Malec, Peter; Schienle, Melanie

Working Paper

Nonparametric Kernel density estimation near the boundary

SFB 649 Discussion Paper, No. 2012-047

Provided in Cooperation with:
Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

Suggested Citation: Malec, Peter; Schienle, Melanie (2012) : Nonparametric Kernel density estimation near the boundary, SFB 649 Discussion Paper, No. 2012-047, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:
https://hdl.handle.net/10419/79594

SFB 649 Discussion Paper 2012-047

Nonparametric Kernel Density Estimation Near the Boundary

Peter Malec*
Melanie Schienle*

* Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

http://sfb649.wiwi.hu-berlin.de
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin

Nonparametric Kernel Density Estimation Near the Boundary *

Peter Malec† Melanie Schienle‡

This Version: August 13, 2012

Abstract

Standard fixed symmetric kernel type density estimators are known to encounter problems for positive random variables with a large probability mass close to zero. We show that in such settings, alternatives of asymmetric gamma kernel estimators are superior but also differ in asymptotic and finite sample performance conditional on the shape of the density near zero and the exact form of the chosen kernel. We therefore suggest a refined version of the gamma kernel with an additional tuning parameter according to the shape of the density close to the boundary. We also provide a data-driven method for the appropriate choice of the modified gamma kernel estimator. In an extensive simulation study we compare the performance of this refined estimator to standard gamma kernel estimates and standard boundary corrected and adjusted fixed kernels. We find that the finite sample performance of the proposed new estimator is superior in all settings. Two empirical applications based on high-frequency stock trading volumes and realized volatility forecasts demonstrate the usefulness of the proposed methodology in practice.

Keywords: Kernel density estimation; boundary correction; asymmetric kernel
JEL Codes: C14, C51

*This research is supported by the Deutsche Forschungsgemeinschaft via the Collaborative Research Center 649 "Economic Risk".

†Institute for Statistics and Econometrics, Humboldt-Universität zu Berlin. Email: malecpet@hu-berlin.de. Address: Spandauer Str. 1, D-10178 Berlin, Germany.

‡Corresponding author. Institute for Statistics and Econometrics and Center for Applied Statistics and Economics (CASE), Humboldt-Universität zu Berlin. Email: melanie.schienle@wiwi.hu-berlin.de. Address: Spandauer Str. 1, D-10178 Berlin, Germany.

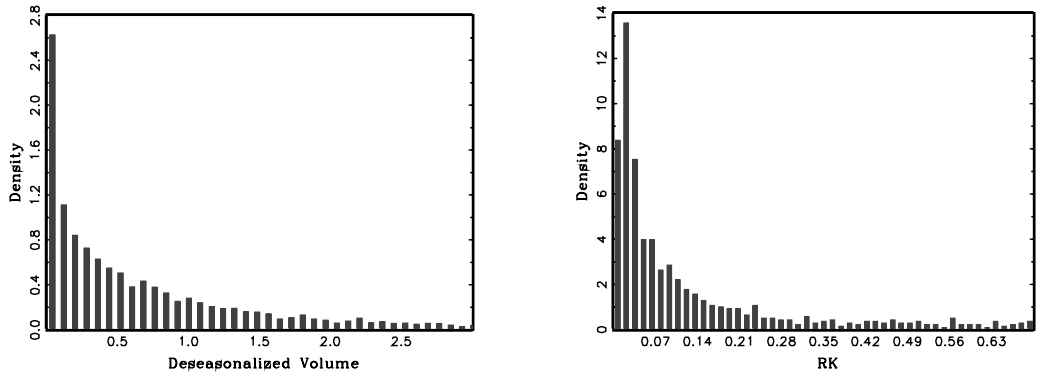
1 Introduction

There are many applications in particular in economics where densities of positive random variables are the object of interest or an essential model ingredient to be estimated from data. Compare e.g. income data, financial transaction data, volatility models but also duration and survival times data. In a lot of these situations, however, appropriate functional forms are unknown or controversial, such that a nonparametric estimate is needed. And it is often the point estimates close to the boundary which are in the focus of practical interest thus require good precision.

For cases of densities where most of the data is concentrated away from the boundary, there is a huge literature on boundary correction techniques of the standard symmetric fixed kernel density estimator. Such adjustments are needed at points close to the boundary since fixed kernels might assign positive weight outside the support yielding inconsistent results. Among these techniques count e.g. the cut- and normalized kernel, see Gasser and Müller (1979), and the reflection method, see Schuster (1958).

If, however, the true density might have substantial mass close the boundary, there are superior methods such as the boundary kernel of Jones (1993). As this estimator could yield negative point estimates, this is corrected in Jones and Foster (1996) at some minor cost of performance (see Jones (1993)). In comparison, the combination of polynomial transformation followed by reflection as in Marron and Ruppert (1994) is much less flexible working well exclusively at boundaries if the initial transformation is close enough to the density shape near zero.

Nonparametric kernel density estimators with asymmetric kernels such as gamma kernels have been introduced to improve upon the performance of fixed kernels at the boundary. In particular for positive random variables their flexible shape avoids the boundary consistency problem and directly yields positive estimates by construction. We illustrate in an extensive simulation study that especially in cases where the true density f approaches the boundary with a derivative f' significantly different from zero, gamma kernel estimates yield superior results to corrected or adjusted fixed kernels. Such density shapes naturally appear in high-frequency data e.g., when studying aggregated trading volumes (see Figure 1) but also in many other applications such as spectral density estimation of long memory time series or when modeling volatilities in particular on the intra-daily level (see e.g. Robinson and Henry (2003) and Corradi et al. (2009)). But we also show that depending on the underlying shape of the true density, the two existing gamma kernel estimators, the so called standard and modified version as introduced in Chen (2000), might also differ substantially in boundary performance and still leave significant room for improvement. In particular, our asymptotic considerations



(a) Volume (b) Realized Kernel

Figure 1: Histograms of Intraday Trading Volume and Realized Kernel Estimates
We consider deseasonalized nonzero 15-Second trading volumes of Citigroup and realized kernel estimates for JP Morgan. Sample period: February 2009 (trading volumes), January 2006 – December 2009 (realized kernel). For details on the seasonal adjustment of trading volumes and computation of the realized kernel, see Section 4.

and a thorough finite sample study suggest that in particular in cases with a pole at zero the standard gamma kernel significantly outperforms the modified version, which is generally used in practice. We therefore introduce a simple data-driven criterion identifying such extreme settings. For all other situations, we propose a refined gamma kernel, which introduces a modification parameter according to the shape of f and its first two derivatives close to the boundary. For determining the appropriate specification of this refined gamma kernel estimator in practice we also provide an automatic procedure.

In an extensive simulation study we compare the performance of this refined estimator to standard gamma kernel estimates and standard boundary corrected and adjusted fixed kernels. We find that the finite sample performance of the proposed new estimator is superior in all settings. Our two applications on high frequency stock trading volumes and realized variance forecasts demonstrate the usefulness of our proposed methodology.

2 Kernel Density Estimation at the Boundary

Throughout the paper, we study density estimation for the case that the support $S_X \subset \mathbb{R}$ of an unknown density is bounded from one side. Without loss of generality, we take this bound to be a lower bound and equal to zero as in many applications as e.g. wage distributions, distribution of trading volumes, etc.. Obtained results, however, can be easily generalized by appropriate translations and reflections at the y-axis. Note also that we restrict our exposition to the case of univariate densities for ease of notation. Multivariate extensions are systematically straightforward via product kernels.

For a random sample $\{X_i\}_{i=1}^n$ from a distribution with unknown density $f_X(x)$, the conventional kernel density estimator has the form

$$\hat{f}_X(x) = \frac{1}{nb} \sum_{i=1}^{n} K\left(\frac{x - X_i}{b}\right), \tag{1}$$

where is b a smoothing bandwidth with $b \to 0$ and $nb \to \infty$ as $n \to \infty$ and K is a kernel function which integrates to unity i.e., $\int K(u)\, du = 1$. If the shape of K is symmetric and fixed across the support, estimation and inference are generally simplified for unbounded support. But if zero bounds the support S_X from below, $\hat{f}_X$ is inconsistent at the boundary $[0, b)$ for such simplistic choices of K. The literature has therefore provided many suggestions for adjustments in fixed kernel estimation, which we will outline in more detail when they appear as benchmarks in the simulation section 3. What characterizes all these approaches, however, is that they mostly work well only for specific forms of f_x in the boundary region and/or can yield negative estimates. In particular, for densities with non-vanishing probability mass close to zero as in Figures 1 these standard correction methods perform poorly at the boundary. Though in applications, it is exactly this boundary region which is in the focus of attention and requires precise estimates.

2.1 Standard Asymmetric Kernel Density Estimators

Density estimators with locally varying form kernels have shown good performance for a wide range of shapes of the underlying true density. Such kernels are nonnegative, but no longer symmetric adjusting in skewness along the support. For the considered one-sided boundary problem, gamma kernel estimators are the simplest and most popular forms of such flexible estimators. In case of a two-sided boundary which is not our focus here, beta kernels would be the appropriate choice (see Chen (1999)). There are two alternative specifications of gamma kernel estimators proposed by Chen (2000) of which the first kind is defined as

$$\hat{f}_X^{\gamma}(x) = \frac{1}{n} \sum_{i=1}^{n} K_{x/b+1,b}^{\gamma}(X_i), \tag{2}$$

where $K_{x/b+1,b}^{\gamma}$ denotes the density of the gamma distribution with shape parameter $x/b + 1$ and scale parameter b, i.e.

$$K_{x/b+1,b}^{\gamma}(u) = \frac{u^{x/b} \exp(-u/b)}{b^{x/b+1}\, \Gamma(x/b + 1)}. \tag{3}$$

Consistency and asymptotic normality of the above estimator are straightforward to derive under standard assumptions (see e.g. Chen (2000) for the pointwise, and Hagmann and Scaillet (2007) for the uniform version). For time series observations, consistency can also be obtained under mixing assumptions in Bouezmarni and Rombouts (2010). In particular, for a sufficiently smooth density $f_X \in C^2(S_X)$, it can be shown that bias and variance vanish asymptotically for $b \to 0$ and $nb \to \infty$. Their asymptotic forms are

$$\text{Bias}\left\{\hat{f}_X^{\gamma}(x)\right\} = b\left\{f_X'(x) + \frac{1}{2}xf_X''(x)\right\} + o(b)\,, \tag{4}$$

and

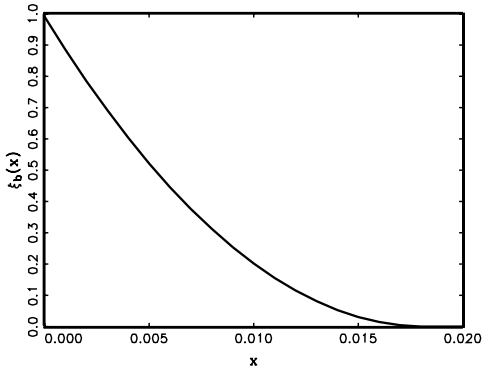
$$\text{Var}\left\{\hat{f}_X^{\gamma}(x)\right\} \approx \begin{cases} \frac{f_X(x)}{nb}\,\mathcal{C}_b(x) & \text{if } x/b \to \kappa; \\ \frac{f_X(x)}{2\sqrt{\pi}}\,(xb)^{-1/2}\,n^{-1} & \text{if } x/b \to \infty, \end{cases} \tag{5}$$

where κ is a nonnegative constant and $\mathcal{C}_b(x) = \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\,\Gamma^2(\kappa+1)}$. Accordingly, the asymptotic mean squared error is

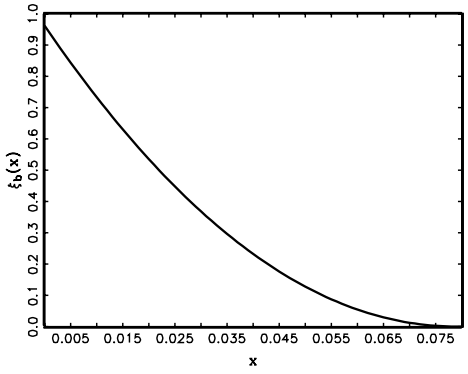
$$\text{MSE}\left\{\hat{f}_X^{\gamma}(x)\right\} \approx \begin{cases} b^2\,\left\{f_X'(x) + \frac{1}{2}xf_X''(x)\right\}^2 + \frac{f_X(x)}{nb}\,\mathcal{C}_b(x) & \text{if } x/b \to \kappa; \\ b^2\,\left\{f_X'(x) + \frac{1}{2}xf_X''(x)\right\}^2 + \frac{f_X(x)}{2\sqrt{\pi}}\,(xb)^{-1/2}\,n^{-1} & \text{if } x/b \to \infty. \end{cases} \tag{6}$$

