Gamma(l) \alpha(i)^{1/3} \quad (19)$$

for some constant  $\Gamma(l)$ .

On the other hand, from Billingsley's inequality

$$\text{cov}((K_b(t, X_1), K_b(t, X_{1+i}))) \leq 4 \|K_b\|_\infty^2 \alpha(i). \quad (20)$$

Using Stirling's formula and the fact that  $x$  is the mode of the kernel  $K_b(x)$

$$\|K_b\|_\infty^2 \leq \frac{1}{2\pi} x^{-1} b^{-1}. \quad (21)$$

Combining (19), (20) and (21), we have for some sequence  $v_n$

$$\begin{aligned}
\frac{n}{2}C(x) &\leq \Gamma(l) \sum_{i=1}^{v_n} (1 - n^{-1}i)\alpha^{1/3}(i) + \frac{2}{\pi}x^{-1}b^{-1} \sum_{i>v_n}^{n-1} (1 - n^{-1}i)\alpha(i) \\
&\leq \Gamma(l) \sum_{i=1}^{v_n} \alpha^{1/3}(i) + \frac{2}{\pi}x^{-1}b^{-1} \sum_{i>v_n}^{n-1} \alpha(i), \quad (1 - n^{-1}i) \leq 1 \\
&\leq \gamma\Gamma(l) \sum_{i=1}^{v_n} i^{-\beta/3} + \frac{2}{\pi}x^{-1}b^{-1}\gamma \sum_{i>v_n}^{n-1} i^{-\beta}, \quad \text{from (4)}.
\end{aligned}$$

We choose  $v_n \cong cb^{-\frac{3}{2\beta}}$ , we have the following results  $nb^{1/2}C(x) = o(1)$  since  $\beta > 3/2$ . This completes the proof of Theorem 1.

## Proof of Theorem 2

We start with

$$n^{1/2}b^{1/4} \left( E(\hat{f}(x)) - f(x) \right) = n^{1/2}b^{1/4}O(b) = O((\log \log(n))^{-5/4}).$$

It remains to show that

$$n^{1/2}b^{1/4} \left( \frac{\hat{f}(x) - E(\hat{f}(x))}{V^*(x)} \right) \rightarrow N(0, 1) \quad (22)$$

Without loss of generality we prove the result for  $x > 2b$ . Almost the same argument can be used to state the asymptotic normality at the boundary region.

For  $x > 2b$ , let for  $i = 1, \dots, n$ ,

$$Y_i = \frac{K_{x,b}(X_i) - EK_{x,b}(X_i)}{\sqrt{V^*(x)}}$$

and

$$S_n = \sum_{i=1}^n Y_i.$$

Note that

$$\left( n^{1/2}b^{1/4} \frac{\hat{f}(x) - E(\hat{f}(x))}{\sqrt{V^*(x)}} \right) = n^{-1/2}b^{1/4} S_n.$$

Now consider the blocks

$$V_1 = Y_1 + \dots + Y_p, \quad V_1^* = Y_{p+1} + \dots + Y_{p+q}, \dots$$

$$V_r = Y_{(r-1)(p+q)+1} + \dots + Y_{rp+(r-1)q}, \quad V_r^* = Y_{rp+(r-1)q+1} + \dots + Y_{r(p+q)},$$

where

$$r(p+q) \leq n \leq r(p+q+1).$$

Then

$$S_n = \sum_{i=1}^r V_i + \sum_{i=1}^r V_i^* + \delta_n.$$

If we choose  $r \sim n^a, p \sim n^{1-a}, q \sim n^c$ , where  $0 < a < 1$  and  $0 < c < 1 - a$ , we obtain

$$\text{var} \left( \frac{\sum_{i=1}^r V_i^*}{(rp)^{1/2} b^{-1/4}} \right) = O(n^{c-1+a})$$

and

$$\text{var} \left( \frac{\delta_n}{(rp)^{1/2} b^{-1/4}} \right) = O(n^{a-1}).$$

Then

$$(rp)^{-1/2} b^{1/4} \sum_{i=1}^r V_i^* \xrightarrow{p} 0.$$

and

$$(rp)^{-1/2} b^{1/4} \delta_n \xrightarrow{p} 0$$

On the other hand, from Bradley's lemma, there exist i.i.d random variables  $W_1, \dots, W_r$  such that  $P_{W_i} = P_{V_i}$  and

$$P(|W_i - V_i| > \epsilon_n) \leq 11 (\epsilon_n^{-1} \|V_i + c_n\|_\infty)^{1/2} \alpha(q) \quad (23)$$

where

$$c_n = \frac{2p b^{-1/2}}{f^{1/2}(x)(\pi x)^{1/4}}, \quad \epsilon_n = \epsilon (rp)^{1/2} b^{-1/4}. \quad (24)$$

