Local Multiplicative Bias Correction for Asymmetric Kernel Density Estimators

M. Hagmann[a] O. Scaillet[b]

First version: September 2003
This version: October 2005

Abstract

We consider semiparametric asymmetric kernel density estimators when the unknown density has support on $[0, \infty)$. We provide a unifying framework which relies on a local multiplicative bias correction, and contains asymmetric kernel versions of several semiparametric density estimators considered previously in the literature. This framework allows us to use popular parametric models in a nonparametric fashion and yields estimators which are robust to misspecification. We further develop a specification test to determine if a density belongs to a particular parametric family. The proposed estimators outperform rival non- and semiparametric estimators in finite samples and are easy to implement. We provide applications to loss data from a large Swiss health insurer and Brazilian income data.

Key words and phrases: semiparametric density estimation, asymmetric kernel, income distribution, loss distribution, health insurance, specification testing.

JEL Classification: C13, C14.

[a] HEC Lausanne and FAME, 33 Route de Chavannes, CH-1007 Lausanne.
E-mail: matthias.hagmann@unil.ch.

[b] HEC Genève and FAME, UNIMAIL, 102 Bd Carl Vogt, CH-1211 Genève.
E-mail: olivier.scaillet@hec.unige.ch.

1 Introduction

One of the major concerns of insurance companies is the study of a group of risks. For insurers, a good understanding of the size of a single claim is of most importance. Loss distributions describe the probability distribution of a payment to the insured. Traditional methods in the actuarial literature use parametric specifications to model single claims. The most popular specifications are the lognormal, Weibull and Pareto distributions. Hogg and Klugman (1984) and Klugman, Panjer and Willmot (1998) describe a set of continuous parametric distributions which can be used for modelling a single claim size. It is, however, unlikely that something as complex as the generating process of insurance claims can be described by just a few parameters. A wrong parametric specification may lead to an inadequate measurement of the risk contained in the insurance portfolio and consequently to a mispricing of insurance contracts. These remarks also apply to financial losses, portfolio selection and risk management procedures in a banking context.

In a totally different area of research, economists studying income distributions and income inequality use similar parametric models to estimate the distribution of income and its evolution over time. Popular models are the gamma, lognormal and Pareto distributions, see Cowell (1999). Whereas the lognormal distribution is thought to have the best overall shape, the Pareto is considered to be a more suitable distribution for individuals in the upper end of the income distribution. Although these densities may capture some stylised facts of income distributions, it is again unlikely that income distributions can be described by just a few parameters. The imposition of a wrong parametric model may lead to inconsistent estimates and misleading inference, as well as to disputable conclusions in inequality measurement for example.

A method which does not require the specification of a parametric model is nonparametric kernel smoothing. This method provides valid inference under a much broader class of structures than those imposed by parametric models. Unfortunately, this robustness comes at a price. The convergence rate of nonparametric estimators is slower than the parametric rate, and the bias induced by the smoothing procedure can be substantial even for moderate sample sizes. Since both income and losses are positive variables, the standard kernel estimator proposed by Rosenblatt (1956) has a boundary bias. This boundary bias is due to weight allocation by the fixed symmetric kernel outside the support of the distribution when smoothing close to the boundary is carried out. As a result, the mode close to the boundary typical for income and loss distributions is often missed. Additionally, standard kernel methods yield wiggly estimation in the tail of the distribution since the mitigation of the boundary bias leads to favour a small bandwidth which prevents pooling enough data. Precise tail measurement of loss distributions is however of particular importance to get appropriate risk measures when designing an efficient risk management system.

We propose a semiparametric estimation framework for the estimation of densities which have support on $[0, \infty)$. Our estimation procedure can deal with all problems of the standard kernel estimator mentioned previously, and this in a single way. Although the above parametric models may be inaccurate, they can be used in a nonparametric fashion to help to decrease the bias induced by nonparametric smoothing. If the parametric model is accurate, the performance of our semiparametric estimator can be close to pure parametric estimation. Following Hjort and Glad (1995) (H&G), we start with a parametric estimator of the unknown density (economic theory may help in providing the parametric start), and then correct nonparametrically for possible misspecification. To decrease the bias even further, we give some local parametric guidance to this nonparametric correction in the spirit of Hjort and Jones (1996) (H&J). This is achieved by employing either local polynomial or log polynomial models, where the latter method results always in nonnegative density estimates. We call this approach *local multiplicative bias correction*, or LMBC to be short.