Note that the asymptotic variance decreases for large x which is offset by an increasing bias. In contrast to fixed kernel estimators, the asymptotic bias contains the first derivative of the density f' which is due to the fact that the chosen flexible kernel shape has its mode rather than its mean at the point of estimation x. The modified gamma kernel estimator improves on this for most of the support without generating convergence problems in the boundary region. In particular, it uses the pdf of a gamma distribution with shape parameter x/b and scale parameter b as kernel function in the interior of the support. This has mean x, but is unbounded for x approaching zero. Therefore the kernel function consists of two regimes where the boundary form is chosen ad hoc to smoothly connect to the desired interior shape while avoiding unboundedness problems. According to Chen (2000) the estimator is thus defined as

$$\hat{f}_X^{\gamma m}(x) = \frac{1}{n}\sum_{i=1}^{n} K^{\gamma}_{\rho_b(x),b}\,(X_i)\,, \tag{7}$$



(a) $b = 0.0091$



(b) $b = 0.0396$

Figure 2: $\xi_b(x)$
Scale factor $\xi_b(x) = (1-x)\{\rho_b(x) - x/b\} / \{1 + b\rho_b(x) - x\}$ entering asymptotic bias and variance of the modified gamma kernel estimator. Bandwidths of two DGPs from the simulation study in Section 3 are used.

where

$$\rho_b(x) = \begin{cases} \frac{1}{4}\left(\frac{x}{b}\right)^2 + 1 & \text{if } x \in [0; 2b)\,; \\ x/b & \text{if } x \in [2b; \infty)\,. \end{cases} \tag{8}$$

Note that the estimator fixes the size of the boundary region to the area from 0 to $2b$ independent of the shape of the underlying true density. The asymptotic bias of the modified gamma kernel estimator has the desired leading term

$$\text{Bias}\left\{\hat{f}_X^{\gamma m}(x)\right\} = \begin{cases} \xi_b(x)\, b f_X'(x) + o(b) & \text{if } x \in [0; 2b)\,; \\ \frac{1}{2} x f_X''(x)\, b + o(b) & \text{if } x \in [2b; \infty)\ , \end{cases} \tag{9}$$

where $\xi_b(x) = (1-x)\{\rho_b(x) - x/b\} / \{1 + b\rho_b(x) - x\}$ which is in $[0, 1]$ for standard choices of $b < 0.5$ for all $x \in [0, 2b)$ (see Figure 2). Its variance can be shown to have the same structure as in (5) with modified constant $\widetilde{\mathcal{C}}_b(x) = \frac{\Gamma\left(2\kappa^2+1\right)}{2^{1+2\kappa^2}\,\Gamma^2(\kappa^2+1)}$ and

$$\text{MSE}\left\{\hat{f}_X^{\gamma m}(x)\right\} \approx \begin{cases} \left\{\xi_b(x)\, b f_X'(x)\right\}^2 + \frac{f_X(x)}{nb}\,\widetilde{\mathcal{C}}_b(x) & \text{if } x/b \to \kappa; \\ \left\{\frac{1}{2} x f_X''(x)\, b\right\}^2 + \frac{f_X(x)}{2\sqrt{\pi}}\,(xb)^{-1/2}\, n^{-1} & \text{if } x/b \to \infty. \end{cases} \tag{10}$$

See Chen (2000) for details on the derivations.

2.2 Choice of Estimators for Different Density Shapes Near Zero

In general in the literature, the modified gamma kernel estimator has been strictly preferred to the standard gamma kernel version. While a simple comparison of their asymptotic variances, however, reveals that the constant for the modified estimator $\widetilde{\mathcal{C}}_b$ is strictly larger than the one for the standard gamma kernel $\mathcal{C}_b$ close to the boundary (for all $\kappa < 1$), this has been justified by the similarity to fixed kernels in asymptotic bias behavior of the modified gamma kernel as displayed in (9). Though, when carefully comparing the leading asymptotic bias terms of both gamma type estimators, we find that there are also cases where it is asymptotically more favorable to use the standard gamma kernel estimator. For all $x > 2b$ in the interior of the support with

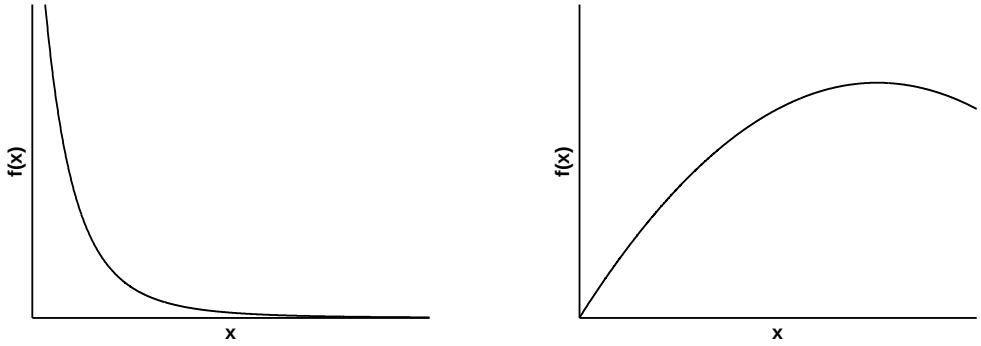
$$|0.5xf''(x)| > |f'(x) + 0.5xf''(x)| \tag{11}$$

the standard gamma kernel is preferable to the modified version. This occurs in particular for areas where the density satisfies the shape restrictions

$$0 < -f'(x)/f''(x) < x\ . \tag{12}$$

The lower bound is fulfilled for values x where f' and f'' have different sign, i.e. where the density f is either decreasing and convex or where it is concave and increasing. In the first case, it can be shown that if f has a pole at zero, then trivially also the upper bound of (12) is satisfied. If additionally f does not have any local maxima, the standard gamma kernel should be preferred to the modified version for the entire interior support (see Figure 3). Our simulation study below confirms that this is also of significant importance in finite samples in particular for smaller sample sizes. It can be easily shown, that a pole is a sufficient condition, but the same logic also applies to all densities with $f' < -c < 0$, c not too small and $f'' \geq 0$ close to the boundary.

Apart from these pronounced cases at the boundary, any density whose support is unbounded from the right will be convex and decreasing for large x in order to be integrable. In this situation, the asymptotic variance regimes are identical for both gamma type estimators. And in the asymptotic bias, independent from the rate of decay of f, the upper bound of (12) always holds in these regions. For very large x, however, slopes and curvature values are generally small yielding overall small biases for any kernel type estimator such that a measurable advantage of the standard versus modified gamma kernel estimator might disappear. Besides these convex cases, unimodal densities are concave around the mode, and increasing to the left of the mode (see Figure 3). Also in this area, the use of the standard gamma kernel estimator might be



(a) Pole at $x = 0$ and no local maxima (b) Unimodal

Figure 3: Density Shapes Favouring the Standard Gamma Kernel Estimator
Schematic densities for which the standard gamma kernel estimator in (2) and (3) should be preferred over the modified version in (2) and (8) according to the shape restriction (12). Left figure: condition (12) *is satisfied globally* for $x > 2b$. Right figure: condition (12) *can be satisfied locally* to the left of the mode.

recommendable. In finite samples, however, observed differences are rather small even in the extreme case of a strictly concave density between zero and the mode.

Moreover, on the boundary for $x \in [0, 2b)$, it can happen that

$$|0.5xf''(x)| > |f'(x)\xi_b(x)| \tag{13}$$

where the standard gamma kernel estimator performs better than the modified one. This happens in particular if $|f'(x)/f''(x)| < 0.5\,x$, since $\xi < 1$ for $x > 0$. Though as this area is vanishingly small, its influence on the overall estimation results is negligible (compare the simulation results in Section 3).

In practice, it is therefore important to detect pole situations in advance in order to choose the best performing estimator among standard and modified gamma kernel estimators. We propose a simple but reliable measure to check for poles as opposed to standard cases. If f has a pole at zero, it is the relative convergence and consistency of the estimator $\hat{f}$ which is of main importance in order to judge if the correct order of decay is detected. See e.g. Robinson and Henry (2003) for how this is important regarding consistent estimation of the long memory parameter in long range dependent time series. Thus it must hold that $|\hat{f}(x)/f(x) - 1| = o_P(1)$. The governing term in the stochastic expansion for the right hand side controlling convergence is $x\frac{f'_X(x)}{f_X(x)}$, which we write as $xD(x)$ (See the proof of Theorem 5.3. in Bouezmarni and Scaillet (2005)). The practically most important pole situations occur for densities which have or can be bounded by densities with hypergeometric decay from zero, i.e. $f(x) = bx^{-\alpha}$ with $b < 0$ and $0 < \alpha < 1$ (the cases with $\alpha > 1$ are excluded by f being a density). Here the quantity $xD(x)$ equals the constant $-\alpha$ irrespective of the scaling b.

For distinguishing a pole situation from a no pole situation, it is favorable to study $D(x)$ directly to get sufficient power of the criterion against alternatives. Therefore we estimate $D(x)$ by exploiting the simple relation

$$D(x) = \frac{f'_X(x)}{f_X(x)} = \frac{d}{dx} \ln f_X(x). \tag{14}$$

Note that for x approaching 0, in a pole situation $D(x)$ is significantly negative, approaching infinity at rate $\frac{-\alpha}{x}$ in case of densities decreasing with hypergeometric speed and -1 for exponential type behavior. In all other settings where the modified gamma kernel is the method of choice $D(x)$ is significantly positive. As a criterion, $D(x)$ combines properties of the density and its slope to distinguish the pole situation from other density shapes. This is more powerful than checking density and slope separately in isolation. In practice, $D(x)$ can be estimated by the difference quotient based on modified gamma kernels

$$\widehat{D}(x) = \frac{\ln \hat{f}_X^{\gamma m}(x+b) - \ln \hat{f}_X^{\gamma m}(x)}{b}, \tag{15}$$

where $b > 0$ is the same bandwidth as for the density estimates at x and $x + b$. For the practical scope of this paper it is sufficient to work with a rough criterion checking if $\widehat{D}(x)$ is significantly negative or not. Developing a novel formal test for H_0 of a hypergeometric pole situation is beyond the scope of this paper. Though, we conjecture that using the results in Fernandes and Grammig (2005) for specification testing in the simple density case, the corresponding asymptotic distribution of the centered test statistic $nb^2 \left(\widehat{D}(x) + \frac{\alpha}{x}\right)$ could be derived. However, as calculations are quite involved and should be complemented with a valid bootstrap approximation scheme for finite samples, we leave this for future research and a paper on its own.

2.3 Refined Estimation with Modified Gamma Kernels

In cases where we can exclude a pole at the boundary, the modified gamma kernel generally should be the method of choice in terms of best asymptotic performance. Though in the literature, its chosen form in particular in the boundary region has mainly been justified by (computational) convenience. Our simulation results, however, clearly indicate that alternative slightly more flexible specifications can significantly improve upon the performance of standard modified gamma kernels.