Now we prove the asymptotic normality of

$$(rp)^{-1/2} b^{1/4} \sum_{i=1}^r W_i$$

and that

$$P \left( (rp)^{-1/2} b^{1/4} \left| \sum_{i=1}^r (W_i - V_i) \right| > \epsilon \right) \xrightarrow{p} 0. \quad (25)$$

We start with 25, from 24 and 23 we obtain

$$P \left( (rp)^{-1/2} b^{1/4} \left| \sum_{i=1}^r (W_i - V_i) \right| > \epsilon \right) = O(p^{1/4} r^{3/4} b^{-1/8} \alpha(q)).$$

Using the fact that  $\alpha(i) \leq \gamma i^{-\beta}, i \geq 1$  and  $b = \frac{c}{\log \log n} n^{-2/5}$ , we have

$$P \left( (rp)^{-1/2} b^{-1/4} \left| \sum_{i=1}^r (W_i - V_i) \right| > \epsilon \right) = O(n^{\frac{3}{10} + \frac{1}{2}a - \beta c} (\log \log n)^{1/8})$$

which converges to zero provided that  $c > \frac{1}{\beta}(\frac{1}{2}a + \frac{3}{10})$ .

To complete the proof, we state the asymptotic normality of

$$(rp)^{-1/2} b^{1/4} \sum_{i=1}^r W_i.$$

In fact it suffices to prove that

$$z_n = \frac{\sum_{i=1}^r |EW_i|^3}{(r \text{var}(W_1))^{3/2}} \longrightarrow 0$$

using Schwartz inequality and the boundedness of  $f_{t_1, \dots, t_4}$

$$\begin{aligned} z_n &\leq \frac{\left( \sum_{i=1}^r |EW_i|^4 \right)^{1/2} \left( \sum_{i=1}^r |EW_i|^2 \right)^{1/2}}{(r \text{var}(W_1))^{3/2}} \\ &= O(n^{\frac{4}{5} - \frac{3}{2}a} (\log \log n)^{-1/2}). \end{aligned}$$

A sufficient condition to insure the convergence of  $z_n$  to zero is  $a > 8/15$ . If we choose  $a < \frac{2\beta}{1+2\beta} - \frac{3}{5\beta} \frac{1}{1+2\beta}$ , then we can have  $0 < c < 1 - a$ ,  $a > 8/15$  and  $c > \frac{1}{\beta}(\frac{1}{2}a + \frac{3}{10})$  for  $\beta > 2$ .

### Proof of Theorem 3

The continuity of  $f$  at  $x$  ensure that

$$E(\hat{f}(x)) \rightarrow f(x).$$

Then it remains to prove that

$$|\hat{f}(x) - E(\hat{f}(x))| \rightarrow 0, \quad a.s.$$

For this we apply large inequality for dependent data (see Bosq (1996) ), to the random variable

$$Y_j = K_{x,b}(X_j) - E(K_{x,b}(X_j)), \quad 1 \leq j \leq n,$$

Throughout we denote by  $[\cdot]$  the integer part. Note also that

$$\|Y_j\|_{\infty} \leq \sqrt{\frac{2}{\pi}} x^{-1/2} b^{-1/2}.$$

Using this inequality and Theorem 1.3 in Bosq (1996), for an integer  $q = \lfloor n^{1/2}b^{-1/4} \rfloor$  we have,

$$\begin{aligned} P\left(|\hat{f}(x) - f(x)| > \epsilon\right) &\leq 4 \exp\left(-\frac{\epsilon^2}{8v^2(q)}q\right) \\ &+ 22 \left(1 + \sqrt{\frac{32}{\pi} \frac{x^{-1/2}b^{-1/2}}{\epsilon}}\right)^{1/2} q\alpha\left(\left[\frac{n}{2q}\right]\right) \end{aligned}$$

where

$$v^2(q) = \frac{2}{p^2}\sigma^2(q) + \frac{\epsilon\|Y_j\|_\infty}{2}$$

with  $p = \frac{n}{2q}$  and

$$\begin{aligned} \sigma^2(q) &= \max_{0 \leq j \leq 2q-1} E\left(\left([jp] + 1 - jp\right)Y_{[jp]+1} + Y_{[jp]+1} + \dots + Y_{[(j+1)p]+1} + \left([j+1)p+1 - (j+1)p\right)Y_{[(j+1)p]+1}\right)^2 \\ &= O(pb^{-1/2}). \end{aligned}$$

