ANNALS OF ECONOMICS AND FINANCE **4**, 103–124 (2003)

Birnbaum-Saunders and Lognormal Kernel Estimators for Modelling Durations in High Frequency Financial Data

Xiaodong Jin

Department of Mathematics
UNC at Charlotte, Charlotte, NC, USA 28223
E-mail: xjin@math.uncc.edu

and

Janusz Kawczak

Department of Mathematics
UNC at Charlotte, Charlotte, NC, USA 28223
E-mail: jkawczak@math.uncc.edu

In this article we extend the class of non-negative, asymmetric kernel density estimators and propose Birnbaum-Saunders (BS) and lognormal (LN) kernel density functions. The density functions have bounded support on [0,∞). Both BS and LN kernel estimators are free of boundary bias, non-negative, with natural varying shape, and achieve the optimal rate of convergence for the mean integrated squared error. We apply BS and LN kernel density estimators to high frequency intraday time duration data. The comparisons are made on several nonparametric kernel density estimators. BS and LN kernels perform better near the boundary in terms of bias reduction.

Key Words: Birnbaum-Saunders kernel; Lognormal kernel; High frequency; ACD model; Durations.
JEL Classification Numbers: C13, C14, C15, C41.

1. INTRODUCTION

With the rocketing development of electronic equipment with tremendous storage capacity, high frequency or ultra high frequency tick-by-tick data are recorded everywhere in the world, such as financial markets, credit card services or any hi-tech companies employing scanning devices. The term "tick" is not in traditional sense of equally spaced time unit but a time stamp whenever there is a piece of quantity coming in such as trading

1529-7373/2002

price. Among various interesting components, time duration (TD) catches special attention from both researchers and practitioners. Let $X_i = t_i - t_{i-1}$ denote TD between two ticks or events occurring at times t_{i-1} and t_i. One example of TD studied later in this paper is the time interval between two consecutive transactions. Numerous papers are presented in modelling TD. Engle (1998) first advocates the autoregressive conditional duration (ACD) model. By adding in the conditional information, the parametric ACD models are capable of capturing the clustering phenomenon and serial dependence of TD. Fernandes and Grammig (2001) further bring in Box-Cox transformation to develop a family of ACD models. In order to select the best model out of ACD family and justify the choice, one needs to gauge the distance between the parametric probability density function (pdf) implied by TD and its non-parametric estimate. With this objective in mind, Fernandes and Grammig (2000) propose nice non-parametric D and H specification tests and recommend choosing the model with the minimum distance. One upcoming task invoking our main concern throughout this paper is how to obtain the most accurate non-parametric pdf estimate of X_i for the finite sample size.

Let $X_1, \ldots, X_n$ be a univariate random TD data from a distribution with an unknown pdf f with bounded support on $[0,\infty)$. The curve of this certain type of pdf declines sharply at the beginning and carries long tail at the end (namely DSLT). DSLT feature is clearly displayed in Table 1. Many methods have been created to handle the issue of boundary support. Schuster (1985) introduces data reflection and Silverman (1986) proposes negative-reflection, but they do not remove the asymptotic bias caused by the discontinuity in the first derivative (Cowling (1996)). Eubank and Speckman (1990) suggests semi-parametric models in which a parametric function is first employed to fit the data, and a nonparametric method is then adopted to the residuals. However, such a global parametric model may not help to reduce the boundary bias or edge effect due to the allocation of weight outside the support when smoothing is applied near the boundary. Same argument is applied to the standard Rosenblatt-Parzen estimator. A fairly complete list of methods available for removing boundary bias can be found in Chen (2000).

Recently the asymmetric kernel estimators are of wide interests. The most popular choices include the beta kernel estimators in Brown (1999), the Gamma kernel estimators GAM1 & GAM2 in Chen (2000) and the inverse and the reciprocal inverse gaussian kernel estimators IG & RIG in Scaillet (2001). They are free of boundary bias and achieve the optimal rate of convergence in the mean integrated square error. Beta kernel has its unique bounded support on [0,1] and the others have bounded support on $[0,\infty)$.

This article develops more flexible Birnbaum-Saunders (BS) and Lognormal (LN) kernel estimators. Thus it extends the class of non-negative kernel density estimators. LN distribution with scaling parameter α and shaping parameter β is well known. It is a transformation of the normal distribution. BS distribution is introduced by Birnbaum and Saunders (1969) to represent lifetimes with shaping parameter α and scaling parameter β. The BS pdf is a mixture (with equal weights) of the Inverse Gaussian IG$(\beta, \beta\alpha^{-2})$ pdf and the Reciprocal Inverse Gaussian RIG$(\beta, \beta^{-1}\alpha^{2})$ pdf. BS and LN kernel estimators are free of boundary bias, non-negative, with natural varying shape, and achieve the optimal rate of convergence for the mean integrated squared error. Furthermore, BS and LN kernels capture DSLT quite well. Moreover, in estimating pdf of high frequency data, BS and LN kernels take advantages over other existing kernels since they reduce the bias near the boundary. Thus BS and LN kernel estimators should deserve more wide attention in estimating pdf of high-frequency or ultra high-frequency data.

1.1. Outline of paper

The paper is organized as follows. In Section 2 we define BS and LN kernel estimators and their properties such as bias, variance, mean squared error and integrated mean square error. Two propositions are given for bias and variance. In Section 3 we apply BS and LN kernel estimators in estimating pdf of high frequency TD data from New York Stock Exchange's (NYSE) Trade and Quote (TAQ) database. In Section 4 we apply various kernel estimators in estimating the pdf of a simulated Burr(1,3,1) data and compare the corresponding point-wise properties, while Section 5 concludes and outlooks. We provide the proofs of two propositions presented in Section 2 in the Appendix.

2. BS AND LN KERNEL ESTIMATORS

Let $X_1, \ldots, X_n$ be univariate random duration data from a distribution with unknown pdf f which has bounded support on $[0,\infty)$. One representation of pdf of BS(α, β) distribution is

$$f_{BS(\alpha,\beta)}(y) = \frac{1}{2\alpha}\left(\sqrt{\frac{1}{\beta y}} + \sqrt{\frac{\beta}{y^3}}\right)\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2\alpha^2}\left(\frac{y}{\beta} - 2 + \frac{\beta}{y}\right)\right]$$

where $y > 0$, $\alpha > 0$, and $\beta > 0$. Its mean and variance are $\beta\left(1 + \frac{1}{2}\alpha^2\right)$ and $(\alpha\beta)^2(1 + \frac{5}{4}\alpha^2)$, respectively. The pdf of LN(α, β) distribution is

$$f_{LN(\alpha,\beta)}(y) = \frac{1}{\sqrt{2\pi\beta}y}\exp\left[-\frac{1}{2\beta}(\ln y - \alpha)^2\right] \qquad y > 0, \beta > 0$$

Its mean and variance are $\exp(\alpha + \beta/2)$ and $\exp(2\alpha + \beta)\,(\exp(\beta) - 1)$, respectively.