In particular, we propose simple refined versions of the modified gamma kernel, where an additional specification parameter c allows for higher accuracy if appropriately chosen in a data-driven way. We study two types of refined modified gamma kernels, i.e.

$$\rho_b^{v_{\text{I}}}(x) = \begin{cases} \left[\frac{1}{4}\left(\frac{x}{bc}\right)^2 + 1\right]\,[c + 2b\,(1-c)] & \text{if } x \in [0; 2bc)\,; \\ \frac{x}{bc}\,(c + 2b - x) & \text{if } x \in [2bc; 2b)\,; \\ x/b & \text{if } x \in [2b; \infty)\,, \end{cases} \tag{16}$$

and

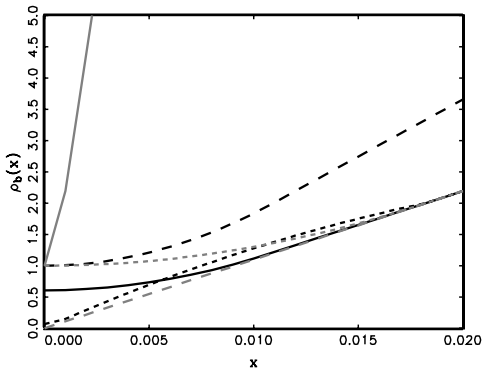
$$\rho_b^{v_{\text{II}}}(x) = \begin{cases} \frac{1}{4}\left(\frac{x}{bc}\right)^2 + 1 & \text{if } x \in [0; 2bc)\,; \\ x/(bc) & \text{if } x \in [2bc; \infty)\,, \end{cases} \tag{17}$$

where $c \in (0, 1]$ with $c = 1$ yielding the original parametrization in both cases. Specification v_{I} shifts the boundary regime below one and introduces a flexible quadratic middle part. In the latter regime, for $\rho_b(x) > x/b$ we have that $x/b < \rho_b^{v_{\text{I}}}(x) < \rho_b(x)$, if

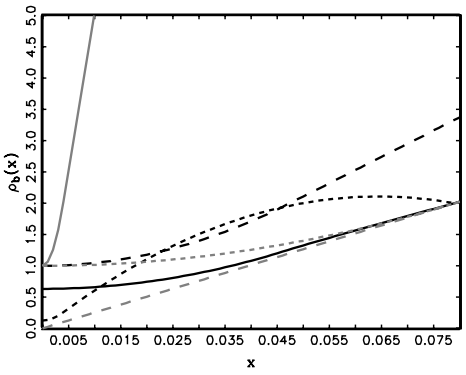
$$\frac{x}{b}\,\frac{2b - x}{\rho_b(x) - x/b} < c < 1; \qquad x \in [2bc; 2b)\,, \tag{18}$$

where $\rho_b(x)$ is defined as in (8). Importantly, fulfilment of the condition implies that specification v_{I} is closer to the theoretically optimal situation with the mean of the kernel being at the observation point as compared to the original modified gamma kernel. The second alternative, v_{II}, keeps two regimes and the general structure of the original specification but shrinks the boundary region proportionally to the value of the tuning parameter c. This modification also affects asymptotics in the interior of the support, as the mean of the kernel equals x/c and, hence, only in the trivial case $c = 1$ coincides with the point of estimation.

Figure 4 shows plots of $\rho_b(x)$ based on the specification proposed by Chen (2000) along with the above refined versions for different values of the constant c and using the bandwidths of two DGPs from the simulation study in Section 3. In addition, we include x/b, which corresponds to the interior component of the original specification and implies a gamma kernel with mean at the point of estimation. In its middle regime, $\rho_b^{v_{\text{I}}}$ is closer to x/b than the original specification for $c = 0.6$ in the right and for both values of c in the left figure, as in these cases condition (18) is satisfied. Close to the boundary, the shape function of specification v_{I} takes values below one, implying that the resulting gamma densities and thus, gamma kernels are unbounded at the origin (see Figure 5). But the finite sample study below clearly reveals that this specification outperforms the original modified and the refined version v_{II} in all settings where a modified gamma kernel should be applied.

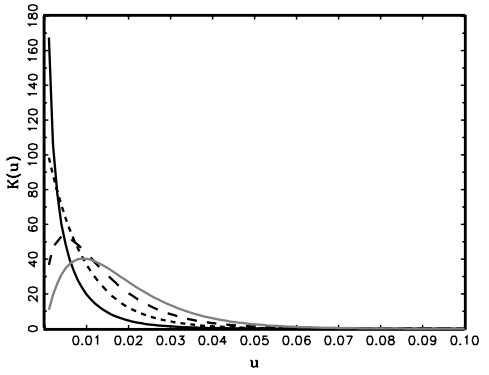


(a) $b = 0.0091$

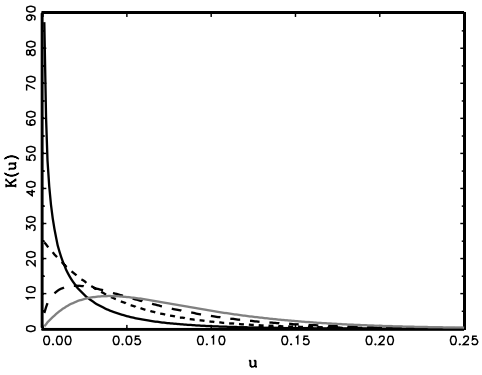


(b) $b = 0.0396$

Figure 4: Shape Parameter $\rho_{\mathbf{b}}(\mathbf{x})$ of Modified Gamma Kernel
Black solid and short-dashed line: $c = 0.6$ and $c = 0.05$ for refined kernel v_{I} (see (16)). Grey solid and short-dashed line: $c = 0.6$ and $c = 0.05$ for refined kernel v_{II} (see (17)). Black long-dashed line: original modified kernel (see (8)). Grey long-dashed line: interior regime of original specification and refined version v_{I}, x/b. Bandwidths of the modified gamma kernel estimator for two DGPs from the simulation study in Section 3 are used.



(a) $b = 0.0091$



(b) $b = 0.0396$

Figure 5: Gamma Kernel Depending on Shape Parameter
Gamma kernel $K^{\gamma}_{\rho_b(x),b}(u)$ for different values of shape parameter ρ_b. Black solid line: $\rho_b = 0.5$. Black short-dashed line: $\rho_b = 1$. Black long-dashed line: $\rho_b = 1.5$. Grey solid line: $\rho_b = 2$. Bandwidths of the modified gamma kernel estimator for two DGPs from the simulation study in Section 3 are used.

For a feasible implementation of these refined estimators, we provide an automatic procedure to select the tuning parameter c: for a fixed bandwidth b, we determine the threshold $x^c = b\,\kappa$ for which the two MSE expressions of the modified gamma kernel in (10) coincide. Then the optimal value of c can then be obtained as $c^* = \kappa/2 = x^c/(2\,b)$. In practice, this approach requires minimizing the objective function in $0 \leq x \leq 2\,b$

$$\mathcal{M}(x) = \left\{ \left[\xi_b(x)\, b f'_X(x)\right]^2 + \frac{f_X(x)}{nb}\, \mathcal{C}_b(x) - \left[\frac{1}{2} x f''_X(x)\, b\right]^2 \right. \\ \left. - \frac{f_X(x)}{2\sqrt{\pi}}\, (xb)^{-1/2}\, n^{-1} \right\}^2 . \tag{19}$$

Evaluation of the objective function requires estimates of the unknown density and its first two derivatives. $f_X(x)$ and $f'_X(x) = D(x)\ f_X(x)$ can be estimated using the original modified gamma kernel. An estimate of $f''_X(x)$ can be obtained by differentiating, e.g., the simple gamma kernel estimator:

$$\begin{aligned} \hat{f}''^{\gamma}_X(x) &= \frac{1}{n\,b^2} \sum_{i=1}^{n} \frac{\partial^2}{\partial x^2}\ K^{\gamma}_{x/b+1,b}\,(X_i)\,, \\ &= \frac{1}{n\,b^2} \sum_{i=1}^{n} K^{\gamma}_{x/b+1,b}\,(X_i)\ \left\{ \left[\ln(X_i/b) - \psi(x/b+1)\right]^2 - \psi_1(x/b+1) \right\}, \end{aligned} \tag{20}$$

where $\psi(u) = (d/du)\,\ln\Gamma(u)$ and $\psi_1(u) = \left(d^2/du^2\right)\,\ln\Gamma(u)$ denote the digamma and trigamma function, respectively.

3 Simulation Study

For a complete picture, we compare basic, modified and refined gamma kernel estimators for a wide range of test densities representing all potential types of shapes near the boundary to standard boundary corrected versions of the symmetric fixed kernel density estimator (1). This also complements simulation studies in the literature for the two standard gamma kernels such as Chen (2000) which only focusses on very specific density settings and Hagmann and Scaillet (2007) which is restrictive in the range of fixed boundary kernel competitors.

All fixed kernels are based on the Epanechnikov kernel $K(u) = 3/4(1-u^2)\mathbb{I}(-1 \leq u \leq 1)$, where $\mathbb{I}(\cdot)$ denotes an indicator function limiting the support of K to $[-1, 1]$. In particular we report results for the following five competing fixed kernel adjustments. The reflection

estimator proposed by Schuster (1958) has the form

$$\hat{f}_X^{\text{Refl}}(x) = \sum_{i=1} K\left(\frac{x - X_i}{b}\right) + K\left(\frac{x + X_i}{b}\right). \tag{21}$$

In the inside of the support for $x \geq 2h$, it coincides with the standard kernel density estimator $\hat{f}_X^{\text{Fixed}}$ in (1). In the cut-and-normalized estimator $\hat{f}_X^{\text{CaN}}$ introduced by Gasser and Müller (1979), the kernel function K on the boundary is truncated at $\nu := x/b$ and normalized ensuring integration to unity. For the Epanechnikov kernel, it has the form

$$K^{\text{CaN}}(u) = \frac{\left(1 - u^2\right)}{\int_{-1}^{\nu}\left(1 - u^2\right) du} \mathbb{1}_{\{-1 \leq u \leq \nu\}}. \tag{22}$$

General boundary corrected estimators $\hat{f}_X^{\text{Bound}}$ (see, e.g., Jones, 1993) replace the standard kernel function on the boundary by a modified version K^{Bound}, which is chosen to meet the following conditions

$$\int_{\nu}^{1} K^{\text{Bound}}(u)\, du = 0, \quad \int_{-1}^{\nu} K^{\text{Bound}}(u)\, du < \infty, \quad \int_{-1}^{\nu} K^{\text{Bound}}(u)\, u\, du = 0. \tag{23}$$

We use the boundary kernel based on the Epanechnikov kernel, which has the following form

$$K^{\text{Bound}}(u) = 12\frac{(1+u)}{(1+\nu)^4}\left[\frac{3\nu^2 - 2\nu + 1}{2} + u\,(1 - 2u)\right] \mathbb{1}_{\{-1 \leq u \leq \nu\}}. \tag{24}$$

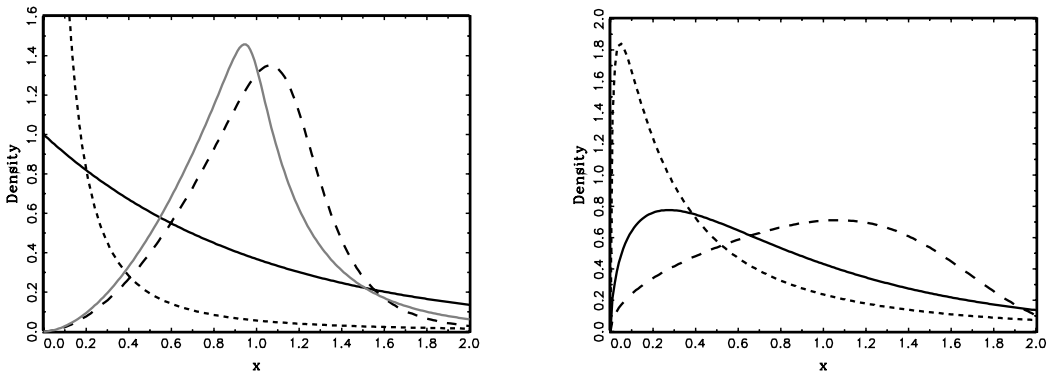
A method that corrects for the possible negativity of the boundary kernel estimates was proposed e.g. by Jones and Foster (1996). The estimator has the following form

$$\hat{f}_X^{\text{JF}}(x) = \hat{f}_X^{\text{CaN}}(x)\, \exp\left\{\frac{\hat{f}_X^{\text{Bound}}(x)}{\hat{f}_X^{\text{CaN}}(x)} - 1\right\}. \tag{25}$$

We compare the performance of the estimators for seven different density functions with nonnegative support, which reflect the variety of practically relevant types of shapes on left-bounded support. The densities of DGP 1 and DGP 2 are entirely decreasing and convex with DGP 2 exhibiting pole behavior at zero. The remaining densities are increasing near the boundary. For DGP 3 and 4, the density is locally convex in the boundary region, while for 5,6 and 7 it is concave with varying degree of steepness. The corresponding density shapes are depicted in Figure 6. All DGPs are generated from different specifications of the flexible generalized F distribution, which is based on a gamma mixture of the generalized gamma

Table 1: Data Generating Processes for Simulation Study
DGPs are generated from i.i.d. samples of different specifications of the generalized F distribution (26). We use the following tuples of shape parameters a, m and η. The scale parameter λ is chosen such that the expectation of each DGP is normalized to one. Corresponding shapes of the densities are depicted in Figure 6.