We emphasize that appropriate boundary bias correction is more important in a semiparametric than a pure nonparametric setting. This is because the bias reduction achieved by semiparametric techniques allows us to increase the bandwidth and thus to pool more data. This, however, increases the boundary region where the symmetric kernel allocates weight to the negative part of the real line. This motivates us to develop LMBC in an asymmetric kernel framework which eliminates the boundary issue completely[1]. Asymmetric kernel estimators were recently proposed by Chen (2000) as a convenient way to solve the boundary bias problem. The symmetric kernel is replaced by an asymmetric gamma kernel which matches the support of the unknown density[2]. As an alternative to the gamma kernel, Scaillet (2004) introduced kernels based on the inverse Gaussian and reciprocal inverse Gaussian density. All of these kernel functions have flexible form, are located on the nonnegative real line and produce nonnegative density estimates. Also, they change the amount of smoothing in a natural way as one moves away from the boundary. This is particularly attractive when estimating densities which have areas sparse in data because more data points can be pooled. As pointed out by Cowell (1999) "Empirical income distributions typically have long tails with sparse data". The same holds true for empirical loss distributions and we therefore think that these kernels are very well suited in this context. The variance advantage of the asymmetric kernel comes, however, at the cost of a slightly increased bias as one moves away from the boundary compared to symmetric kernels, which highlights the importance of effective bias reduction techniques in the tails. In a comprehensive

[1] The theoretical results derived in this paper show that the form of the bias reduction achieved through LMBC is analogous in the symmetric and asymmetric kernel case, although the mathematics and the strategy of the proof yielding these results are totally different. Obviously we cannot exploit symmetry in the derivation of the results for asymmetric kernels.

[2] Other remedies include the use of particular boundary kernels or bandwidths, see e.g. Rice (1984), Schuster (1985), Jones (1993), Müller (1991) and Jones and Foster (1996).

simulation study, Scaillet (2004) obtains attractive finite sample performance of these asymmetric kernel estimators. He also reports that boundary kernel estimators lead too often to negative density estimates without outperforming asymmetric kernel density estimators. Chen (2000) reports superior performance of the gamma kernel estimator compared to other remedies proposed in the literature as the local linear estimator of Jones (1993). A particular advantage of the gamma kernel estimator is its consistency when the true density is unbounded at $x = 0$, which is important for the estimation of highly skewed loss and income distributions. This is shown in Bouezmarni and Scaillet (2005) who also establish uniform and L_1 convergence results for asymmetric kernel density estimators. Furthermore they report nice finite sample performance of the asymmetric kernel density estimator w.r.t. the L_1 norm.

Our simulation study underlines the importance of efficient boundary correction in a semiparametric framework. We find that LMBC in connection with asymmetric kernels yields excellent results. These estimators perform better in a mean integrated squared error (MISE) sense than pure nonparametric estimators. If the parametric information provided is accurate, we find that a MISE reduction of 50-80% can be reasonably expected. Even if the misspecification considered is large, our LMBC estimator still achieves a MISE reduction of around 25%. Asymmetric kernel based LMBC estimators outperform their symmetric rivals: first because they eliminate the boundary bias issue more successfully (allowing a larger bandwidth), and second because they have an intrinsic advantage in the tails of the density. Furthermore, they are often easier to implement.

As a by-product of our approach, we propose a new attractive semiparametric specification test to determine whether a particular unknown density belongs to a parametric class of densities. The test is very simple to implement and should prove useful in empirical applications. We also explain how this statistic can be used to determine which density can be felt as a suitable parametric start.

Although we concentrate in the empirical part on loss and income distributions, similar issues as discussed above are also important in the finance literature. Aït-Sahalia (1996a) develops an estimation procedure for diffusion models of the short term interest rate. Based on Bickel and Rosenblatt's (1973) work on density matching, Aït-Sahalia (1996b) also proposes a way to test various parametric specifications for diffusion models of the short rate. In his estimation and specification framework, the nonparametric estimation of the stationary distribution of the interest rate process plays a key role. The intertemporal general equilibrium asset pricing model of Cox, Ingersoll and Ross (1985) implies that this distribution follows a gamma probability law. Although this interest rate model may again be overly restrictive, it gives some economic guidance about the likely form of the stationary distribution of the short rate. This information can be incorporated in a semiparametric estimator like ours. Furthermore, our results are potentially important for estimation and specification testing of the baseline hazard function in financial duration analysis. In this literature parametric models like the Burr and generalized gamma distribution are popular specifications for the baseline hazard. We refer to Engle

(2000) for an overview of autoregressive conditional duration models (ACD), Tyurin (2003) for a recent application of the competing risk model to the foreign exchange market, and Fernandes and Grammig (2005) for exploitation of asymmetric kernels in financial duration analysis. Clearly the standard kernel estimator is again not appropriate in these contexts, since it does not take into account that the underlying variables, interest rates and durations, are nonnegative.