For  $\epsilon > 0$  small enough,

$$v^2(q) \leq \sqrt{\frac{2}{\pi}}x^{-1/2}b^{-1/2}\epsilon.$$

Therefore

$$\begin{aligned} P\left(|\hat{f}(x) - f(x)| > \epsilon\right) &\leq 4 \exp\left(-\frac{1}{8}\sqrt{\frac{\pi}{2}}\epsilon x^{1/2}n^{1/2}b^{1/4}\right) \\ &+ 22 \left(1 + \sqrt{\frac{32}{\pi} \frac{x^{-1/2}b^{-1/2}}{\epsilon}}\right)^{1/2} n^{1/2}b^{-1/4}\gamma\left(\left[\frac{1}{2}n^{1/2}b^{1/4}\right]\right)^{-\beta} \\ &\leq \xi \exp\left(-\zeta n^{1/2}b^{1/4}\right) \end{aligned}$$

for some positive constants  $\xi$  and  $\zeta$  depending on  $\epsilon$  and  $x$ .

Let  $u_n = \frac{nb^{1/2}}{(\log n)^2}$ , then

$$\sum_{n \geq 1} \xi \exp\left(-\zeta n^{1/2}b^{1/4}\right) = \sum_{n \geq 1} \frac{\xi}{n^{\zeta\sqrt{u_n}}} < \infty.$$

Therefore, Borel-Cantelli lemma concludes the proof the Theorem. 3

■

## References

ANDERSEN, T., T. BOLLERSLEV, F. DIEBOLD, AND H. EBENS (2001): “The Distribution of Realized Return Volatility,” *Journal of Financial Economics*, 61, 43–76.

- ANDERSEN, T., T. BOLLERSLEV, F. DIEBOLD, AND P. LABYS (2001): “The Distribution of Realized Exchange Rate Volatility,” *Journal of the American Statistical Association*, 96, 42–55.
- (2003): “Modeling and Forecasting Realized Volatility,” *Econometrica*, 71, 529–626.
- BARNDORFF-NIELSEN, O., AND N. SHEPHARD (2001): “Non-Gaussian Ornstein-Uhlenbeck based Models and some of their uses in Financial Economics,” *Journal of the Royal Statistical Society, Series B*, 63, 167–241.
- (2002): “Econometric Analysis of Realized Volatility and its use in Estimating Stochastic Volatility Models,” *Journal of the Royal Statistical Society, Series B*, 64, 253–280.
- BAUWENS, L., AND P. GIOT (2001): *Econometric Modelling of Stock Market Intraday Activity*. Kluwer Academic Publishers, Boston, Dordrecht, London.
- BAUWENS, L., P. GIOT, J. GRAMMIG, AND D. VEREDAS (2004): “A Comparison of Financial Duration Models Via Density Forecasts,” *International Journal of Forecasting*, 20, 589–609.
- BAUWENS, L., AND D. VEREDAS (2004): “The Stochastic Conditional Duration Model: A Latent Factor Model for the Analysis of Financial Durations,” *Journal of Econometrics*, 119, 381–412.
- BOSQ, D. (1996): *Nonparametric Statistics for Stochastic Processes, Estimation and Prediction*. Springer-Verlag, New York.
- BOUEZMARNI, T., AND O. SCAILLET (2003): “Consistency of Asymmetric Kernel Density Estimators and Smoothed Histograms with Application to Income Data,” *Econometric Theory*, 21, 390–412.
- CARASCO, M., AND X. CHEN (2002): “Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models,” *Econometric Theory*, 18, 17–39.
- CHEN, S. (1999): “A Beta Kernel Estimation for Density Functions,” *Computational Statistics and Data Analysis*, 31, 131–145.
- (2000): “Probability Density Functions Estimation Using Gamma Kernels,” *Annals of the Institute of Statistical Mathematics*, 52, 471–480.
- (2002): “Local linear smoothers using asymmetric kernels,” *Annals of the Institute of Statistical Mathematics*, 54, 312–323.