As $\alpha = b^{1/2}$ and $\beta = x$, the class of BS kernels considered is

$$K_{BS(b^{1/2},x)}(s) = \frac{1}{2\sqrt{2b\pi}}\left(\sqrt{\frac{1}{xs}} + \sqrt{\frac{x}{s^3}}\right)\exp\left[-\frac{1}{2b}\left(\frac{s}{x} - 2 + \frac{x}{s}\right)\right] \quad (1)$$

As $\alpha = \ln x$ and $\beta = 4\ln(1+b)$, the class of LN kernels considered is

$$K_{LN(\ln x, 4\ln(1+b))}(s) = \frac{1}{\sqrt{8\pi\ln(1+b)}s}\exp\left[-\frac{(\ln s - \ln x)^2}{8\ln(1+b)}\right] \quad (2)$$

where b is the bandwidth satisfying the condition that $b \to 0$ and $nb \to \infty$ as $n \to \infty$.

The corresponding BS and LN kernel estimators are

$$\hat{f}_{BS}(x) = n^{-1}\sum_{i=1}^{n} K_{BS(b^{1/2},x)}(X_i). \quad (3)$$

and

$$\hat{f}_{LN}(x) = n^{-1}\sum_{i=1}^{n} K_{LN(\ln x, 4\ln(1+b))}(X_i). \quad (4)$$

For the comparison purpose, we list out the other kernels as follows. GAM1 and GAM2 kernels (Chen (2000)) are

$$K_{Gam1(x/b+1,b)}(s) = \frac{s^{x/b}\exp\{-s/b\}}{\Gamma\{x/b+1\}b^{x/b+1}}, \quad (5)$$

and

$$K_{Gam2(\rho_b,b)}(s) = \frac{s^{\rho_b(x)-1}\exp\{-s/b\}}{\Gamma\{\rho_b(x)\}b^{\rho_b(x)}}. \quad (6)$$

where

$$\rho_b(x) = \begin{cases} x/b & \text{if } x \in [2b, \infty); \\ \frac{1}{4}(x/b)^2 + 1 & \text{if } x \in [0, 2b). \end{cases}$$

IG and RIG kernels (Scaillet (2001)) are

$$K_{IG(x,1/b)}(s) = \frac{1}{\sqrt{2\pi b s^3}}\exp\left[-\frac{1}{2bx}\left(\frac{s}{x} - 2 + \frac{x}{s}\right)\right], \quad (7)$$