The outline of the paper is as follows. In Section 2 we introduce our semiparametric estimation framework and relate it to the relevant associated literature. This unified framework embeds semiparametric density estimators developed by H&G, H&J, and Loader (1996), and allows us to fully clarify the interdependence between these approaches. Section 3 recalls asymmetric kernel methods. Section 4 contains the main contribution of the paper, namely the extension of the LMBC framework to the asymmetric kernel case. We develop several examples, which show that the estimation procedure is user friendly and remarkably simple to implement in most cases[3]. The procedure is therefore appealing for applied work. We also discuss bandwidth choice and model diagnostic tests. In Section 5 we compare the performance of our estimators through an extensive simulation study. To the best of our knowledge, it is the first time that the various semiparametric approaches mentioned above are compared on a finite sample basis. In Section 6 we provide two empirical applications: the first one to loss data from a large Swiss health insurer, the second one to Brazilian income data. Section 7 contains some concluding remarks. An appendix gathers the proofs and technicalities related to the properties of the various estimators considered in the text.

2 Local multiplicative bias correction

In nonparametric regression, local polynomial fitting is a very popular approach, e.g. Fan and Gijbels (1996) and the references therein. Gozalo and Linton (2000) are the first to consider local fitting of a general functional using a least squares criterion, a normal error distribution version of a local likelihood estimator. For density estimation, the local likelihood approach was independently developed at the same time by H&J and Loader (1996). For a recent extension of the approach with application to Value-at-Risk (VaR) in risk management, see Gourieroux and Jasiak (2001). Whereas Loader (1996) concentrates on local polynomial fitting to the logarithm of the density, H&J allow for general local functionals like Gozalo and Linton (2000) in regression estimation.

In this section, we briefly introduce an estimation framework based on familiar symmetric kernel methods, which contains as special cases the local likelihood and multiplicative bias correction approach as described in H&J, Loader (1996), and H&G, respectively. This framework will allow us to derive the

[3]MATLAB code is available on request.

properties of these methods using asymmetric kernels in a single step, instead of treating the methods separately. Furthermore it allows us to shed some light on the intertwinning of these methods.

Let $X_1, ..., X_n$ be a random sample from a probability distribution F with an unknown density function $f(x)$ where x has support on $[0, \infty)$. We propose the following local model as a basis to estimate the true density function $f(x)$:

$$m(x, \theta_1, \theta_2(x)) = f(x, \theta_1) r(x, \theta_2(x)). \tag{2.1}$$

The first part of the local model m consists of $f(x, \theta_1)$, which is a parametric family of densities indexed by the global parameter $\theta_1 = (\theta_{11}, ..., \theta_{1p}) \in \Theta_1 \in \mathbf{R}^p$. This term serves as a global parametric start and is assumed to provide a meaningful but potentially inaccurate description of the true density $f(x)$. The second part of m denoted by $r(x, \theta_2(x))$ with $\theta_2(x) = (\theta_{21}(x), ..., \theta_{2q}(x)) \in \Theta_2 \in \mathbf{R}^q$ serves as the local parametric model for the unknown function $r(x) = f(x)/f(x, \theta_1)$. The role of this 'correction function' is, as the name says, to correct the potentially misspecified global start density $f(x, \theta_1)$ towards the true density $f(x)$. H&J briefly discuss this local model as a particular example in their paper, whereas we use it to provide a general framework for several semiparametric estimators proposed in the literature. We call this local multiplicative bias correction (LMBC) since only the multiplicative correction factor is modelled locally. Note that the correction function $r(x)$ is uniformly equal to one if the parametric start is well specified. Hence when the degree of misspecification is not too severe it is intuitively more natural to model the correction factor locally than the unknown density itself.

The procedure is as follows: first, estimate the parameter θ_1, which does not depend on x, by maximum likelihood. It is well known that when the parametric model $f(x, \theta_1)$ is misspecified, θ_1 converges in probability to the pseudo true value θ_1^0 which minimizes the Kullback-Leibler distance of $f(x, \theta_1)$ from the true $f(x)$, see e.g. White (1982) and Gourieroux, Monfort and Trognon (1984). Second, choose $\theta_2(x)$ such that

$$\frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_h(X_i - x) v(x, X_i, \theta_2) - \int \mathcal{K}_h(t - x) v(x, t, \theta_2) f\left(t, \hat{\theta}_1\right) r(t, \theta_2)\, dt = 0 \tag{2.2}$$

holds, where $\mathcal{K}_h(z) = (1/h)\mathcal{K}(z/h)$ is a symmetric kernel function, h is the bandwidth parameter and $v(x, t, \theta_2)$ is a $q \times 1$ vector of weighting functions. We omit for notational simplicity the possible dependence of the weighting function on θ_1. If we choose the score $\partial \log m(x, \theta_1, \theta_2(x)) / \partial\theta_2$ as the weighting function, then Equation (2.2) is just the first order condition of the local likelihood function given in H&J. In general, the form of the weighting function is driven by the tractability of the implied resulting estimator and is discussed in more detail as we proceed in the paper. The local multiplicatively bias corrected density estimator is

$$\hat{f}(x) = f\left(x, \hat{\theta}_1\right) r\left(x, \hat{\theta}_2(x)\right). \tag{2.3}$$