DGP	a	m	η
1	1	1	∞
2	0.9	0.7	1.2
3	14	0.2	0.5
4	35	0.08	0.1
5	0.8	2	∞
6	0.55	3	5
7	5	0.3	∞



(a) DGPs 1 to 4 (b) DGPs 5 to 7

Figure 6: Densities Corresponding to Different DGPs
Densities corresponding to tuples of shape parameters in Table 1. Left: DGP 1 (black solid), DGP 2 (black short-dashed), DGP 3 (black long-dashed) and DGP 4 (grey solid). Right: DGP 5 (black solid), DGP 6 (black short-dashed) and DGP 7 (black long-dashed).

distribution (see, e.g., Lancaster, 1997). Its marginal density function is given by

$$f_x(x) = \frac{a\, x^{a\,m-1}\left[\eta + (x/\lambda)^a\right]^{(-\eta-m)}\eta^\eta}{\lambda^{a\,m}\,\mathcal{B}(m,\eta)}, \tag{26}$$

where $a > 0, m > 0, \eta > 0$ and $\lambda > 0$. $\mathcal{B}(\cdot)$ describes the full Beta function with $\mathcal{B}(m,\eta) := \frac{\Gamma(m)\Gamma(\eta)}{\Gamma(m+\eta)}$. Table 1 shows the values of the shape parameters a, m and η for the seven DGPs considered. To ensure comparability across the different DGPs, the expectation is restricted to one by setting the scale parameter λ equal to

$$\lambda^{-1} = \eta^{-1/a}\frac{\Gamma(m)\,\Gamma(\eta)}{\Gamma(m+1/a)\,\Gamma(\eta-1/a)}. \tag{27}$$

Table 2: Bandwidths
Bandwidths chosen by minimizing the mean integrated squared error (28) using simulated samples with $n = 200$. The following estimators are used. Gamma and $\text{Gamma}_{\text{mod}}$: basic and modified gamma kernel estimator. Fixed: fixed kernel estimator based on the epanechnikov kernel. Refl: reflection estimator. CaN: cut-and-normalized estimator. Bound: boundary kernel estimator. JF: Jones-Foster estimator.

Est.	DGP 1	DGP 2	DGP 3	DGP 4	DGP 5	DGP 6	DGP 7
				$\mathbf{n = 400}$			
Gam	0.0768	0.0042	0.0096	0.0088	0.0571	0.0319	0.0308
Gam_{m}	0.1163	0.0166	0.0099	0.0091	0.0634	0.0396	0.0336
Fixed	0.1821	0.0176	0.2058	0.1820	0.2678	0.0888	0.4015
Refl	0.4643	0.0391	0.2054	0.1818	0.3569	0.2144	0.3609
CaN	0.4223	0.0307	0.2044	0.1808	0.4308	0.1868	0.3494
Bound	0.7471	0.0086	0.2064	0.1825	0.3824	0.3141	0.4024
JF	0.4223	0.0307	0.2044	0.1808	0.4308	0.1868	0.3494
				$\mathbf{n = 4000}$			
Gam	0.0485	0.0027	0.0061	0.0055	0.0360	0.0201	0.0195
Gam_{m}	0.0734	0.0104	0.0062	0.0058	0.0400	0.0250	0.0212
Fixed	0.1149	0.0111	0.1299	0.1148	0.1689	0.0561	0.2533
Refl	0.2930	0.0247	0.1296	0.1147	0.2252	0.1353	0.2277
CaN	0.2664	0.0194	0.1290	0.1141	0.2718	0.1179	0.2204
Bound	0.4714	0.0054	0.1302	0.1152	0.2413	0.1982	0.2539
JF	0.2664	0.0194	0.1290	0.1141	0.2718	0.1179	0.2204

From each DGP, we draw 1000 random samples $\{X_i\}_{i=1}^n$ of size $n = 400$ and $n = 4000$. To minimize the effects of sampling variation, we follow Zhang (2010) and select the optimal bandwidth for each estimator and DGP by minimizing the integrated mean squared error (IMSE)

$$\text{IMSE}\Big\{\hat{f}_X(x)\Big\} = \frac{1}{1000}\sum_{r=1}^{1000}\int_\tau^\infty \Big\{f_X(x) - \hat{f}_X^r(x)\Big\}^2\, dx, \tag{28}$$

where τ is a small number and $\hat{f}_X^r(x)$ denotes the density estimate for the rth simulated sample. Bandwidth selection is conducted using the sample size $n_b = 200$, which requires multiplying the resulting bandwidths by the factor $(n/n_b)^{-1/5}$ for the subsequent analysis. The rescaled bandwidths for $n = 400$ and $n = 4000$ are reported in Table 2. The two gamma kernel estimators estimators exhibit noticeably smaller bandwidths in comparison to the other estimators, which can be explained by the reduced variance of the former in the interior part of the support.

Table 3 and 4 report the IMSEs of the different estimators for the seven DGPs and two samples sizes. IMSEs are computed over the interval $[0, 2]$. For DGPs 3 to 7, we additionally consider shorter intervals that encompass and exclude the mode of the distribution, respectively. Four major results are apparent. First, in a general comparison with the standard fixed kernel

adjustments, gamma kernel estimators appear to offer a satisfactory performance. They are clearly more precise for DGPs 2, 5 and 6, while yielding similar (or only slightly higher) IMSEs in the remaining cases. In particular, the single largest improvement in favor of the (basic) gamma kernel is achieved in the pole scenario of DGP 2. Note that when the applied polynomial transformation for the method of Marron and Ruppert (1994) was close to the true pole behavior, we could also construct a fixed kernel estimator with a similar or even better precision for DGP 2. Corresponding results, however, were not robust to deviations of the transformation from the true density shape near zero implying a high risk of extremely large IMSEs in practice. Due to the tailored construction of the above method for pole situations only, also the IMSE records for any other form of the density were largely inferior to the rest. We therefore do not report results for this estimator.

Second, the simulation evidence confirms the relationship between the performance of the basic and modified gamma kernel estimator and the shape of the underlying density. If the latter has first and second derivatives of opposing sign in the interior of the support, as is the case for DGPs 3, 4, 6 and 7 in the subinterval to the right of the mode, the basic gamma kernel yields noticeably lower IMSEs (see bottom panel). When considering the entire interval $[0, 2]$, the basic gamma kernel is more precise for DGPs 2 and 6 with the most striking gains occurring in the former scenario, as it corresponds to a globally convex density with pole at zero. Finally, the above relation breaks down within the boundary region due to the involvement of the factor $\xi_b(x)$ in the asymptotic bias (see (9)). For DGPs 5 and 6, the modified gamma kernel implies lower IMSEs over the leftmost subinterval in which the corresponding densities are increasing and concave (see lower top panel).

The simulation results stress the importance of determining pole situations in advance, which can be achieved by examining the normalized density derivative $D(x)$ in the boundary region. We estimate the latter as in (15) using the modified gamma kernel for the points $x \in \{0, b, 2b\}$, where b is the bandwidth of the corresponding estimator. Table 5 reports descriptive statistics of the estimates for $n = 400$. In case of DGP 2, these estimates are highly negative at all three points, demonstrating that our simple method is able to detect a pole at zero. We obtain negative estimates at all or at distinct points also for DGPs 1 and 6 but their magnitude is considerably lower than in the above true pole scenario.

As was argued in Section 2.2, whenever no pole situation has been detected, the modified gamma kernel in its original or refined form should be used. The IMSEs of the three corresponding estimators are displayed in Table 6. For the refined kernels v_{I} and v_{II}, a set of values for the threshold c is considered. To ensure comparability, we apply the bandwidths b of the original modified gamma kernel to all estimators and also use $2b$ as the upper integration limit in the IMSE calculations. The main finding is that the refined kernel v_{I}, exhibits a high precision in all

Table 3: Integrated Mean-Squared Errors (n=400)
For DGP 3, 4 and 7, limits x_1 and x_2 are chosen such that $[x_1, x_2]$ encompasses the mode. $x_1 \in \{0.888, 0.600, 0.500\}$ and $x_2 \in \{1.292, 1.090, 1.700\}$. For DGP 5 and 6, $x_1 = x_2$ is chosen such that $[0, x_1]$ includes the mode, where $x_1 = x_2 \in \{0.600, 0.200\}$. The following estimators are used. Gamma and Gamma$_{\text{mod}}$: basic and modified gamma kernel estimator. Fixed: fixed kernel estimator based on the epanechnikov kernel. Refl: reflection estimator. CaN: cut-and-normalized estimator. Bound: boundary kernel estimator. JF: Jones-Foster estimator. Results are re-scaled by the factor 10^3.

Estimator	**DGP 1**	**DGP 2**	**DGP 3**	**DGP 4**	**DGP 5**	**DGP 6**	**DGP 7**
				$0 \leq x \leq 2$			
Gamma	4.185	75.151	8.857	9.759	5.828	14.572	5.286
Gamma$_{\text{mod}}$	3.575	279.733	8.720	9.623	5.422	14.879	4.575
Fixed	17.659	287.118	7.293	8.63	7.182	26.125	3.462
Refl	3.854	320.621	7.309	8.645	9.815	16.738	4.943
CaN	4.171	177.666	7.340	8.67	10.16	17.46	5.308
Bound	3.039	356.335	7.279	8.621	6.651	18.096	3.594
JF	4.259	923.101	7.274	8.612	7.235	17.629	3.792
				$0 \leq x \leq x_1$			
Gamma			2.880	1.729	4.691	11.902	1.804
Gamma$_{\text{mod}}$			2.553	1.502	4.330	11.59	1.501
Fixed			2.015	1.125	4.991	17.248	0.934
Refl			2.037	1.143	8.285	13.032	2.309
CaN			2.083	1.178	8.956	13.339	2.619
Bound			1.993	1.109	5.25	14.616	1.066
JF			2.017	1.120	6.030	13.508	1.104
				$x_1 \leq x \leq x_2$			
Gamma			4.702	6.416			3.224
Gamma$_{\text{mod}}$			4.304	6.038			2.549
Fixed			3.788	5.316			2.190
Refl			3.783	5.312			2.303
CaN			3.770	5.295			2.358
Bound			3.796	5.326			2.189
JF			3.770	5.295			2.358
				$x_2 \leq x \leq 2$			
Gamma			1.220	1.573	1.136	2.670	0.258
Gamma$_{\text{mod}}$			1.806	2.039	1.092	3.289	0.525
Fixed			1.433	2.134	2.190	8.877	0.339
Refl			1.432	2.136	1.530	3.706	0.330
CaN			1.431	2.141	1.204	4.122	0.331
Bound			1.434	2.131	1.401	3.479	0.339
JF			1.431	2.141	1.204	4.122	0.331

Table 4: Integrated Mean-Squared Errors (n=4000)
For DGP 3, 4 and 7, limits x_1 and x_2 are chosen such that $[x_1, x_2]$ encompasses the mode. $x_1 \in \{0.888, 0.600, 0.500\}$ and $x_2 \in \{1.292, 1.090, 1.700\}$. For DGP 5 and 6, $x_1 = x_2$ is chosen such that $[0, x_1]$ includes the mode, where $x_1 = x_2 \in \{0.600, 0.200\}$. The following estimators are used. Gamma and Gamma$_{\text{mod}}$: basic and modified gamma kernel estimator. Fixed: fixed kernel estimator based on the epanechnikov kernel. Refl: reflection estimator. CaN: cut-and-normalized estimator. Bound: boundary kernel estimator. JF: Jones-Foster estimator. Results are re-scaled by the factor 10^3.