FIG. 1. MSFT 7/14/1999

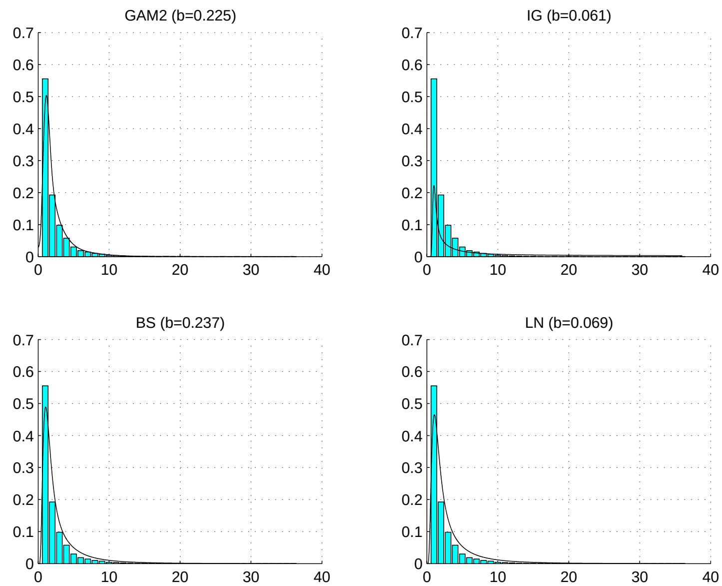


FIG. 2. YHOO 7/16/1999

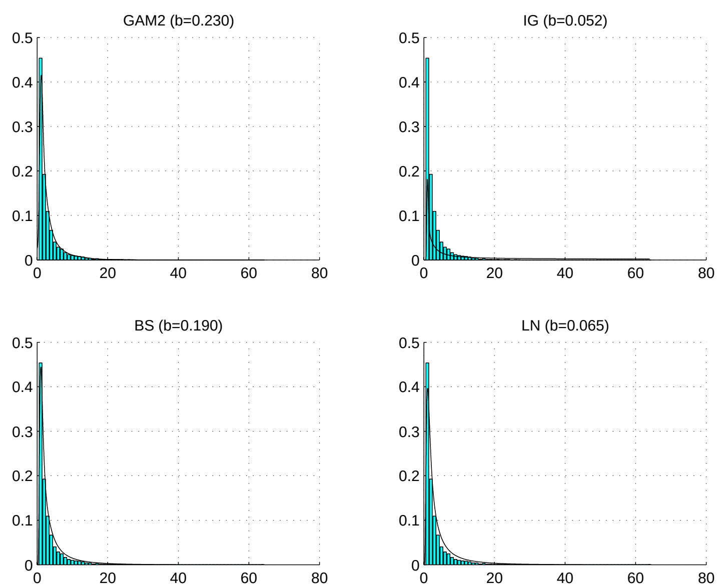


FIG. 3. ORCL 7/9/1999

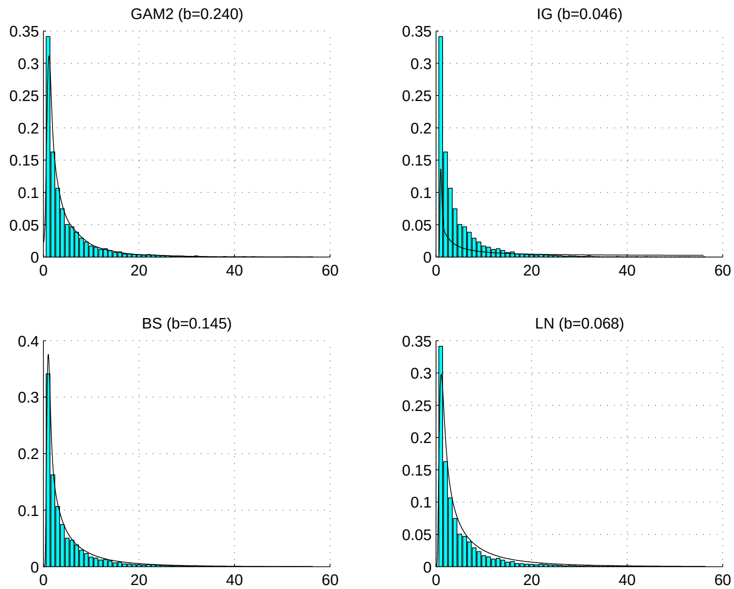


FIG. 4. Kernel Functions for $b = 0.05$

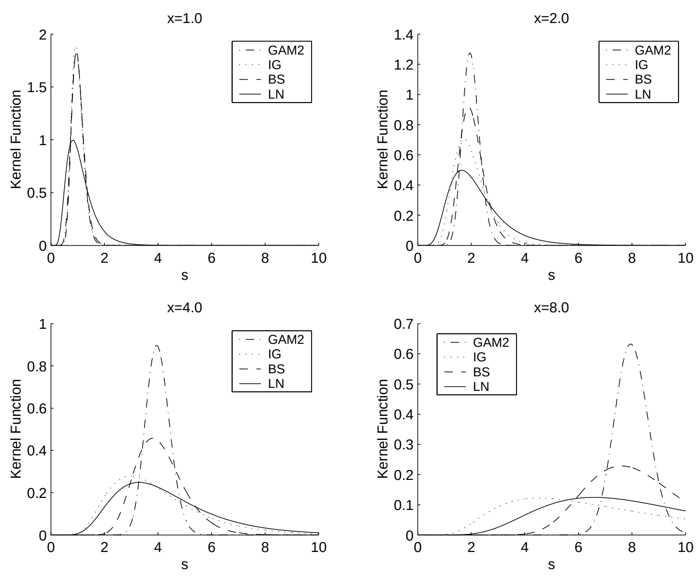


FIG. 5. Kernel Functions for $b = 0.2$

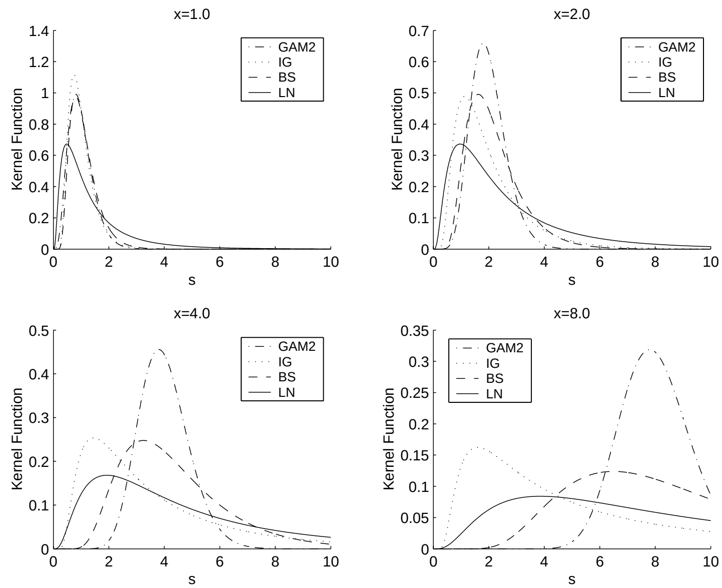


FIG. 6. Kernel Functions for $b = 0.5$

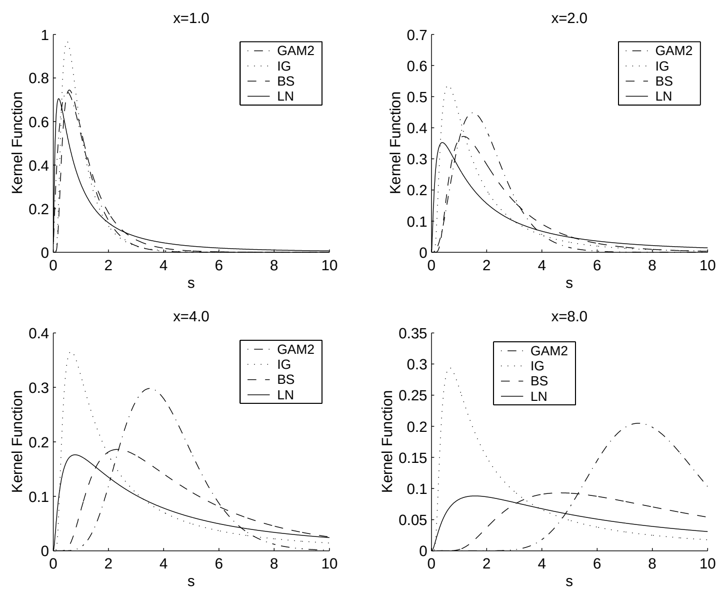
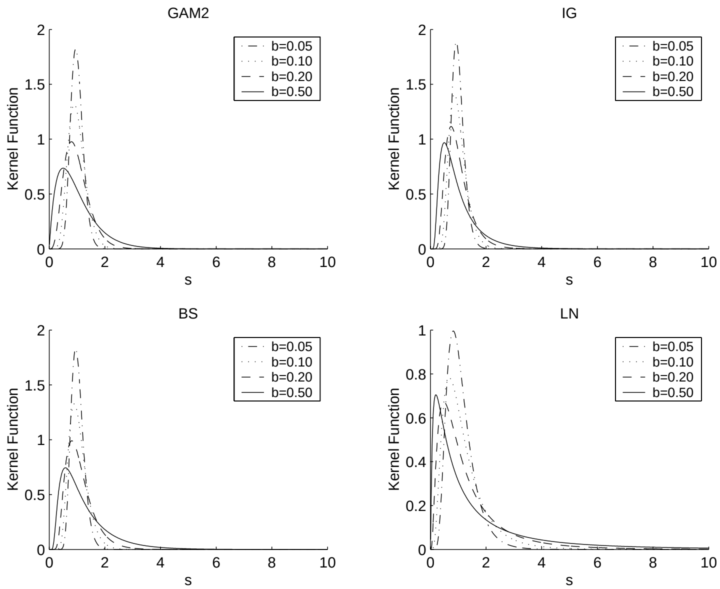


FIG. 7. Kernel Shapes for $x = 1$



and

$$K_{RIG(\ln x, 4\ln(1+b))}(s) = \frac{1}{\sqrt{2\pi bs}} \exp\left[-\frac{x-b}{2b}\left(\frac{s}{x-b} - 2 + \frac{x-b}{s}\right)\right]. \quad (8)$$

As claimed in Chen (2000), GAM2 has a better global performance due to the smaller mean integrated squared error (MISE). And there is no significant difference between IG and RIG kernels, so hereafter we choose BS, LN, GAM2 and IG for comparison.