From the theoretical results concerning bias and variance of the local likelihood estimator given in H&J, it immediately follows that this estimator has the same variance as the standard kernel density estimator introduced by Rosenblatt (1956). The bias is however different. Compared to H&J, we prefer to state the bias in terms of the correction factor. This is more intuitive and simplifies the comparison between different estimation strategies. To ease notation, we write $f_0(x) = f\left(x, \theta_1^0\right)$ and $r_0(x) = f(x)/f\left(x, \theta_1^0\right)$. When θ_2 is one dimensional, the bias is[4]

$$\begin{aligned} \text{Bias}\left(\hat{f}(x)\right) &= \sigma_{\mathcal{K}}^2 h^2 \left(\frac{1}{2} f_0(x) \left[r_0^{(2)}(x) - r^{(2)}(x, \theta_2^0)\right]\right. \\ &\left. + \{\frac{v_0^{(1)}(x)}{v_0(x)} f_0(x) + f_0^{(1)}(x)\} \left[r_0^{(1)}(x) - r^{(1)}(x, \theta_2^0)\right]\right), \end{aligned} \tag{2.4}$$

where $\sigma_{\mathcal{K}}^2 = \int z^2 \mathcal{K}(z)\, dz$ and $v_0(x)$ denotes $v(x, x, \theta_2^0)$. The magnitude of this bias term depends on how well the correction function can be approximated locally by a suitable parametric model. This is so if $r(x)$ is smooth, or equivalently, if the global parametric start is close to the true density. In the single parameter case, the bias also depends on the weighting function and on the distance between the slopes of the correction function and its local model. If $\dim(\theta_2) \geq 2$, the bias is free of this term and only the first term in the brackets appears. For further details we refer to H&J.

Direct local modelling of the density can be obtained by choosing the parametric start density as an improper uniform distribution. W.l.o.g. set $f_0(x)$ to one. Then the only source of bias reduction is provided by the local model $r(x, \theta)$. The bias is $\frac{1}{2}\sigma_{\mathcal{K}}^2 h^2 \left[r_0^{(2)}(x) - r^{(2)}(x, \theta_2^0)\right]$ as in H&J.

The multiplicatively corrected kernel estimator of H&G emerges from choosing the weighting function as $1/f\left(x, \hat{\theta}_1\right)$, and choosing the local model as a constant. From Equations (2.2) and (2.3) it follows that the estimator is

$$\hat{f}(x) = \frac{f\left(x, \hat{\theta}_1\right)}{\int \mathcal{K}_h(t-x)dt}\frac{1}{n}\sum_{i=1}^{n} \mathcal{K}_h(X_i - x)\frac{1}{f\left(X_i, \hat{\theta}_1\right)}. \tag{2.5}$$

From (2.4), the bias is $(1/2)\, \sigma_{\mathcal{K}}^2 h^2 f_0(x)\, r^{(2)}(x)$, which is the bias obtained by H&G and does not depend on the chosen weighting function. We remark that this is the only possible choice of weighting function which sets the second bracket term in Equation (2.4) to zero. Assuming that $\mathcal{K}$ has support $[-1, 1]$[5], the term in the denominator of Equation (2.5) integrates to one if x lies in the interior, meaning that $x/h \to \kappa > 1$. However, close to the boundary where $0 \leq \kappa < 1$, this integral term normalizes the density estimate and therefore adjusts for the undesirable weight allocation of the symmetric kernel

[4]Since $\hat{\theta}_1$ exhibits $\sqrt{n}$-convergence which is faster than the nonparametric rate, the additional variability introduced through the first step estimation of θ_1 does not influence the bias and variance of $\hat{f}(x)$ up to negligible higher order terms.