Estimator	**DGP 1**	**DGP 2**	**DGP 3**	**DGP 4**	**DGP 5**	**DGP 6**	**DGP 7**
				$0 \leq x \leq 2$			
Gamma	1.011	21.601	2.246	2.758	2.069	6.924	1.218
Gamma$_{\text{mod}}$	0.679	161.086	2.237	2.76	1.734	7.307	0.986
Fixed	8.122	46.145	1.331	1.683	1.960	8.504	0.734
Refl	0.803	154.026	1.332	1.684	4.415	10.797	1.203
CaN	0.890	56.600	1.334	1.682	5.443	10.075	1.318
Bound	0.471	62.457	1.329	1.684	2.106	10.593	0.684
JF	0.585	616.087	1.327	1.676	2.663	8.142	0.721
				$0 \leq x \leq x_1$			
Gamma			0.450	0.271	1.831	6.484	0.322
Gamma$_{\text{mod}}$			0.349	0.207	1.532	6.567	0.216
Fixed			0.360	0.193	1.592	7.029	0.267
Refl			0.362	0.194	4.149	10.148	0.723
CaN			0.367	0.198	5.227	9.364	0.830
Bound			0.357	0.191	1.859	9.886	0.216
JF			0.360	0.192	2.447	7.431	0.234
				$x_1 \leq x \leq x_2$			
Gamma			1.495	2.200			0.826
Gamma$_{\text{mod}}$			1.337	2.055			0.567
Fixed			0.721	1.118			0.405
Refl			0.720	1.116			0.421
CaN			0.717	1.11			0.430
Bound			0.722	1.121			0.405
JF			0.717	1.110			0.430
				$x_2 \leq x \leq 2$			
Gamma			0.293	0.281	0.238	0.440	0.071
Gamma$_{\text{mod}}$			0.542	0.488	0.202	0.741	0.202
Fixed			0.239	0.361	0.369	1.475	0.062
Refl			0.239	0.361	0.266	0.650	0.058
CaN			0.239	0.362	0.216	0.711	0.058
Bound			0.239	0.361	0.247	0.706	0.062
JF			0.239	0.362	0.216	0.711	0.058

Table 5: Summary Statistics of Normalized Density Derivative
Descriptives for estimate of the ratio $D(x) := f'(x)/f(x)$ based on the modified gamma kernel. The estimator from equation (15) is used. $n = 400$.

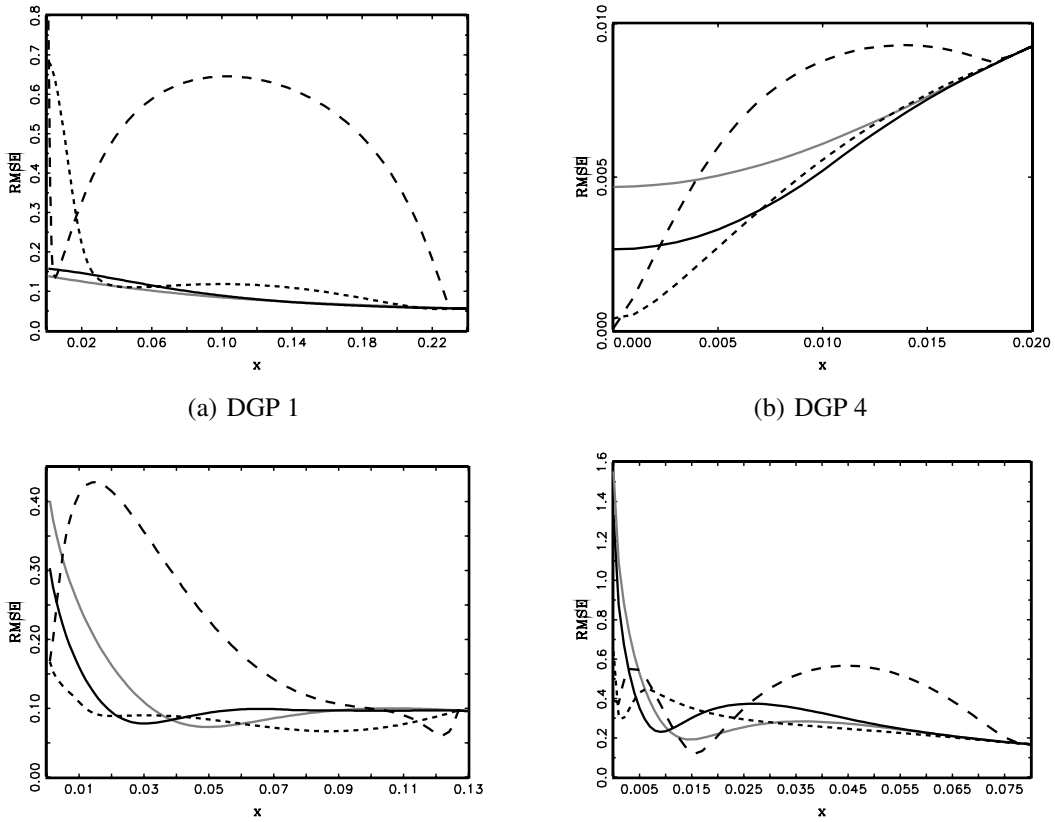
x	DGP 1	DGP 2	DGP 3	DGP 4	DGP 5	DGP 6	DGP 7
				Mean			
0	-0.226	-13.982	73.923	80.433	1.899	1.265	4.71
b	-0.711	-27.341	182.844	205.189	3.26	0.756	8.741
$2b$	-0.941	-23.57	185.301	198.408	1.892	-1.558	6.763
				1st Quartile			
0	-0.399	-15.110	67.995	73.900	1.473	0.737	2.707
b	-1.017	-29.185	165.094	184.470	2.467	-0.257	4.562
$2b$	-1.180	-25.188	161.913	172.373	1.247	-2.386	3.800
				Median			
0	-0.220	-13.928	75.557	82.861	1.922	1.284	4.89
b	-0.727	-27.382	187.737	212.998	3.209	0.710	8.395
$2b$	-0.939	-23.557	191.768	208.193	1.865	-1.568	6.629
				3rd Quartile			
0	-0.059	-12.794	81.790	89.863	2.350	1.798	7.020
b	-0.421	-25.602	206.611	235.238	4.025	1.764	12.854
$2b$	-0.703	-21.906	216.830	236.042	2.493	-0.786	9.496

situations, for which the modified kernel should be considered, i.e. all DGPs except the second one. The improvement with respect to the original specification is particularly pronounced, accompanied by low optimal values of the constant c, in case of densities with concave shape near the boundary, as in DGPs 5,6 and 7. Further, the refined kernel v_{II} is at roughly the same level as the traditional parameterization and even yields the lowest IMSE for DGP 1 when $n = 400$. However, recall that this specification makes the boundary region smaller and has neither its mean nor mode at the point of estimation for $x > 2bc$ (see Section 2.3). These properties cause a vastly lower precision compared to the other specifications in the interior part of the support. Corresponding simulation results are available upon request.

Finally, Table 6 shows that the performance of the refined modified gamma kernel estimators is highly dependent on the value of the threshold c. This is underlined by Figure 7, which depicts plots of the root mean squared errors (RMSEs) of the estimators based on the original modified gamma kernel and the refined version v_{I} for several values of c. The plots also illustrate that the choice of c determines for which part of the support the original estimator can or cannot be outperformed. E.g. in case of DGP 4, specification v_{I} almost consistently exhibits lower RMSEs for $c = 0.6$ or $c = 0.1$, while providing precise estimates only in a small neighborhood of $x = 0$ if $c = 0.01$.

Table 6: Integrated MSE for refined Modified Gamma KDE $v_{\mathbf{I}}$ & $v_{\mathbf{II}}$
Refined modified gamma kernel estimators as defined in (7) and (16) or (17). $c = 1^*$ denotes original modified gamma kernel from (7) and (8). IMSEs are computed from 0 to $2b$. Bandwidths of the original modified gamma kernel are used. Results for $n = 400$ and $n = 4000$ are rescaled by the factor 10^4 and 10^5, respectively.

	c	DGP 1	DGP 2	DGP 3	DGP 4	DGP 5	DGP 6	DGP 7
					$\mathbf{n = 400}$			
	1^*	18.230	1971.501	0.005	0.009	24.452	99.402	2.713
$v_{\underline{\mathbf{I}}}$	0.9	18.389	2238.201	0.005	0.008	22.112	98.176	2.555
	0.8	18.927	3280.673	0.005	0.008	19.936	97.139	2.416
	0.7	19.976	6045.533	0.004	0.007	17.94	96.462	2.299
	0.6	21.776	13836.526	0.004	0.007	16.123	96.310	2.201
	0.3	39.301	16309119	0.004	0.007	11.443	96.476	1.991
	0.1	74.542	3193544.7	0.004	0.007	9.150	69.093	1.872
	0.05	99.107	29681669	0.004	0.007	18.179	45.087	2.162
	0.01	623.499	3640.61	0.008	0.012	72.511	144.333	10.431
$v_{\underline{\mathbf{II}}}$	0.9	17.878	1453.286	0.006	0.01	23.301	97.185	2.722
	0.8	18.157	1040.206	0.007	0.011	22.507	95.373	2.78
	0.7	19.768	837.209	0.008	0.013	22.47	94.715	2.937
	0.6	24.05	996.348	0.009	0.015	23.755	96.878	3.285
	0.3	104.500	6653.866	0.027	0.042	43.068	195.018	8.734
	0.1	629.167	24190.131	0.364	0.530	81.172	874.974	48.989
	0.05	1115.938	32319.507	3.483	5.224	170.476	1460.694	105.8
	0.01	1688.05	38771.282	69.245	70.217	321.568	2185.416	48.598
					$\mathbf{n = 4000}$			
	1^*	25.397	10185.573	0.002	0.002	94.207	594.231	5.050
$v_{\underline{\mathbf{I}}}$	0.9	23.982	16235.41	0.002	0.002	80.85	562.997	4.439
	0.8	23.446	29262.905	0.002	0.001	68.844	533.361	3.912
	0.7	24.057	55516.877	0.001	0.001	58.242	506.235	3.472
	0.6	26.498	110274.880	0.001	0.001	49.013	483.124	3.115
	0.3	109.444	16058082	0.001	0.001	27.206	431.343	2.493
	0.1	721.532	561573.060	0.001	0.001	11.743	283.302	2.029
	0.05	740.051	57227048	0.001	0.001	31.993	131.333	2.011
	0.01	1716.071	6091.886	0.002	0.002	523.996	244.364	16.592
$v_{\underline{\mathbf{II}}}$	0.9	25.709	5653.425	0.002	0.002	88.817	571.585	5.026
	0.8	27.903	2185.219	0.002	0.002	86.931	550.769	5.212
	0.7	33.997	682.415	0.003	0.003	91.895	534.382	5.827
	0.6	47.993	2474.760	0.003	0.003	108.945	526.934	7.271
	0.3	306.934	53804.758	0.009	0.011	334.800	750.698	31.256
	0.1	2502.053	207647.900	0.145	0.209	634.147	3412.288	212.589
	0.05	5517.404	277217.360	1.424	2.168	757.340	6670.104	540.434
	0.01	10682.698	335603.280	373.201	424.930	1566.361	12161.535	433.552



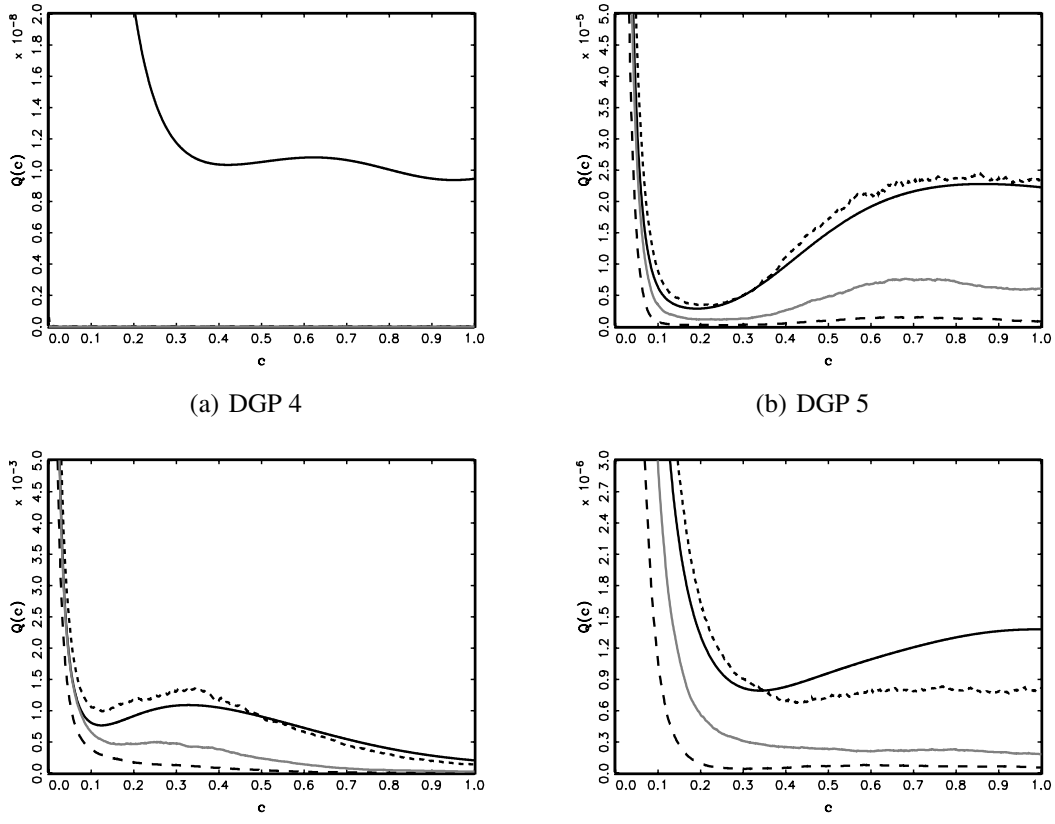
(a) DGP 1

(b) DGP 4

(c) DGP 5

(d) DGP 6

Figure 7: RMSE of Refined Modified Gamma KDE v_{I}
Refined modified gamma kernel v_{I} as defined in (7) and (16). Black solid line: $c = 0.6$. Black short-dashed line: $c = 0.1$. Black long-dashed line: $c = 0.01$. Grey solid line: $c = 1^*$ (original modified gamma kernel). $n = 400$. Bandwidths of the original modified gamma kernel are used.