From Figure 1 to Figure 3 we can see that high frequency TD data usually result in small optimal bandwidths via methods such as biased cross validation. By choosing b to be 0.05, 0.1, 0.2 and 0.5, we plot the four kernels' comparative shapes from Figure 4 to Figure 6, respectively. As x goes through the value of 1, 2, 4 and 8 so it gives four plots on each page. It's clear that GAM2 kernel looks very stable and IG kernel is not working at all (at least for the bandwidths stated). At the same time, BS & LN tend to perform quite well but are sensitive to the deviations in the bandwidths. However, if the bandwidth is close to the optimal, they cover the x-values very well. For small values (close to 0), they perform better in the coverage (smoothness) sense due to the multiplier of x before f' and x^2 before $f''(x)$. This statement can be confirmed by Figure 12 in section 4. Fix x to be 1, 2, 4 and 8, in Figure 7 to Figure 10 we also provide the four

FIG. 8. Kernel Shapes for $x = 2$

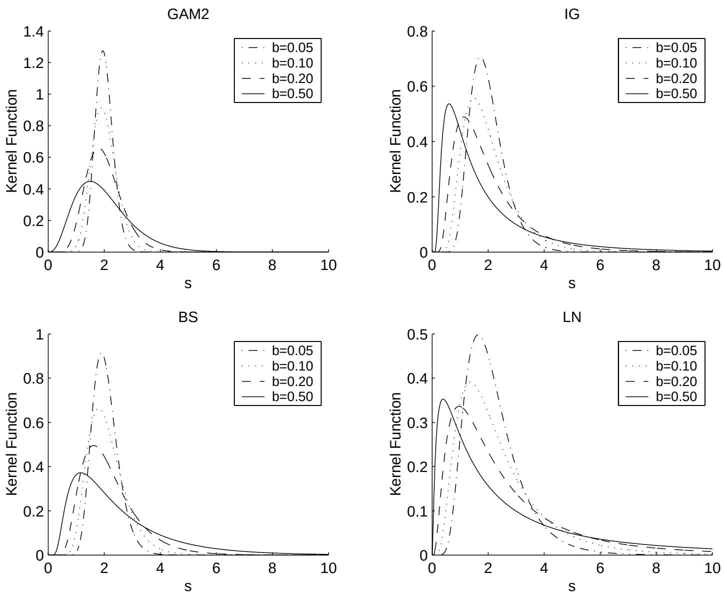


FIG. 9. Kernel Shapes for $x = 4$

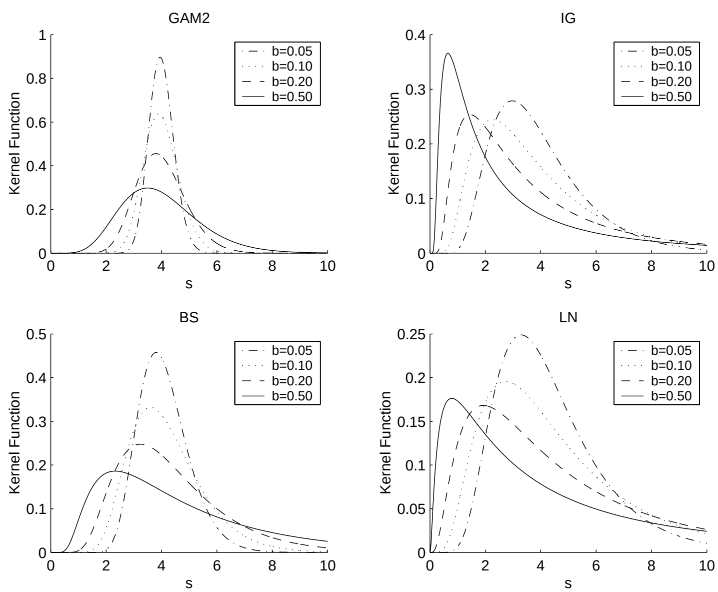


FIG. 10. Kernel Shapes for $x = 8$

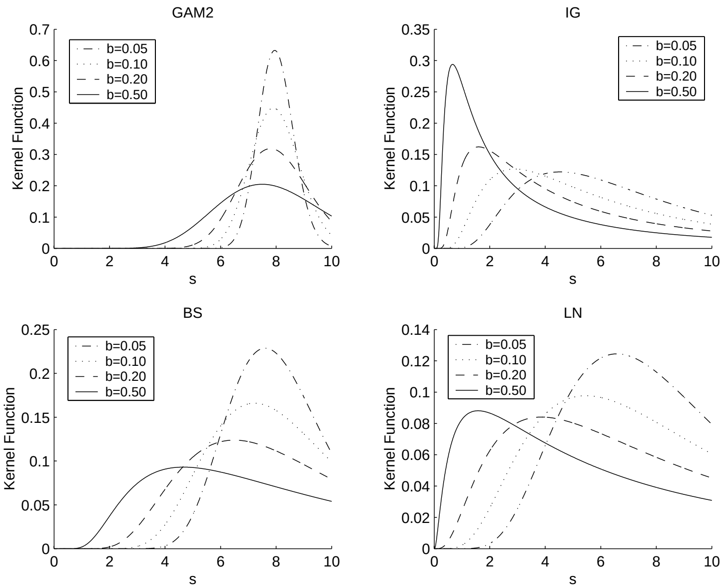


FIG. 11. Burr(1,3,1), $n = 10002$

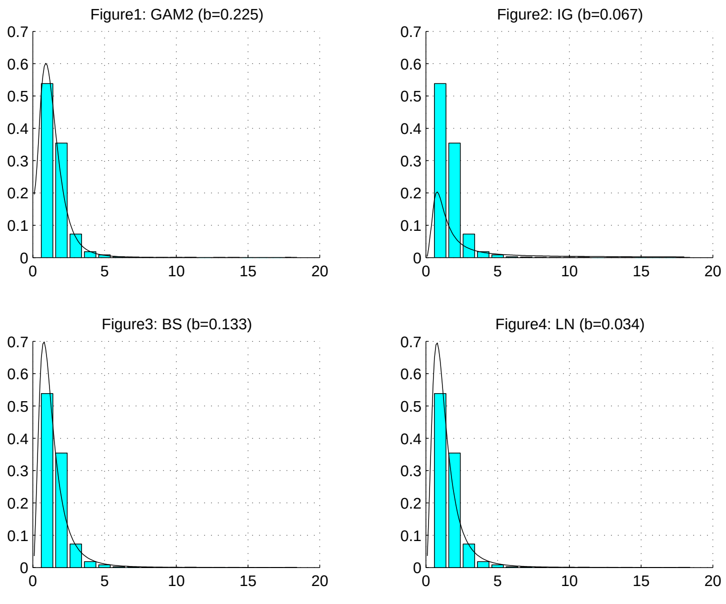
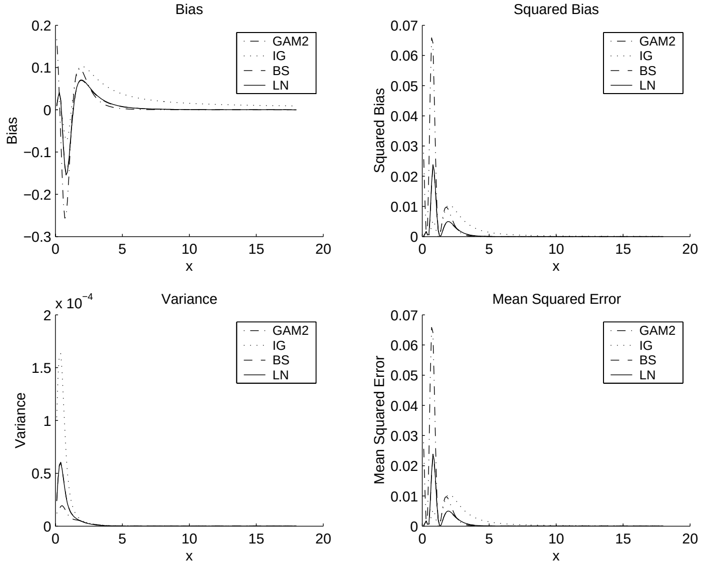


FIG. 12. Point-wise bias, squared bias, variance and mse for estimating pdf of Burr(1,3,1)



comparative shape plots as bandwidths go through 0.05, 0.20 and 0.50 for each kernel. Again, the small bandwidth 0.05 exposes the x-values quite well for all kernels.