[5]This setup can easily be extended to infinite support kernels. However, finite support is a standard assumption, delineating boundary and interior regions.

outside the support of the density. This adjustment is not optimal and boundary bias is still of the undesirable order $O(h)$. Like in nonparametric regression, see e.g. Fan and Gijbels (1992), one of the possible boundary bias correction methods which achieves an $O(h^2)$ order is the popular local linear estimator, see Jones (1993) for the density case. To obtain a local linear H&G version we propose to choose the local model as $r(t,\theta_2) = \theta_{21} + \theta_{22}(t-x)$ and the weight functions as $1/f(t,\theta_1)$ and $(t-x)/f(t,\theta_1)$. The resulting estimator is equivalent to the H&G estimator in the interior of the density. Close to the boundary it provides however again a correction due to weight allocation of the symmetric kernel to the negative part of the real line. Define $\alpha_j(\kappa) = \int_{-1}^{\kappa} \mathcal{K}(u)u^j du$, then the local linear estimator in the boundary region is

$$\hat{f}(x) = f\left(x, \hat{\theta}_1\right) \frac{\hat{r}(x) - [\alpha_1(\kappa)/h\alpha_2(\kappa)]\,\hat{g}(x)}{\left(\alpha_0(\kappa) - \alpha_1(\kappa)^2/\alpha_2(\kappa)\right)}, \tag{2.6}$$

where $\hat{g}(x)$ is the sample average of $\mathcal{K}_h(X_i - x)\left[(X_i - x)/f\left(X_i, \hat{\theta}_1\right)\right]$.

After presenting this unifying framework for previously proposed estimators, we now turn to an asymmetric kernel version of the above approach. Since the support of these kernels matches the support of the density under consideration, no boundary correction of the type presented above is necessary by construction. We first briefly review asymmetric kernel estimators for densities defined on the nonnegative real line introduced by Chen (2000) and Scaillet (2004). In Section 4 we will treat the LMBC case.

3 Asymmetric kernel methods

The asymmetric kernel density estimator is

$$\hat{f}_b(x) = \frac{1}{n}\sum_{i=1}^{n} K(X_i; x, b), \tag{3.1}$$

where b is a smoothing parameter satisfying $b \to 0$ and $bn \to \infty$ as $n \to \infty$. The asymmetric weighting function K is either a gamma density K_G with parameters $(x/b+1, b)$ as proposed by Chen (2000), an inverse Gaussian density K_{IG} with parameters $(x, 1/b)$, or a reciprocal inverse Gaussian density K_{RIG} with parameters $(1/(x-b), 1/b)$ as proposed by Scaillet (2004). These kernel densities are

$$\begin{aligned}
K_G(t; \frac{x}{b}+1, b) &= \frac{t^{x/b}\exp(-t/b)}{\Gamma(x/b+1)\,b^{x/b+1}}, \\
K_{IG}(t; x, \frac{1}{b}) &= \frac{1}{\sqrt{2\pi b t^3}}\exp\left(-\frac{1}{2bx}\left(\frac{t}{x} - 2 + \frac{x}{t}\right)\right), \\
K_{RIG}(t; \frac{1}{x-b}, \frac{1}{b}) &= \frac{1}{\sqrt{2\pi b t}}\exp\left(-\frac{x-b}{2b}\left(\frac{t}{x-b} - 2 + \frac{x-b}{t}\right)\right).
\end{aligned}$$

Note that these asymmetric kernels do not take the form $\omega(x-t,b)$ where ω is an asymmetric function (instead of a symmetric one), and thus do not belong to the class of asymmetric kernels studied by Abadir and Lawford (2004). Figure 1 displays the gamma kernel for some selected x-values. All asymmetric kernels share the property that the shape of the kernel changes according to the value of x. This varying kernel shape changes the amount of smoothing applied by the asymmetric kernel since the variance of, for instance, $K_G(t;\frac{x}{b}+1,b)$ is $xb+b^2$, which is increasing in x as we move away from the boundary. This is also reflected in the bias and variance expressions, which we give here for the gamma kernel estimator, and on which we concentrate below:

$$\begin{aligned} \text{Bias}\left(\hat{f}_b^G(x)\right) &= \left\{f^{(1)}(x)+\frac{1}{2}xf^{(2)}(x)\right\}b+o(b), \\ \text{Var}\left(\hat{f}_b^G(x)\right) &= \begin{cases} \frac{1}{2\sqrt{\pi}}n^{-1}b^{-1/2}x^{-1/2}f(x) & \text{if } x/b\to\infty, \\ \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)}n^{-1}b^{-1}f(x) & \text{if } x/b\to\kappa. \end{cases} \end{aligned} \tag{3.2}$$

This estimator is not subject to boundary bias, but involves the first derivative of the unknown density. This is because x is not the mean of the gamma kernel $K_G(t;\frac{x}{b}+1,b)$, rather its mode. This is different for the inverse Gaussian and reciprocal inverse Gaussian kernel estimators, whose biases only involve the second order derivative of the unknown density. To circumvent the first derivative in the bias expression, Chen (2000) also proposes a second gamma kernel which is, as Scaillet (2004) reports, similar in shape as the reciprocal inverse Gaussian kernel but has a slightly inferior finite sample performance.