(a) DGP 4

(b) DGP 5

(c) DGP 6

(d) DGP 7

Figure 8: Objective Function for Choice of c
Mean (black solid), median (grey solid), first (black long-dashed) and third (black short-dashed) quartile of (transformed) objective function for choice of the constant c in the refined modified gamma kernel v_{I} as defined in (7) and (16). The transformed objective function is $\mathcal{Q}(c) := \mathcal{M}(2bc)$, where $\mathcal{M}(x)$ is given in (19) and b denotes the bandwidth of the original modified gamma kernel. $n = 400$.

Since, in practice, the constant c has to be chosen ex-ante, we examine how well the data-driven method introduced in Section 2.3 can "track" the optimal values according to Table 6. We estimate the unknown quantities entering the objective function (19) as was outlined above. Figure 8 displays averages, medians and quartiles of the resulting estimates of the (transformed) objective function $\mathcal{Q}(c) := \mathcal{M}(2bc)$, where b is the bandwidth of the modified gamma kernel. A comparison with the IMSEs from Table 6 shows that for DGPs 5, 6 and 7 the means, in particular, have local minima close to the values of c yielding the lowest IMSEs of the estimator based on the refined modified kernel v_{I}. For DGP 4, finding a unique minimum is more difficult, which corresponds to the fact that several values of c imply equal IMSEs. These results suggest that, if suitable starting values are chosen, the above approach can determine the optimal value of c with reasonable precision.

4 Application: Intraday Trading Volumes and Return Volatility

To demonstrate the practical relevance of the above methodology, we employ the latter to compute semiparametric estimates of the conditional distributions of high-frequency trading volumes and return volatilities of stocks traded at the New York Stock Exchange (NYSE). Modeling high-frequency trading volumes is, for instance, relevant for trading strategies replicating the (daily) volume weighted average price (VWAP). Estimates of conditional volatility distributions are crucial for the pricing of volatility derivatives. Examples include options and futures on the CBOE Volatility Index (VIX) trading at the Chicago Board Options Exchange (CBOE).

4.1 Modeling Intraday Trading Volumes

We consider transaction data for Citigroup from the last trading week of February 2009. The raw sample is filtered by deleting transactions that occurred outside regular trading hours from 9:30 am to 4:00 pm, computing cumulated trading volumes over 15 second intervals and removing zero observations, which yields a sample size of 7452.[1] To capture the well-known intraday seasonalities of high-frequency trading variables (see, e.g., Hautsch (2004) for an overview), we divide the cumulated volumes by a seasonality component which is pre-estimated employing a cubic spline function.

An important property of the resulting (deseasonalized) trading volumes is the strong persistence, as evidenced by the highly significant Ljung-Box statistics in Table 7. The most widely-used parametric framework for this type of data, see, e.g., Brownlees et al. (2010), is the multiplicative error model (MEM) originally proposed by Engle (2002). Accordingly, we decompose the t-th trading volume, $x_t^{(v)}$, as

$$x_t^{(v)} = \mu_t^{(v)} \varepsilon_t^{(v)}, \quad \varepsilon_t^{(v)} \sim \text{i.i.d. } \mathcal{D}(1)\,, \tag{29}$$

where $\mu_t^{(v)}$ denotes the conditional mean given the past information set $\mathcal{F}_{t-1}^{(v)}$ and is assumed to evolve according to the dynamics described in Appendix A. $\varepsilon_t^{(v)}$ is a disturbance following an unspecified distribution $\mathcal{D}(1)$ with positive support and $E\big[\varepsilon_t^{(v)}\big] = 1$. Assuming MEM-type dynamics would allow to apply gamma kernel estimators to trading volumes directly and estimate their unconditional density $f_x\big(x_t^{(v)}\big)$ consistently (see Bouezmarni and Rombouts, 2010). Our object of interest, the conditional density given the past information set $\mathcal{F}_{t-1}^{(v)}$, can be estimated semiparametrically in a straightforward way, as the MEM structure implies the

[1] For a detailed discussion of the treatment of zero observations in the context of financial high-frequency data, see Hautsch et al. (2010).

Table 7: Ljung-Box Statistics for Intraday Trading Volume and Realized Kernel Estimates
$Q(l)$: Ljung-Box statistic associated with l lags. The 5% (1%) critical values associated with lag lengths 20, 50 and 100 are 31.41 (37.57), 67.51 (76.15) and 124.34 (135.81). We consider deseasonalized nonzero 15-Second trading volumes of Citigroup and realized kernel (RK) estimates for JP Morgan.

	Volume	**RK**
$Q(20)$	10349.281	5045.309
$Q(50)$	19447.096	9834.944
$Q(100)$	31353.699	14012.591

basic relationship

$$f_x\big(x_t^{(v)}|\mathcal{F}_{t-1}^{(v)}\big) = f_\varepsilon\big(x_t^{(v)}/\mu_t^{(v)}\big)/\mu_t^{(v)}. \tag{30}$$

We consider a two-step approach. First, we estimate $\mu_t^{(v)}$ by exponential QML and generate residuals $\hat{\varepsilon}_t^{(v)} := x_t^{(v)}/\hat{\mu}_t^{(v)}$, which are consistent estimates of the i.i.d errors $\varepsilon_t^{(v)}$ (see, e.g., Drost and Werker, 2004). Second, we estimate $f_\varepsilon\big(x_t^{(v)}/\mu_t^{(v)}\big)$ nonparametrically employing gamma kernels. The consistency and parametric rate of convergence of the conditional mean estimates enable us to use the MEM residuals as inputs without affecting the asymptotics of the kernel density estimators.

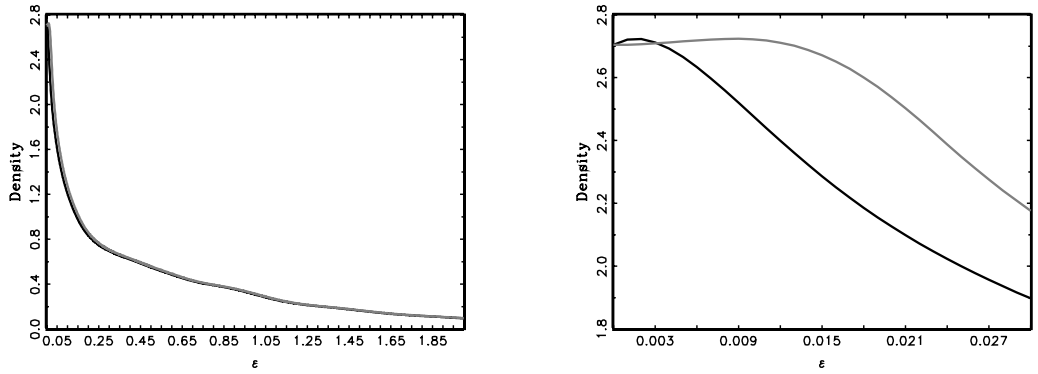
Nonparametric estimation of the error density requires the choice of the appropriate type of gamma kernel, i.e. basic or modified in the original and refined version (specification v_{I}). To ensure comparability and boundary regions of equal size, we consider the least-squares cross-validation (LSCV) bandwidth of the basic gamma kernel estimator in all cases. In particular, we use the bandwidth b^* that minimizes a nearly unbiased estimate of the integrated mean-squared error, i.e.

$$\begin{aligned} CV(b) = \frac{1}{n^2}\sum_i\sum_j \int_\tau^\infty K^\gamma_{x/b+1,b}\left(\hat{\varepsilon}_i^{(v)}\right) K^\gamma_{x/b+1,b}\left(\hat{\varepsilon}_j^{(v)}\right) dx \\ - \frac{2}{n\,(n-1)}\sum_i\sum_{j\neq i} K^\gamma_{x_i/b+1,b}\left(\hat{\varepsilon}_j^{(v)}\right), \end{aligned} \tag{31}$$

which yields the bandwidth $b^* = 0.0118$. See Hjort and Glad (1995) for details on (nearly) unbiased cross-validation. Further, we estimate the normalized density derivative $D\big(\varepsilon_t^{(v)}\big)$ for $\varepsilon_t^{(v)} \in \{0, b^*, 2b^*\}$ as in (15) based on the modified gamma kernel. The corresponding results in Table 8 show that two out of three estimates are considerably negative, which indicates a possible pole situation and suggests the use of the basic gamma kernel. Figure 9 displays estimates of the error density $f_\varepsilon\big(\varepsilon_t^{(v)}\big)$ based on the basic and, for comparison, modified gamma kernel for the boundary region and a larger part of the support. While for both density estimates, the probability mass is quite concentrated close to the origin, the basic gamma kernel, being the

Table 8: Estimates of Normalized Density Derivative for MEM Errors Based on Intraday Trading Volume and Realized Kernel Values
Estimates of the ratio $D\big(\varepsilon_t^{(m)}\big) := f'_\varepsilon\big(\varepsilon_t^{(m)}\big)/f_\varepsilon\big(\varepsilon_t^{(m)}\big)$, $m = v, rk$, based on the modified gamma kernel in the boundary region as in (15). $\varepsilon_t^{(m)}$ are errors from the MEM structure (29) fitted to deseasonalized nonzero 15-second trading volumes of Citigroup and realized kernel (RK) estimates for JP Morgan. b^* is LSCV bandwidth of the basic gamma kernel estimator: 0.0118 for trading volumes and 0.0206 for realized kernel estimates.

$\varepsilon_t^{(m)}$	**Volume**	**RK**
0	0.293	26.283
b^*	-10.100	59.915
$2b^*$	-14.399	53.235

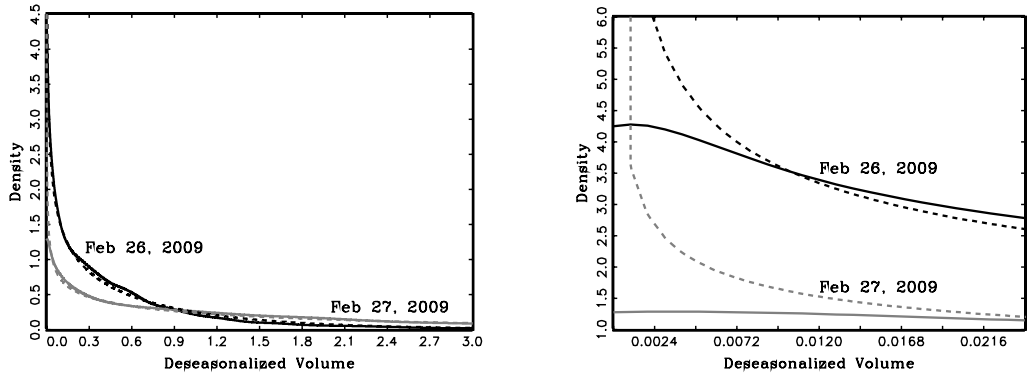


(a) Full Support (b) Boundary

Figure 9: Estimates of MEM Error Density for Intraday Trading Volumes (Citigroup)
Estimates of the density $f_\varepsilon\big(\varepsilon_t^{(v)}\big)$ from the MEM structure (29) fitted to deseasonalized nonzero 15-second trading volumes of Citigroup. Black solid line: basic gamma kernel. Grey solid line: modified gamma kernel. LSCV bandwidth of the basic gamma kernel, $b^* = 0.0118$, is used for both estimators.

method of choice, yields an estimate that lies clearly below the density implied by the modified kernel for the major part of the boundary region.