TABLE 1.

Simple Statistics for Original and Box-Cox $^{(*)}$ Transformed Data.

Statistics	MSFT1	-1.2(∗)	YHOO2	-0.6(∗)	ORCL3	-0.4(∗)
N	10720	10720	7745	7745	5048	5048
Max	36	0.822	64	1.529	56	2.000
Min	1	0.000	1	0.000	1	0.000
Median	1	0.000	2	0.567	2	0.605
Mode	1	0.000	1	0.000	1	0.000
Mean	2.190	0.261	3.030	0.469	4.645	0.729
Std Dev	2.197	0.302	3.596	0.469	5.681	0.609
Skewness	3.766	0.413	3.865	0.312	2.989	0.075
Kurtosis	26.272	1.395	26.139	1.617	15.523	1.641
L1 moment	2.190	0.261	3.030	0.469	4.645	0.729
L2 moment	0.886	0.157	1.464	0.256	2.478	0.342
L3 moment	0.497	0.037	0.795	0.043	1.227	0.021
L4 moment	3.707	0.531	5.189	0.967	8.175	1.524

The integrated squared biases (IB) for BS and LN kernel estimators are

$$IB_{BS}(b) = \frac{1}{4}b^2 \int_0^\infty \left(xf'(x) + x^2 f''(x)\right)^2 dx + o(b^2). \tag{9}$$

and

$$IB_{LN}(b) = 4b^2 \int_0^\infty \left(xf'(x) + x^2 f''(x)\right)^2 dx + o(b^2). \tag{10}$$

Since we are selecting the bandwidth using MISE criterion the following assumptions about the unknown probability density function are made:

(A1) f is twice continuously differentiable.
(A2) $\int_0^\infty [xf'(x)]^2 \, dx < \infty$ and $\int_0^\infty \left[x^2 f''(x)\right]^2 dx < \infty$.
(A3)

(a) For BS,

$$\frac{3}{16}f(x) - \frac{1}{4}xf'(x) + \frac{1}{4}x^2 f''(x) = O(x^{1/2}); \tag{11}$$

(b) For LN,

$$2f(x) - 2xf'(x) + x^2 f''(x) = O(x). \tag{12}$$

Note: Assumptions (A1) and (A2) are necessary for the Taylor expansion and the finiteness of the second moment respectively. The equations in (A3) resemble first three terms of the Taylor expansion of f around x except for the negative sign for the first derivative. It is important to notice that the behavior of the function f in the vicinity of the origin should be "smooth" and the linear functional based on the function itself and first and second derivatives should grow at the prescribed rate. Of course, higher order expansion would require a bit modified conditions.

These assumptions will be used in the calculations related to the bias and the variance for both kernels.

Assumption (A2) ensures that IBs for BS and LN are finite. Next, we state two propositions with the proofs shifted to the Appendix.

PROPOSITION 2.1 (Bias). *The biases of BS and LN kernel estimators are*

$$Bias\left[\hat{f}_{BS}(x)\right] = b\left[\frac{1}{2}xf'(x) + \frac{1}{2}x^2 f''(x)\right] + o(b) \tag{13}$$

and

$$Bias\left[\hat{f}_{LN}(x)\right] = b\left[2xf'(x) + 2x^2 f''(x)\right] + o(b) \tag{14}$$

PROPOSITION 2.2 (Variance). *The variances of BS and LN kernel estimators are*

$$Var\left[\hat{f}_{BS}(x)\right] = \frac{1}{\sqrt{2\pi}} n^{-1} b^{-1/2} x^{-1} f(x) + o(n^{-1} b^{-1/2}). \tag{15}$$

and

$$Var\left[\hat{f}_{LN}(x)\right] = \frac{1}{4\sqrt{\pi}} n^{-1} b^{-1/2} x^{-1} f(x) + o(n^{-1} b^{-1/2}). \tag{16}$$

One interesting feature is that all variances are of order $o(n^{-1}b^{-1/2})$, except for GAM1 in the case of $x/b \to c$ for a non-negative constant c, whose variance is of order $o(n^{-1}b^{-1})$. The integrated variances (IV) of BS and LN kernel estimators are

$$IV_{BS}(b) = \frac{1}{\sqrt{2\pi}} n^{-1} b^{-1/2} \int_0^\infty \left(x^{-1} f(x)\right) dx + o(n^{-1} b^{-1/2}). \tag{17}$$

and

$$IV_{LN}(b) = \frac{1}{4\sqrt{\pi}} n^{-1} b^{-1/2} \int_0^\infty \left(x^{-1} f(x)\right) dx + o(n^{-1} b^{-1/2}). \tag{18}$$

From (13) and (15), as well as (14) and (16), the mean squared errors (MSE) of BS and LN kernel estimators are

$$\begin{aligned} MSE[\hat{f}_{BS}(x)] &= Bias^2\left[\hat{f}_{BS}(x)\right] + var\left[\hat{f}(x)\right] \\ &= \frac{1}{4} b^2 \left(x f'(x) + x^2 f''(x)\right)^2 + \frac{1}{\sqrt{2\pi}} n^{-1} b^{-1/2} x^{-1} f(x) \\ &\quad + o(n^{-1} b^{-1/2} + b^2) \end{aligned} \tag{19}$$

and

$$\begin{aligned} MSE[\hat{f}_{LN}(x)] &= Bias^2\left[\hat{f}_{BS}(x)\right] + var\left[\hat{f}(x)\right] \\ &= 4b^2 \left(x f'(x) + x^2 f''(x)\right)^2 + \frac{1}{4\sqrt{\pi}} n^{-1} b^{-1/2} x^{-1} f(x) \\ &\quad + o(n^{-1} b^{-1/2} + b^2) \end{aligned} \tag{20}$$

And from (9) and (17), as well as (10) and (18), the integrated mean squared errors (MISE) of BS and LN kernel estimators are

$$\begin{aligned} MISE[\hat{f}_{BS}(x)] &= IB_{BS}(b) + IV_{BS}(b) \\ &= \frac{1}{4}b^2 \int_0^\infty \big(xf'(x) + x^2 f''(x)\big)^2\, dx \\ &\quad + \frac{1}{\sqrt{2\pi}} n^{-1} b^{-1/2} \int_0^\infty \big(x^{-1} f(x)\big)\, dx \\ &\quad + o(n^{-1}b^{-1/2} + b^2) \end{aligned} \tag{21}$$

and

$$\begin{aligned} MISE[\hat{f}_{LN}(x)] &= IB_{LN}(b) + IV_{LN}(b) \\ &= 4b^2 \int_0^\infty \big(xf'(x) + x^2 f''(x)\big)^2\, dx \\ &\quad + \frac{1}{4\sqrt{\pi}} n^{-1} b^{-1/2} \int_0^\infty \big(x^{-1} f(x)\big)\, dx + o(n^{-1}b^{-1/2} + b^2) \end{aligned} \tag{22}$$