Compared to other boundary correction techniques, the bias of gamma kernel estimators may be larger as x increases, this is however compensated by a reduced variance. In the interior where $x/b\to\infty$, it is apparent from (3.2) that the variance of the gamma kernel estimator decreases as x gets larger. This is in contrast to symmetric kernel estimators whose variance coefficients remain constant outside the boundary area. Also asymmetric kernels have a larger effective sample size than kernels with compact support. This is desirable for estimating densities with sparse areas as more data points can be pooled.

In the following we address the question of semiparametric bias reduction techniques for asymmetric kernel methods. This is important since as just reported, the bias may be larger than for standard symmetric kernel methods. Effective bias reduction techniques combined with a variance decreasing as we move away from the boundary is giving us hope for promising performance of our estimators for the estimation of loss and income distributions. This will be confirmed later in the paper.

4 Local multiplicative bias correction with asymmetric kernels

Apart from being an attractive semiparametric bias reduction framework, LMBC allows us to implement a popular boundary bias reduction by choosing a local linear model for the density or correction factor. This boundary bias reduction is not necessary *per se* in the asymmetric kernel framework since no weight is allocated outside the support of the unknown density. The effect of LMBC in an asymmetric framework is just to reduce the potentially larger bias for asymmetric kernel techniques. In addition, LMBC in connection with asymmetric kernels allows for the construction of user friendly estimators despite the apparent complexity of the approach. Estimators based on symmetric kernels, like e.g. the Gaussian kernel, often require numerical integration and optimization procedures. Bolancé et. al. (2003) mention that nonparametric methods for loss distribution estimation are seldom applied in practice because of implementation difficulties. Candidate estimators must be easy to implement to have a chance of being applied in the non-academic world. Numerical tractability is also a key advantage when resampling methods such as the bootstrap are used for inferential purposes.

We now extend the LMBC approach to the asymmetric kernel case, compute bias and variance of the estimator and discuss the choice of bandwidth. We also consider the special cases of H&J, Loader (1996), and H&G, and show how these methods can be applied to the estimation of income and loss distributions.

4.1 Definition of the estimator

We follow the notation introduced in Section 2. The estimator is $\tilde{f}_b(x) = f\left(x, \hat{\theta}_1\right) r\left(x, \hat{\theta}_2(x)\right)$, where $\hat{\theta}_1$ is the global maximum likelihood estimator which does not depend on x, and $\hat{\theta}_2(x)$ is chosen by maximizing the local likelihood function

$$\begin{aligned}
\log L_n\left(x, \hat{\theta}_1, \theta_2\right) &= \int_0^\infty K(t; x, b)\left\{\log m\left(t, \hat{\theta}_1, \theta_2\right) dF_n(t) - m\left(t, \hat{\theta}_1, \theta_2\right) dt\right\} \\
&= \frac{1}{n}\sum_{i=1}^n K(X_i; x, b)\log m(X_i, \hat{\theta}_1, \theta_2) - \int_0^\infty K(t; x, b)\, m(t, \hat{\theta}_1, \theta_2) dt, \qquad (4.1)
\end{aligned}$$

with F_n denoting the empirical distribution function. This criterion function is equivalent to the one of H&J. However, the symmetric kernel is replaced by an asymmetric kernel, whose support matches the support of the density we wish to estimate. For notational simplicity we omit the local dependency of θ_2 on x. The first term in (4.1) is the standard log-likelihood function weighted by an asymmetric kernel function. Maximizing this term alone would lead to inconsistent results because the expectation of its score is not equal to zero at the true parameter value θ_2^0. The second term guarantees that this is the case; we refer to H&J. From (4.1), when b is very large, $K(t;x,b)$ is independent of t and the above expression is a constant times the ordinary, normalized log-likelihood function. The maximization of the local likelihood then becomes the same as ordinary likelihood maximization. But if b is small, the maximization of $\log L_n\left(x,\hat{\theta}_1,\theta_2\right)$ will provide the best local estimator of $f(x)$. This follows since under some regularity conditions,

$$\log L_n\left(x,\hat{\theta}_1,\theta_2\right)\xrightarrow{p}\pi(x,\theta_1^0,\theta_2)=\int_0^\infty K(t;x,b)\{f(t)\log m(t,\theta_1^0,\theta_2)-m(t,\theta_1^0,\theta_2)\}dt,$$

as n grows. Hence $\hat{\theta}_2$, the maximizer of (4.1), aims at the parameter value $\theta_2^0(x)$ that maximizes $\pi(x,\theta_1^0,\theta_2)$ when the above convergence is uniform over the parameter space. See for example Linton and Pakes (2001). The solution to the above problem minimizes the following distance measure which is a localized version of the Kullback-Leibler distance of $f(x)$ from $m(x,\theta_1^0,\theta_2)$:

$$d[f,m(\cdot,\theta_1^0,\theta_2)]=\int_0^\infty K(t;x,b)\left[f(t)\log\frac{f(t)}{m(t,\theta_1^0,\theta_2)}-\{f(t)-m(t,\theta_1^0,\theta_2)\}\right]dt.$$

This shows that $\hat{\theta}_2$ aims at the best local parametric approximant to the true f. The estimator depends on the chosen smoothing parameter. For further details and justifications of the local likelihood approach, see H&J, Loader (1996), and the references therein.