- DEVROYE, L., AND L. GYORFI (1985): *Nonparametric Density Estimation: The L1 View*. Wiley, New York, USA.
- DIGGLE, P. (1985): “A Kernel Method for Smoothing Point Process Data,” *Applied Statistics*, 34, 138–147.
- DROST, F., AND B. WERKER (2004): “Semiparametric Duration Models,” *Journal of Business and Economic Statistics*, 22, 40–50.
- ENGLE, R. (2000): “The Econometrics of Ultra High Frequency Data,” *Econometrica*, 68, 1–22.
- ENGLE, R., AND J. RUSSELL (1998): “Autoregressive Conditional Duration: A New Model for Irregularly spaced Transaction Data,” *Econometrica*, 66, 1127–1162.
- FAN, J., AND Q. YAO (2003): *Nonlinear Time Series*. Springer-Verlag, New York.
- FARE, R., S. GROSSKOPF, D. NOH, AND W. WEBER (2005): “Characteristics of a Polluting Technology: Theory and Practice,” *Journal of Econometrics*, 119, 381–412.
- FERNANDEZ, M., AND J. GRAMMIG (2005): “Nonparametric specification tests for conditional duration models,” *Journal of Econometrics*, 127, 35–68.
- (2006): “A Family of Autoregressive Conditional Duration Models,” *Journal of Econometrics*, 130, 1–23.
- FERNANDEZ, M., AND P. MONTEIRO (2005): “Central Limit Theorem for Asymmetric Kernel Functionals,” *Annals of the Institute of Statistical Mathematics*, 57, 425–442.
- GHYSELS, E., C. GOURIEROUX, AND J. JASIAK (2004): “Stochastic Volatility Duration Models,” *Journal of Econometrics*, 119, 413–433.
- HALDRUP, N., AND M. NIELSEN (2006): “A Regime Switching Long Memory Model for Electricity Prices,” *Journal of Econometrics*, 135, 349–376.
- HAMILTON, J., AND O. JORDA (2002): “A Model for the Federal Funds Rate Target,” *Journal of Political Economy*, 110, 1135–1167.
- HANSEN, P., AND A. LUNDE (2005): “A Forecast Comparison of Volatility Models: Does Anything Beat a GARCH(1,1)?,” *Journal of Applied Econometrics*, 20, 873–889.



- HÄRDLE, W. (1990): *Applied Nonparametric Regression*. Cambridge University Press, UK.
- HARVEY, A., E. RUIZ, AND N. SHEPHARD (1994): “Multivariate Stochastic Variance Models,” *Review of Economic Studies*, 61, 247–264.
- JONES, M. (1993): “Simple Boundary Correction for Kernel Density Estimation,” *Statistical Computing*, 3, 135–146.
- JONES, M., AND P. FOSTER (1996): “A Simple Nonnegative Boundary Correction Method for Kernel Density Estimation,” *Statistica Sinica*, 6, 1005–1013.
- LEJEUNE, M., AND P. SARDA (1992): “Smooth Estimators of Distribution and Density Functions,” *Computational Statistics and Data Analysis*, 14, 457–471.
- MARRON, J., AND D. RUPPERT (1994): “Transformations to reduce boundary bias in kernel density estimation,” *Journal of the Royal Statistical Society, Series B*, 56, 653–671.
- MÜLLER, H. (1991): “Smooth Optimum Kernel Estimators near Endpoints,” *Biometrika*, 78, 521–530.
- PAGAN, A., AND A. ULLAH (1999): *Nonparametric Econometrics*. Cambridge University Press, UK.
- PARK, B., AND L. SIMAR (1994): “Efficient Semiparametric Estimation in a Stochastic Frontier Model,” *Journal of the American Statistical Association*, 89, 929–936.
- SCHUSTER, E. (1985): “Incorporating Support Constraints into Nonparametric Estimators of Densities,” *Communications in Statistics - Theory and Methods*, 14, 1123–1136.
- SHEPHARD, N. (2005): *Stochastic Volatility: Selected Readings*. Oxford University Press, UK.
- SILVERMAN, B. (1986): *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, New York.
- TAYLOR, S. (1986): *Modelling Financial Time series*. Chicester: Wiley.