By minimizing the leading terms in (21) and (22), we obtain the optimal bandwidths for BS and LN kernel estimators

$$b^*_{BS} = \frac{\left[\frac{1}{\sqrt{2\pi}} \int_0^\infty \big(x^{-1} f(x)\big)\, dx\right]^{2/5}}{\left[\int_0^\infty \big(xf'(x) + x^2 f''(x)\big)^2\, dx\right]^{2/5}} n^{-2/5} \tag{23}$$

and

$$b^*_{LN} = \frac{\left[\frac{1}{4\sqrt{\pi}} \int_0^\infty \big(x^{-1} f(x)\big)\, dx\right]^{2/5}}{2^{8/5}\left[\int_0^\infty \big(xf'(x) + x^2 f''(x)\big)^2\, dx\right]^{2/5}} n^{-2/5} \tag{24}$$

The optimal bandwidths are of order $O(n^{-2/5})$ for all kernels. Plugging in the above optimal bandwidths to (21) and (22) we obtain the optimal MISE

$$\begin{aligned} MISE^*_{BS} &= \frac{5}{4}\left[\int_0^\infty \big(xf'(x) + x^2 f''(x)\big)^2\, dx\right]^{1/5} \\ &\quad \times \left[\frac{1}{\sqrt{2\pi}} \int_0^\infty \big(x^{-1} f(x)\big)\, dx\right]^{4/5} n^{-4/5} \end{aligned} \tag{25}$$

and

$$MISE^*_{LN} = \frac{5}{2^{3/5}} \left[\int_0^\infty \left(xf'(x) + x^2 f''(x) \right)^2 dx \right]^{1/5} \times \left[\frac{1}{4\sqrt{\pi}} \int_0^\infty \left(x^{-1} f(x) \right) dx \right]^{4/5} n^{-4/5} \tag{26}$$

3. EMPIRICAL APPLICATION

We apply BS and LN kernel estimators on high frequency time duration (TD) data. Three transaction level TD data, Microsoft 7/14/1999 (MSFT1), Yahoo 7/16/1999 (YHOO2), and Oracle 7/9/1999 (ORCL3) are selected from the New York Stock Exchange's (NYSE) Trade and Quote (TAQ) database. Some special conditions are imposed based on TAQ documentation and the data are trimmed to 9:30AM-16:00PM trading hours. TDs are measured in seconds between successive trades. If the recording system is more advanced, TDs would be measured more accurately in smaller grids such that they would behave much more in a continuous fashion. As mentioned in Giot (2000), price durations feature a strong time-of-the-day effect. Lots of researchers consider diurnally adjusted TDs. Since in determining the diurnal factor, it's subjective to choose the time length for fitting a spline with nodes at each time period. Therefore without loss of generality, we still use the raw TD data without transformation in our paper. Similar work follows for diurnally adjusted TD data.

Out of 19826, 10995 and 7298 total duration observations from MSFT1, YHOO2, and ORCL3, there are 9106, 2250 and 2250 zero durations (namely D_0), respectively. High percentage of D_0 reflects the high intensity of trading. Denote $\tilde{\pi}$ the proportion of D_0 counts out of total duration counts. First we exclude D_0 and obtain kernel estimator $\hat{f}(x)$ for the rest of the data D_1, then by defining weighted kernel estimator $\tilde{f}(\cdot)$ as

$$\tilde{f}(x) = \begin{cases} \tilde{\pi} & \text{if } x = 0; \\ (1 - \tilde{\pi})\hat{f}(x) & \text{if } x \in (0, \infty). \end{cases}$$

we get the whole picture.

High frequency characteristic is obvious from the simple statistics for D_1 displayed in Table 1. Table 1 also provides L-moments describe in Hosking (1990). L-moments are a linear combination of order statistics. Significant advantages of L-moments over conventional central moments in our case are: robustness to outliers and characterization of an unknown distribution. Sample L1 moments are equivalent to sample mean. The

2nd, 3rd and 4th sample L-moments are much more smaller than standard deviation, skewness and kurtosis in the original data, respectively. In Table 1, simple statistics are also presented for D_1 after Box-Cox transformation. Apparently, Table 1 reveals the DSLT characteristic in high frequency data.

Bandwidth selection is an immediate issue directly related to applying asymmetric kernels. Turlach (1993) and Chiu (1996) provide quite nice reviews on bandwidth selection. In choosing optimal bandwidth for asymmetric kernels, the practical methods available are least squares cross-validation (LSCV), biased cross-validation (BCV), smoothed cross-validation (SCV), etc. The plug-in method, bootstrapping method and adaptive varying kernel size selection (Katkovnik and Shmulevich (2000)) are not applicable in our case since they are pilot bandwidth and symmetric kernel-driven.

In this paper we apply BCV in our high frequency data to obtain the optimal bandwidths. The duration histograms and kernel estimator functions with those optimal bandwidths are overlayed from Figure 1 to Figure 3. Several important observations are worth mentioning. First, due to the nature of high frequency, the optimal bandwidths b^* are all of small values less than 0.5. Second, although there exist differences among three data, b^* is extremely close for LN and GAM2 kernels. Last but not least, we can see that all kernels except IG are doing good jobs in estimating pdfs. IG obviously underestimate the frequency for small x-values. Therefore IG is not suitable for estimating the pdf of high frequency duration data.

4. SIMULATION RESULTS

In order to compare the performance of BS, LN, GAM2 and IG kernels, we selectively generate a random sample from a "neutral" Burr(1,3,1) distribution. The pdf of Burr($\mu = 1$,k,r) is

$$f_{Burr}(s) = \frac{k{\cdot}s^{k-1}}{\left(1 + r{\cdot}s^k\right)^{1/r+1}},$$

The mean and variance for Burr(1,k,r) are

$$E(S) = \frac{\Gamma\left(1+\frac{1}{k}\right)\cdot\Gamma\left(\frac{1}{r}-\frac{1}{k}\right)}{r^{1+\frac{1}{k}}\cdot\Gamma\left(\frac{1}{r}+1\right)},$$

and

$$V(S) = \frac{\Gamma\left(1+\frac{2}{k}\right)\cdot\Gamma\left(\frac{1}{r}-\frac{2}{k}\right)}{r^{1+\frac{2}{k}}\cdot\Gamma\left(\frac{1}{r}+1\right)} - [E(S)]^2,$$

respectively. And the mean and variance for Burr(1,3,1) are 1.2092 and 0.9562, respectively. The nice feature about Burr is that, compared with