The estimator $\hat{\theta}_2$ for general weight functions $v(x,t,\theta_2)$ is defined to be the solution to

$$V_n\left(x,\hat{\theta}_1,\theta_2\right)=\frac{1}{n}\sum_{i=1}^n K(X_i;x,b)\,v(x,X_i,\theta_2)-\int_0^\infty K(t;x,b)\,v(x,t,\theta_2)m(t,\hat{\theta}_1,\theta_2)dt=0. \tag{4.2}$$

This is identical to the first order condition of (4.1) in the case where v is chosen as the score $u(t,\theta_2)=(\partial/\partial\theta_2)\log r(t,\theta_2)$. For identification reasons, assume that

$$V_n\left(x,\hat{\theta}_1,\theta_2\right)\xrightarrow{p}V\left(x,\theta_1^0,\theta_2\right)=\int K(t;x,b)\,v(x,t,\theta_2)f_0(x)\{r_0(t)-r(t,\theta_2)\}dt=0$$

has a unique solution at $\theta_2=\theta_2^0$. This requires that the q weight functions are functionally independent, and that the correction function $r_0(t)$ is within reach of the parametric model $r(t,\theta_2)$ as θ_2 varies. This is like M-estimation in a possibly misspecified case, since the true correction function does not have to belong to the parametric family $r(t,\theta)$.

4.2 Large sample properties

We now develop the bias and variance of the LMBC estimator. The derivations of all results presented here are given in the appendix. When not stated otherwise, we will focus on results for the gamma kernel developed in Chen (2000) since other kernel choices can be handled in a similar fashion.

If we locally fit one parameter, the bias of the LMBC estimator is

$$\begin{aligned}
\operatorname{Bias}\left(\tilde{f}_b^G(x)\right) &= f_0(x)\left[\{r_0^{(1)}(x)-r^{(1)}(x,\theta_2^0)\}+\frac{1}{2}x\{r_0^{(2)}(x)-r^{(2)}(x,\theta_2^0)\}\right]b \\
&\quad+\left(\frac{v_0^{(1)}(x)}{v_0(x)}f_0(x)+f_0^{(1)}(x)\right)x\{r_0^{(1)}(x)-r^{(1)}(x,\theta_2^0)\}b \\
&\quad+o(b)+O\left(\frac{1}{nb^{1/2}}\right),
\end{aligned} \tag{4.3}$$

where we use the same notation as in the second section of this paper. H&J also note that the first derivative will vanish automatically from the bias expression if the number q of locally fitted parameters is larger than two. Equation (4.3) will then hold for any component $v_{j,0}(x)$ of the weighting function, which can only be true if $\left[r_0^{(1)}(x)-r^{(1)}(x,\theta_2^0)\right]=o(1)$ as $b\to 0$. This is not generally the case with one parameter. This means that for $q\geq 2$, the bias reduces to

$$\operatorname{Bias}\left(\tilde{f}_b^G(x)\right)=\frac{1}{2}f_0(x)\{r_0^{(2)}(x)-r^{(2)}(x,\theta_2^0)\}xb+o(b)+O\left(\frac{1}{nb^{1/2}}\right). \tag{4.4}$$

H&J show that for $q\geq 3$ one can argue that $\{r_0^{(2)}(x)-r^{(2)}(x,\theta_2^0)\}$ is also $o(1)$ and third and fourth order derivatives appear in the bias term. This property also holds for asymmetric kernels. We also note that Chen (2002) introduced a local linear regression estimator based on the first gamma kernel, which has the first derivative removed from the bias expression compared to the standard local constant regression smoother.

There are several worthwhile remarks. First, we obtain the same result as in the symmetric kernel case albeit relying on different proof techniques adapted to our asymmetric framework. Comparing Equations (3.2) and (4.4), the first derivative vanished and the second derivative in the bias of the asymmetric kernel estimator is replaced by $f_0(x)\left[r_0^{(2)}(x)-r^{(2)}(x,\theta_2^0)\right]$. So this estimator performs better than pure asymmetric kernel methods if the latter expression is smaller than the former in absolute values. This is the case if the unknown density exhibits high local curvature or if the parametric start is close to the true density since then $r_0^{(2)}(x)$ is small. Additionally, the local model for the correction factor can make this term even smaller if it can locally capture the curvature of the correction factor. If the model is correct, the local likelihood estimator is unbiased up to the order considered.