Finally, Figure 10 shows estimates of the conditional density of trading volumes for February 26 and 27, 2009, at 11am EST. On the latter day, Citigroup announced that the US treasury would be taking a major equity stake in the company, while the former day is included for comparison. As an alternative to the semiparametric approach, the plot also features the conditional density implied by maximum likelihood estimates of the MEM (29) assuming that the errors follow the widely-used gamma distribution (e.g. Engle and Gallo, 2006). The impact of the announcement on trading activity related to the Citigroup stock is clearly visible, as the conditional volume distribution for February 27 assigns considerably less weight to small transactions. The semiparametric density estimates and their parametric counterparts are quite close to each other in the interior of the support. The major difference occurs at the origin where the parametric densities exhibit a pole, which is not the case for the semiparametric estimates.



(a) Full Support (b) Boundary

Figure 10: (Semi-)Parametric Conditional Density of Intraday Volumes (Citigroup)
Conditional densities at t given past information $\mathcal{F}_{t-1}^{(v)}$ based on the MEM structure (29) and the relationship (30). Parametric estimates (dashed lines) are implied by a ML approach assuming gamma distributed errors $\varepsilon_t^{(v)}$. Semiparametric estimates (solid lines) rely on QML estimates of $\mu_t^{(v)}$ and nonparametric estimates of $f_\varepsilon(x_t^{(v)}/\mu_t^{(v)})$ using the basic gamma kernel. Conditional densities are estimated for 11am EST on February 26 (black lines) and February 27, 2009 (grey lines).

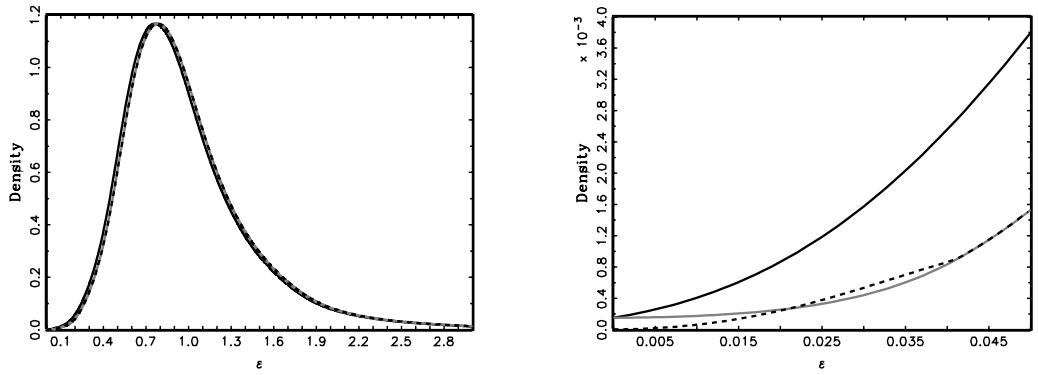
4.2 Forecasting Realized Volatility

Realized volatility measures computed from high-frequency data allow to construct more accurate estimates of the underlying lower frequency volatility (see, e.g., Andersen et al., 2010). We employ mid-quotes for JP Morgan from January 2006 to December 2009, which corresponds to 983 trading days, and clean the raw data as suggested in Barndorff-Nielsen et al. (2008b). The realized volatility for day t is simply defined as the sum of squared (mid-quote) returns $r_{i,t}$, $i = 1, \ldots, N_t$. Barndorff-Nielsen and Shephard (2002) show that, in the absence of noise and with the number of intraday returns approaching infinity, this basic estimator is consistent for the latent integrated volatility, which under regularity conditions provides an unbiased measure of the conditional variance of (daily) returns. In practice, observed prices are contaminated by microstructure effects causing an inconsistency of the basic realized volatility estimator (e.g. Hansen and Lunde, 2006). Hence, we consider the noise-robust realized kernel estimator, which was proposed by Barndorff-Nielsen et al. (2008a) and takes the form

$$x_t^{(rk)} := \gamma_0 + \sum_{h=1}^{H} k\left(\frac{h-1}{H}\right)(\gamma_h + \gamma_{-h}), \qquad \gamma_h := \sum_{i=1}^{n} r_{i,t} r_{i-h,t}, \tag{32}$$

where $k(\cdot)$ is the Parzen kernel and H the bandwidth.[2] Since (filtered) realized kernel estimates are used as inputs for kernel density estimators below, the two bandwidths involved have to

[2]The number of returns used for the computation of the realized kernel, n, is lower than the total number of observations N_t due to the so-called jittering procedure. See Barndorff-Nielsen et al. (2008a) for details.



(a) Full Support (b) Boundary

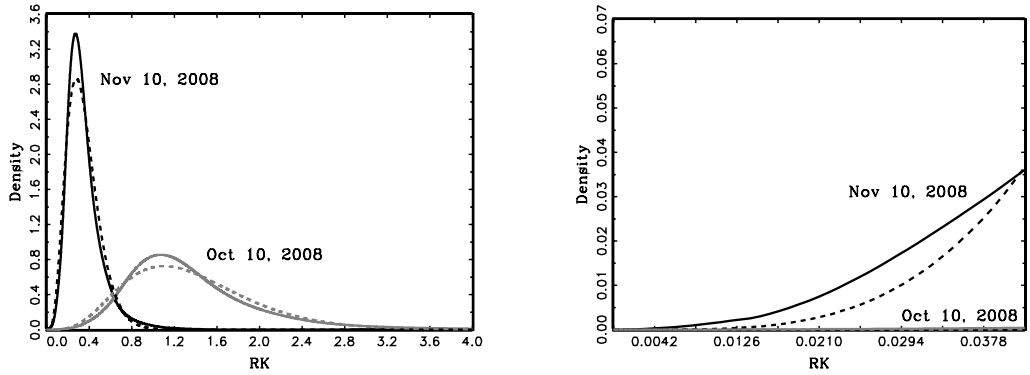
Figure 11: Estimates of MEM Error Density for Realized Kernel Estimates (JP Morgan)
Estimates of the density $f_\varepsilon\big(\varepsilon_t^{(rk)}\big)$ from the MEM structure (29) fitted to realized kernel estimates for JP Morgan. Black solid line: basic gamma kernel. Grey solid line: modified gamma kernel. Black dashed line: refined modified gamma kernel v_{I}. LSCV bandwidth of the basic gamma kernel, $b^* = 0.0206$, is used for all estimators.

be balanced in a way similar to Corradi et al. (2009), who propose nonparametric conditional density estimators for the integrated volatility. We ensure that their assumption A.1 is met by choosing H as in section 4.3 of Barndorff-Nielsen et al. (2008a).[3]

Table 7 shows that the realized kernel estimates exhibit a similar persistence as trading volumes, which we account for by following Engle and Gallo (2006) and imposing a flexible MEM structure. Hence, we model the realized kernel value for day t, $x_t^{(rk)}$, analogously to (29), where the assumptions for the errors $\varepsilon_t^{(rk)}$ remain the same, while a slightly different specification is chosen for the conditional mean $\mu_t^{(rk)}$ (see Appendix A). We compute semi-parametric estimates of the conditional density $f_x\big(x_t^{(rk)}|\mathcal{F}_{t-1}^{(rk)}\big)$ using the same approach as in Section 4.1, which in the given application, can be considered as a simple alternative to the fully nonparametric procedure proposed in Corradi et al. (2009). As Table 8 reports, the estimates of the normalized density derivative for the MEM errors are consistently positive indicating that the corresponding density should be estimated using a modified gamma kernel. Thus, we first determine the optimal value of the constant c for the refined specification v_{I} by minimizing the objective function (19). We compute the required pilot estimates of the unknown density and its first two derivatives as outlined in Section 2.3, which yields the threshold $c^* = 0.0863$.

Estimates of the MEM error density implied by all three types of gamma kernels considered are displayed in Figure 11 and indicate the following major results. First, as compared to the error density based on trading volumes in Figure 9, the mode of the distribution is further to the interior of the support. Second, the density exhibits a similar degree of right-skewness

[3]To estimate the so-called noise-to-signal ratio, we follow Barndorff-Nielsen et al. (2008b).



(a) Full Support (b) Boundary

Figure 12: (Semi-)Parametric Conditional Density of Realized Kernel Estimates (JP Morgan) Conditional densities at t given past information $\mathcal{F}_{t-1}^{(rk)}$ based on the MEM structure (29) and the relationship (30). Parametric estimates (dashed lines) are implied by a ML approach assuming gamma distributed errors $\varepsilon_t^{(rk)}$. Semiparametric estimates (solid lines) rely on QML estimates of $\mu_t^{(rk)}$ and nonparametric estimates of $f_\varepsilon\big(x_t^{(rk)}/\mu_t^{(rk)}\big)$ using the refined modified gamma kernel υ_{I}. Conditional densities are estimated for October 10 (grey lines) and November 10, 2008 (black lines). Realized kernel estimates are annualized.

as was reported for the unconditional distribution of realized volatilities by Andersen et al. (2001). Finally, the density estimate based on the refined modified kernel tends to zero when approaching the boundary, instead of taking a strictly positive value at $\varepsilon_t^{(rk)} = 0$. This effect is caused by the low value of the threshold c^*, which pushes the shape parameter $\rho_b^{\upsilon_{\mathrm{I}}}\big(\varepsilon_t^{(rk)}\big)$ below one when smoothing at the boundary (see eq. (16)). A distribution of stock return volatility with vanishing probability mass close to the boundary is in line with financial theory, since stocks are "risky" assets for which investors demand a volatility premium (e.g. Merton, 1973).

Figure 12 displays conditional density estimates of realized kernel values for two days during the financial crisis 2007 – 2008: October 10, 2008, when the DJIA index fell by 8% at the start of the trading day, and November 10, 2008, when a major restructuring of the AIG bailout plan was announced. The density estimates are based on our semiparametric procedure using the refined modified gamma kernel and the parametric approach from Section 4.1. Except for some discrepancies around the mode and in the boundary region, the parametric estimates roughly match the semiparametric ones indicating that the gamma distribution is a reasonable assumption for the MEM errors. With respect to dynamic changes, the conditional densities reflect the more unstable market environment on October 10, when the volatility distribution has its mode further away from the origin and is more dispersed. Further, as in case of the unconditional error density, the probability mass is vanishing close to the boundary for both days and estimators considered.

5 Conclusion

Gamma kernel estimators vary their shape according to the point of estimation along the support. For positive random variables, this location adaptiveness thus avoids the boundary bias of standard fixed kernel estimators while yielding strictly nonnegative density estimates by construction. We show for various density shapes that in finite samples the two original gamma kernel estimators outperform all boundary and boundary corrected fixed kernel type estimators at the boundary, in particular for settings with a large probability mass close to zero. For all other setups and in the interior of the support, their finite sample performance is comparable to the one of fixed type boundary kernels. Moreover, with asymptotic considerations and finite sample illustrations we find that for pole situations at zero, the two gamma kernel estimators differ substantially. In fact the standard type is superior to the generally used modified version in this case. We therefore suggest a simple criterion to check for such situations. For all other settings, we propose a refined modified version of the gamma kernel estimator, which further improves upon the performance of the modified gamma kernel. Our technique is complemented by a data-driven way for choosing the specification parameters in the new refined gamma kernel. In two application settings, we demonstrate that, in particular in high-frequency finance, the suggested methodology yields superior results of practical impact.

References

ANDERSEN, T. G., T. BOLLERSLEV, AND F. X. DIEBOLD (2010): "Parametric and Nonparametric Measurements of Volatility," in *Handbook of Financial Econometrics*, ed. by Y. Ait-Sahalia and L. Hansen, Amsterdam: North Holland, 67–137.

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2001): "The distribution of realized exchange rate volatility," *Journal of the American Statistical Association*, 96, 42–55.

BARNDORFF-NIELSEN, O., P. HANSEN, A. LUNDE, AND N. SHEPHARD (2008a): "Designing Realized Kernels to Measure the Ex-Post Variation of Equity Prices in the Presence of Noise," *Econometrica*, 76, 1481–1536.

——— (2008b): "Realised Kernels in Practice: Trades and Quotes," *Econometrics Journal*, 4, 1–32.

BARNDORFF-NIELSEN, O. AND N. SHEPHARD (2002): "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society, Ser. B.*, 64, 253–280.

BAUWENS, L. AND P. GIOT (2000): "The logarithmic ACD model: an application to the bid-ask quote process of three NYSE stocks," *Annales D'Economie et de Statistique*, 60, 117–149.