Weibull, Gamma and some other parametric distributions, it captures the DSLT quite well hence it is frequently cited by Fernandes and Grammig (2000). Besides, it has closed form of CDF. First, we generate random U(0,1) numbers with size equal to 10002, then we transform them into Burr(1,3,1) by

$$y = \left(\frac{x}{1-x}\right)^{1/3}$$

The range of Burr(1,3,1) numbers is [0,17.55). We choose the optimal bandwidth for each kernel by minimizing the integrated squared error

$$ISE(b) = \int_0^{17.55} \left[\hat{f}(x) - f(x)\right]^2 dx$$

The histogram and kernel estimator functions for Burr(1,3,1) are overlayed in Figure 11. Again, all kernels except IG do good jobs in estimating the Burr(1,3,1) pdf. All the optimal bandwidths are of small values. In Figure 12 we display the point-wise bias, squared bias, variance and mean square errors of four kernel estimators for x∈[0,17.55) for Burr(1,3,1) density with size n=10002. From Figure 12 we can see that the bias, squared bias, variance and MSEs are all go to zero as x greater than 5. It's hard to distinguish between BS and LN kernels from the graph since they are overlapped. The squared bias for BS and LN are much more smaller than that of GAM2 for the smaller values of x. Since variances are much smaller than their corresponding squared bias, so MSEs of BS and LN is also much smaller than MSE of GAM2. This confirms our statements in section 2. Thus, by choosing BS and LN kernel estimators, we can quite well estimate the pdf of high or ultra high frequency data.

5. CONCLUSION AND OUTLOOK

In this paper we introduced two new kernels BS and LN to estimate the pdf with bounded support $[0, \infty)$. They are free of boundary bias, non-negative, with natural varying shape, and achieve the optimal rate of convergence $O(n^{-2/5})$ for the mean integrated squared error. When we apply these two kernels in high-frequency financial intraday time duration data, we can see that they are better than IG kernel in terms of estimation and better than GAM2 due to smaller squared bias and MSE. Since BS and LN kernels achieve accurate non-parametric pdf estimates, we can apply them in non-parametric specification tests as mentioned in Fernandes and Grammig (2000). Although this paper concentrates on estimating pdf of duration data for the illustration purpose, BS and LN kernels are also suitable for estimating pdf of all other high frequency or ultra frequency data. To further extend the class of non-parametric kernels, more work can be done to explore the Pareto kernel, Burr kernel, etc.

APPENDIX

Proof (**Proof of Proposition 1.**).
Bias of the BS kernel estimator.
Since

$$E\left[\hat{f}_{BS}(x)\right] = \int_0^\infty K_{BS(b^{1/2},x)}(u)f(u)\,du = E\left[f(\delta_x)\right] \tag{A.1}$$

where δ_x is the BS$(b^{1/2}, x)$ random variable with

$$\mu_\delta = E[\delta_x] = x\left(1 + \frac{1}{2}b\right) = x + O(b) \tag{A.2}$$

$$V_\delta = Var[\delta_x] = x^2 b\left(1 + \frac{5}{4}b\right) = x^2 b + O(b^2) \tag{A.3}$$

Employing Taylor expansion while plugging in (A.2) and (A.3) gives

$$\begin{aligned}
E[f(\delta_x)] &= f(\mu_\delta) + \frac{1}{2}f''(x)V_\delta + o(b) \\
&= f\left(x + \frac{x}{2}b\right) + \frac{1}{2}f''(x)x^2 b\left(1 + \frac{5}{4}b\right) + o(b) \\
&= f(x) + f'(x)\frac{1}{2}xb + \frac{1}{2}f''(x)x^2 b\left(1 + \frac{5}{4}b\right) + o(b) \\
&= f(x) + b\left[\frac{1}{2}xf'(x) + \frac{1}{2}f''(x)\right] + o(b) \qquad \text{(A.4)}
\end{aligned}$$

(A.1) and (A.4) lead to (13).
Bias of the LN kernel estimator.
Since

$$E\left[\hat{f}_{LN}(x)\right] = \int_0^\infty K_{LN(\ln x, 4\ln(1+b))}(u)f(u)\,du = E\left[f(\delta_x)\right] \tag{A.5}$$

where δ_x is the LN$(\ln x, 4\ln(1+b))$ random variable and using Taylor expansion while plugging in (A.8) and (A.9) gives

$$\begin{aligned} E[f(\delta_x)] &= f(\mu_\delta) + \frac{1}{2}f''(x)V_\delta + o(b) \\ &= f\left(x(1+b)^2\right) + \frac{1}{2}f''(x)\left[x^2(1+b)^4\left((1+b)^4-1\right)\right] + o(b) \\ &= f(x) + f'(x)b(2x+bx^2) + \frac{1}{2}f''(x)b(4x^2) + o(b) \\ &= f(x) + b\left[2xf'(x) + 2x^2f''(x)\right] + o(b) \end{aligned} \tag{A.6}$$

(A.7)

where

$$\mu_\delta = E[\delta_x] = x(1+b)^2) = x + 2bx + b^2x^2 = x + O(b) \tag{A.8}$$

$$\begin{aligned} V_\delta = Var[\delta_x] = x^2(1+b)^4[(1+b)^4-1] &= 4x^2b + 22x^2b^2 + O(b^3) \\ &= 4x^2b + O(b^2). \end{aligned} \tag{A.9}$$

Finally, (A.5) and (A.6) lead to (14). ▮

Proof (**Proof of Proposition 2.**).
Variance of the BS kernel estimator.
The variance of $\hat{f}_{BS}(x)$ is

$$Var\left[\hat{f}_{BS}(x)\right] = n^{-1}E\left[K^2_{BS(b^{1/2},x)}(X_i)\right] + O(n^{-1}) \tag{A.10}$$

Let λ_x be a BS$\left((b/2)^{1/2}, x\right)$ random variable such that

$$\mu_\lambda = E[\lambda_x] = x\left(1 + \frac{1}{4}b\right) = x + O(b), \tag{A.11}$$

$$V_\lambda = Var[\lambda_x] = \frac{x^2}{2}b\left(1 + \frac{5}{8}b\right) = \frac{x^2}{2}b + O(b^2). \tag{A.12}$$

Then

$$E\left[K^2_{BS\left(b^{1/2},x\right)}(X_i)\right] = C_b^1E\left[\lambda_x^{-1/2}f(\lambda_x)\right] + C_b^2E\left[\lambda_x^{-3/2}f(\lambda_x)\right] \tag{A.13}$$

where $C_b^1 = (8\pi xb)^{-1/2}$ and $C_b^2 = \left(8\pi x^{-1}b\right)^{-1/2}$. Therefore applying Taylor expansion again while plugging in (A.11) and (A.12) gives

$$\begin{aligned}
E\left[\lambda_x^{-1/2} f(\lambda_x)\right] &= \mu_\lambda^{-1/2} f(\mu_\lambda) \\
&\quad + \frac{1}{2}\left(\frac{3}{4}x^{-5/2} f(x) - x^{-3/2} f'(x) + x^{-1/2} f''(x)\right) V_\lambda + o(b) \\
&= x^{-1/2}(1+\frac{b}{4})^{-1/2} f(x+\frac{x}{4}b) \\
&\quad + b\left(\frac{3}{16}x^{-1/2} f(x) - \frac{1}{4}x^{1/2} f'(x) + \frac{1}{4}x^{3/2} f''(x)\right) + o(b) \\
&= x^{-1/2}\left(1 - \frac{1}{8}b + \frac{3}{128}b^2 + o(b^2)\right)\left(f(x) + \frac{x}{4}bf'(x)\right) \\
&\quad + b\left(\frac{3}{16}x^{-1/2} f(x) - \frac{1}{4}x^{1/2} f'(x) + \frac{1}{4}x^{3/2} f''(x)\right) + o(b) \\
&= x^{-1/2} f(x) + O(b). \qquad \text{(A.14)}
\end{aligned}$$

We apply (11) to claim (A.14).