BOUEZMARNI, T. AND J. V. ROMBOUTS (2010): "Nonparametric density estimation for positive time series," *Computational Statistics & Data Analysis*, 54, 245 – 261.

BOUEZMARNI, T. AND O. SCAILLET (2005): "Consistency Of Asymmetric Kernel Density Estimators And Smoothed Histograms With Application To Income Data," *Econometric Theory*, 21, 390–412.

BROWNLEES, C. T., F. CIPOLLINI, AND G. M. GALLO (2010): "Intra-daily volume modeling and prediction for algorithmic trading," *Journal of Financial Econometrics*, 8, 1–30.

CHEN, S. (1999): "Beta kernel estimators for density functions," *Computational Statistics & Data Analysis*, 31, 131–145.

——— (2000): "Probability density function estimation using gamma kernels," *Annals of the Institute of Statistical Mathematics*, 52, 471–480.

CORRADI, V., W. DISTASO, AND N. R. SWANSON (2009): "Predictive density estimators for daily volatility based on the use of realized measures," *Journal of Econometrics*, 150, 119 – 138.

CORSI, F. (2009): "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics*, 174–196.

DROST, F. C. AND B. J. M. WERKER (2004): "Semiparametric duration models," *Journal of Business and Economic Statistics*, 22, 40–50.

ENGLE, R. F. (2002): "New frontiers for ARCH models," *Journal of Applied Econometrics*, 17, 425–446.

ENGLE, R. F. AND G. M. GALLO (2006): "A multiple indicators model for volatility using intra-daily data," *Journal of Econometrics*, 131, 3–27.

FERNANDES, M. AND J. GRAMMIG (2005): "Nonparametric specification tests for conditional duration models," *Journal of Econometrics*, 127, 35–68.

GASSER, T. AND H. MÜLLER (1979): "Kernel estimation of regression functions," in *Lecture Notes in Mathematics 757*, ed. by T. Gasser and M. Rosenblatt, Heidelberg: Springer, 23–68.

HAGMANN, M. AND O. SCAILLET (2007): "Local multiplicative bias correction for asymmetric kernel density estimators," *Journal of Econometrics*, 141, 213–249.

HANSEN, P. R. AND A. LUNDE (2006): "Realized Variance and Market Microstructure Noise," *Journal of Business and Economic Statistics*, 24, 127–161.

HAUTSCH, N. (2004): *Modelling Irregularly Spaced Financial Data: Theory and Practice of Dynamic Duration Models*, Berlin: Springer.

HAUTSCH, N., P. MALEC, AND M. SCHIENLE (2010): "Capturing the zero: a new class of zero-augmented distributions and multiplicative error processes," SFB 649 Discussion Paper 2010-055, Sonderforschungsbereich 649, Humboldt Universität zu Berlin.

HJORT, N. L. AND I. K. GLAD (1995): "Nonparametric density estimation with a parametric start," *The Annals of Statistics*, 23, 882–904.

JONES, M. (1993): "Simple boundary correction for kernel density estimation," *Statistics and Computing*, 3, 135–146.

JONES, M. C. AND P. J. FOSTER (1996): "A simple nonnegative boundary correction method for kernel density estimation," *Statistica Sinica*, 6, 1005–1013.

LANCASTER, T. (1997): *The Econometric Analysis of Transition Data*, Cambridge: Cambridge University Press.

MARRON, J. S. AND D. RUPPERT (1994): "Transformations to reduce boundary bias in kernel density estimation," *Journal of the Royal Statistical Society. Series B*, 56, 653–671.

MERTON, R. (1973): "An Intertemporal Capital Asset Pricing Model," *Econometrica*, 41, 867 – 888.

ROBINSON, P. AND M. HENRY (2003): "Higher-order kernel semiparametric M-estimation of long memory," *Journal of Econometrics*, 114, 1–27.

SCHUSTER, E. (1958): "Incorporating support constraints into nonparametric estimators of densities," *Communications in Statistics, Part A - Theory and Methods*, 14, 1123–1136.

ZHANG, S. (2010): "A note on the performance of gamma kernel estimators at the boundary," *Statistics and Probability Letters*, 80, 548–557.

A MEM Specifications

For trading volumes, we specify the conditional mean $\mu_t^{(v)}$ in (29) using the logarithmic MEM proposed by Bauwens and Giot (2000). The latter does not require parameter constraints to ensure the positivity of $\mu_t^{(v)}$ and implies

$$\ln \mu_t^{(v)} = \omega + \sum_{i=1}^{p} \alpha_i \ln x_{t-i}^{(v)} + \sum_{i=1}^{q} \beta_i \ln \mu_{t-i}^{(v)}, \tag{33}$$

where the lag structure is chosen according to the Schwartz information criterion (SIC).

In case of volatilities, we consider (33) with $p = 1$ but augmented by the lags of (logarithmic) weekly and monthly realized kernel estimates, which are defined as the averages

$$x_{t,w}^{(rk)} =: \frac{1}{5} \sum_{j=0}^{4} x_{t-j}^{(rk)} \quad \text{and} \quad x_{t,m}^{(rk)} =: \frac{1}{20} \sum_{j=0}^{19} x_{t-j}^{(rk)}. \tag{34}$$

This extension is motivated by the widely-used heterogenous autoregressive (HAR) model for realized volatilities proposed by Corsi (2009) and yields

$$\ln \mu_t^{(rk)} = \omega + \alpha^d \ln x_{t-1}^{(rk)} + \alpha^w \ln x_{t-1,w}^{(rk)} + \alpha^m \ln x_{t-1,m}^{(rk)} + \sum_{i=1}^{q} \beta_i \ln \mu_{t-i}^{(rk)}, \tag{35}$$

where q is determined using the SIC.

SFB 649 Discussion Paper Series 2012

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

001 "HMM in dynamic HAC models" by Wolfgang Karl Härdle, Ostap Okhrin and Weining Wang, January 2012.
002 "Dynamic Activity Analysis Model Based Win-Win Development Forecasting Under the Environmental Regulation in China" by Shiyi Chen and Wolfgang Karl Härdle, January 2012.
003 "A Donsker Theorem for Lévy Measures" by Richard Nickl and Markus Reiß, January 2012.
004 "Computational Statistics (Journal)" by Wolfgang Karl Härdle, Yuichi Mori and Jürgen Symanzik, January 2012.
005 "Implementing quotas in university admissions: An experimental analysis" by Sebastian Braun, Nadja Dwenger, Dorothea Kübler and Alexander Westkamp, January 2012.
006 "Quantile Regression in Risk Calibration" by Shih-Kang Chao, Wolfgang Karl Härdle and Weining Wang, January 2012.
007 "Total Work and Gender: Facts and Possible Explanations" by Michael Burda, Daniel S. Hamermesh and Philippe Weil, February 2012.
008 "Does Basel II Pillar 3 Risk Exposure Data help to Identify Risky Banks?" by Ralf Sabiwalsky, February 2012.
009 "Comparability Effects of Mandatory IFRS Adoption" by Stefano Cascino and Joachim Gassen, February 2012.
010 "Fair Value Reclassifications of Financial Assets during the Financial Crisis" by Jannis Bischof, Ulf Brüggemann and Holger Daske, February 2012.
011 "Intended and unintended consequences of mandatory IFRS adoption: A review of extant evidence and suggestions for future research" by Ulf Brüggemann, Jörg-Markus Hitz and Thorsten Sellhorn, February 2012.
012 "Confidence sets in nonparametric calibration of exponential Lévy models" by Jakob Söhl, February 2012.
013 "The Polarization of Employment in German Local Labor Markets" by Charlotte Senftleben and Hanna Wielandt, February 2012.
014 "On the Dark Side of the Market: Identifying and Analyzing Hidden Order Placements" by Nikolaus Hautsch and Ruihong Huang, February 2012.
015 "Existence and Uniqueness of Perturbation Solutions to DSGE Models" by Hong Lan and Alexander Meyer-Gohde, February 2012.
016 "Nonparametric adaptive estimation of linear functionals for low frequency observed Lévy processes" by Johanna Kappus, February 2012.
017 "Option calibration of exponential Lévy models: Implementation and empirical results" by Jakob Söhl und Mathias Trabs, February 2012.
018 "Managerial Overconfidence and Corporate Risk Management" by Tim R. Adam, Chitru S. Fernando and Evgenia Golubeva, February 2012.
019 "Why Do Firms Engage in Selective Hedging?" by Tim R. Adam, Chitru S. Fernando and Jesus M. Salas, February 2012.
020 "A Slab in the Face: Building Quality and Neighborhood Effects" by Rainer Schulz and Martin Wersing, February 2012.
021 "A Strategy Perspective on the Performance Relevance of the CFO" by Andreas Venus and Andreas Engelen, February 2012.
022 "Assessing the Anchoring of Inflation Expectations" by Till Strohsal and Lars Winkelmann, February 2012.

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2012

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

023 "Hidden Liquidity: Determinants and Impact" by Gökhan Cebiroglu and Ulrich Horst, March 2012.
024 "Bye Bye, G.I. - The Impact of the U.S. Military Drawdown on Local German Labor Markets" by Jan Peter aus dem Moore and Alexandra Spitz-Oener, March 2012.
025 "Is socially responsible investing just screening? Evidence from mutual funds" by Markus Hirschberger, Ralph E. Steuer, Sebastian Utz and Maximilian Wimmer, March 2012.
026 "Explaining regional unemployment differences in Germany: a spatial panel data analysis" by Franziska Lottmann, March 2012.
027 "Forecast based Pricing of Weather Derivatives" by Wolfgang Karl Härdle, Brenda López-Cabrera and Matthias Ritter, March 2012.
028 "Does umbrella branding really work? Investigating cross-category brand loyalty" by Nadja Silberhorn and Lutz Hildebrandt, April 2012.
029 "Statistical Modelling of Temperature Risk" by Zografia Anastasiadou, and Brenda López-Cabrera, April 2012.
030 "Support Vector Machines with Evolutionary Feature Selection for Default Prediction" by Wolfgang Karl Härdle, Dedy Dwi Prastyo and Christian Hafner, April 2012.
031 "Local Adaptive Multiplicative Error Models for High-Frequency Forecasts" by Wolfgang Karl Härdle, Nikolaus Hautsch and Andrija Mihoci, April 2012.
032 "Copula Dynamics in CDOs." by Barbara Choroś-Tomczyk, Wolfgang Karl Härdle and Ludger Overbeck, May 2012.
033 "Simultaneous Statistical Inference in Dynamic Factor Models" by Thorsten Dickhaus, May 2012.
034 "Realized Copula" by Matthias R. Fengler and Ostap Okhrin, Mai 2012.
035 "Correlated Trades and Herd Behavior in the Stock Market" by Simon Jurkatis, Stephanie Kremer and Dieter Nautz, May 2012
036 "Hierarchical Archimedean Copulae: The HAC Package" by Ostap Okhrin and Alexander Ristig, May 2012.
037 "Do Japanese Stock Prices Reflect Macro Fundamentals?" by Wenjuan Chen and Anton Velinov, May 2012.
038 "The Aging Investor: Insights from Neuroeconomics" by Peter N. C. Mohr and Hauke R. Heekeren, May 2012.
039 "Volatility of price indices for heterogeneous goods" by Fabian Y.R.P. Bocart and Christian M. Hafner, May 2012.
040 "Location, location, location: Extracting location value from house prices" by Jens Kolbe, Rainer Schulz, Martin Wersing and Axel Werwatz, May 2012.
041 "Multiple point hypothesis test problems and effective numbers of tests" by Thorsten Dickhaus and Jens Stange, June 2012
042 "Generated Covariates in Nonparametric Estimation: A Short Review." by Enno Mammen, Christoph Rothe, and Melanie Schienle, June 2012.
043 "The Signal of Volatility" by Till Strohsal and Enzo Weber, June 2012.
044 "Copula-Based Dynamic Conditional Correlation Multiplicative Error Processes" by Taras Bodnar and Nikolaus Hautsch, July 2012

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

SFB 649 Discussion Paper Series 2012

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

045 "Additive Models: Extensions and Related Models." by Enno Mammen, Byeong U. Park and Melanie Schienle, July 2012.
046 "A uniform central limit theorem and efficiency for deconvolution estimators" by Jakob Söhl and Mathias Trabs, July 2012
047 "Nonparametric Kernel Density Estimation Near the Boundary" by Peter Malec and Melanie Schienle, August 2012

SFB 649, Spandauer Straße 1, D-10178 Berlin
http://sfb649.wiwi.hu-berlin.de

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".