Similarly

$$E\left[\lambda_x^{-3/2} f(\lambda_x)\right] = x^{-3/2} f(x) + O(b). \qquad \text{(A.15)}$$

(A.10), together with (A.13), (A.14) and (A.15), gives (15).

Variance of the LN kernel estimator.

The variance of $\hat{f}_{LN}(x)$ is

$$Var\left[\hat{f}_{LN}(x)\right] = n^{-1} E\left[K^2_{LN(\ln x, 4\ln(1+b))}(X_i)\right] + O(n^{-1}) \qquad \text{(A.16)}$$

Let λ_x be a $LN\,(\ln x, 2\ln(1+b))$ random variable such that

$$\mu_\lambda = E[\lambda_x] = x(1+b) = x + O(b), \qquad \text{(A.17)}$$

$$\begin{aligned}
V_\lambda = Var[\lambda_x] = x^2(1+b)^2\left[(1+b)^2 - 1\right] &= 2x^2b + 5x^2b^2 + O(b^3) \qquad \text{(A.18)} \\
&= 2x^2b + O(b^2).
\end{aligned}$$

Then

$$E\Big[K^2_{LN(\ln x, 4\ln(1+b))}(X_i)\Big] = C_b E\left[\lambda_x^{-1} f(\lambda_x)\right] \qquad \text{(A.19)}$$

where $C_b = (16\pi \ln(1+b))^{-1/2}$. Therefore applying Taylor Expansion again while plugging in (A.17) and (A.18) gives

$$\begin{aligned}
E\left[\lambda_x^{-1} f(\lambda_x)\right] &= \mu_\lambda^{-1} f(\mu_\lambda) + \frac{1}{2}\big(2x^{-3} f(x) - 2x^{-2} f'(x) \\
&\quad + x^{-1} f''(x)\big) V_\lambda + o(b) \\
&= (x+xb)^{-1} f(x+xb) \\
&\quad + b\left(2x^{-1} f(x) - f'(x) + x\frac{1}{2}x^{-1} f''(x)\right) + o(b) \\
&= x^{-1}(1+b)^{-1} f(x) + \frac{b}{1+b} f'(x) + o(b) && \text{(A.20)} \\
&= x^{-1}(1+b)^{-1} f(x) + \big(b - b^2 - O(b^3)\big) f'(x) + o(b) && \text{(A.21)} \\
&= x^{-1}(1+b)^{-1} f(x) + b f'(x) + o(b) \\
&= x^{-1}(1+b)^{-1} f(x) + O(b). && \text{(A.22)}
\end{aligned}$$

We also apply (12) to claim (A.20) and apply series expansion on b to obtain (A.21) from (A.20). Based on the series expansion

$$\begin{aligned}
(\ln(1+b))^{-1/2}(1+b)^{-1} &= b^{-1/2} - \frac{3}{4} b^{1/2} + \frac{65}{96} b^{3/2} + O(b^{5/2}) \\
&= b^{-1/2} + o(b^{-1/2}), && \text{(A.23)}
\end{aligned}$$

together with (A.16), (A.19), (A.22) we have

$$\begin{aligned}
Var\left[\hat{f}_{LN}(x)\right] &= n^{-1}\left[(16\pi \ln(1+b))^{-1/2} x^{-1}(1+b)^{-1} f(x)\right] + O(n^{-1}b) \\
&= \frac{1}{4\sqrt{\pi}} n^{-1} b^{-1/2} x^{-1} f(x) + o(n^{-1} b^{-1/2}). && \text{(A.24)}
\end{aligned}$$

(A.24) is exactly (16). ▮

REFERENCES

Birnbaum, Z. W. and S. C. Saunders, 1969, A new family of life distributions. *Journal of Applied Probability* **6**, 319-327.

Brown, B. M. and S.-X. Chen, 1999, Beta-Bernstein smoothing for regression curves with compact supports. *Scand. J. Statist.* **26**, 47-59.

Chen, S.-X., 2000, Probability density function estimation using gamma kernels. *Ann. Inst. Statist. Math.* **52, No.3**, 471-480.

Chiu, S.-T., 1996, A comparative review of bandwidth selection for kernel density estimation. *Statistica Sinica* **6**, 129-145.

Cowling, A. and P. Hall, 1996, On pseudodata methods for removing boundary effects in kernel density estimation. *J. Roy. Statist. Soc. Ser. B* **58**, 551-563.

Engle, R. F. and J. R. Russell, 1998, Autoregressive conditional duration: A new model for irregularly-spaced transaction data. *Econometrica* **66**, 1127-1162.

Eubank, R. L. and P. Speckman, 1990, Curve fitting by polynomial-trigonometric regression. *Biometrika* **77**, 1-9.

Fernandes, M. and J. Grammig, 2000, Non-parametric specification tests for conditional duration models. *C.O.R.E Discussion Paper* 2000-4, European University Institute.

Fernandes, M. and J. Grammig, 2001, A family of autoregressive conditional duration models. *ECO Working Paper* 2001-36, European University Institute.

Giot, P., 2000, Time transformations, intraday data, and volatility models. *Journal of Computational Finance* **4**, 31-62.

Hosking, J. R. M., 1990, L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. R. Statist. Soc. B* **52, No.1**, 105-124.

Katkovnik, V. and I. Shmulevich, 2000, Nonparametric density estimation with adaptive varying window size. *Conference on Image and Signal Proessing for Remote Sensing VI, European Symposium on Remote Sensing* Barcelona, Spain, Septermber 25-29, 2000.

Scaillet, O., 2001, Density estimation using inverse and reciprocal inverse gaussian kernels. *IRES Discussion Paper*, No. 2001-17, Univ. Catholique de Louvain.

Schuster, E. F., 1985, Incorporating support constraints into nonparametric estimation of densites. *Commun. Statist. Theory Meth.* **14**, 1123-1136.

Silverman, B. W., 1986, *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Turlach, B. A., 1993, Bandwidth selection in kernel density estimation: A review. *C.O.R.E and Institut de Statistique*, Univ. Catholique de Louvain.