The variance of the asymmetric LMBC estimator in the one parameter case is

$$\operatorname{Var}\left(\tilde{f}_b^G(x)\right)=\begin{cases}\frac{1}{2\sqrt{\pi}}n^{-1}b^{-1/2}x^{-1/2}f(x) & \text{if } x/b\to\infty, \\ \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)}n^{-1}b^{-1}f(x) & \text{if } x/b\to\kappa,\end{cases} \tag{4.5}$$

where κ is a positive constant. We therefore obtain exactly the same result for the variance as for the pure nonparametric gamma kernel estimator. Also the variance of the LMBC estimator does not depend on the chosen weighting functions. We therefore have some flexibility to select them to obtain estimators which are tractable to implement.

The variance of the LMBC estimator in the multiple parameter case for a general asymmetric kernel is

$$\operatorname{Var}\left(\tilde{f}(x)\right)=\frac{f(x)}{n}e_1^t\tau\left(K\right)e_1-\frac{f\left(x\right)^2}{n}+O\left(b/n\right), \tag{4.6}$$

where, using some simplified notation, $\tau\left(K\right)$ is given by

$$\left(\int_0^\infty K\left(t\right)V_tV_t^{'}dt\right)^{-1}\left(\int_0^\infty K\left(t\right)^2V_tV_t^{'}dt\right)\left(\int_0^\infty K\left(t\right)V_tV_t^{'}dt\right)^{-1}$$

and V_t is $q\times 1$ vector containing in the j^{th} position the elements $(t-x)^{j-1}$ for $j=1,...,q$. This expression depends on the kernel being used. Independent of the kernel used, the variance of the LMBC estimator in the two parameter case is the same as in the single parameter case. We collect results for all the asymmetric kernel estimators in the following proposition.

Proposition 1 *The bias expressions of the asymmetric LMBC estimator in the cases where K is the gamma, inverse Gaussian and reciprocal inverse Gaussian kernel are given for $q\geq 2$ by:*

$$\begin{aligned}
Bias\left(\tilde{f}_b^G(x)\right) &= \frac{1}{2}xf_0\left(x\right)\left[r_0^{(2)}(x)-r^{(2)}(x,\theta_2^0)\right]b+o(b),\\
Bias\left(\tilde{f}_b^{IG}(x)\right) &= \frac{1}{2}x^3f_0\left(x\right)\left[r_0^{(2)}(x)-r^{(2)}(x,\theta_2^0)\right]b+o(b),\\
Bias\left(\tilde{f}_b^{RIG}(x)\right) &= \frac{1}{2}xf_0\left(x\right)\left[r_0^{(2)}(x)-r^{(2)}(x,\theta_2^0)\right]b+o(b).
\end{aligned}$$

The variances of the asymmetric LMBC kernel estimator are the same as those in the pure nonparametric case.

Note that global performance measures such as MISE are easy to derive from these results (see Section 4.5).

4.3 Special cases

After developing the general LMBC framework, properties of special cases are now derived. As described in Section 2, direct local modelling of the density can be obtained by choosing the parametric start density as an improper uniform distribution. W.l.o.g. we can set $f_0\left(x\right)$ to one. The local model $r\left(t,\theta\right)$ is then the only source of bias reduction. As soon as we fit two or more local parameters ($q\geq 2$), Proposition 1 implies that the bias of the asymmetric local likelihood estimator is $\frac{1}{2}x^a\left[r_0^{(2)}(x)-r^{(2)}(x,\theta_2^0)\right]b$, where a is equal to three for the inverse Gaussian and one for the other asymmetric kernels. The asymmetric version of the multiplicatively corrected kernel estimator of H&G emerges from choosing the weighting function as $1/f\left(t,\hat{\theta}_1\right)$ and choosing the local model as a constant. From Equation (4.2) it follows that the estimator is in this case

$$\tilde{f}_b(x) = f(x,\hat{\theta}_1)\hat{r}(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(X_i;x,b\right)\frac{f(x,\hat{\theta}_1)}{f(X_i,\hat{\theta}_1)}.$$

This estimator has the advantage that it is very simple to implement. Chen's asymmetric kernel estimator has therefore an implicit initial parametric start which is given by an improper uniform distribution. The ratio $f(x,\hat{\theta}_1)/f(X_i,\hat{\theta}_1)$ equals one in this case. This time, no boundary correction terms are needed which contrasts with symmetric kernels. This is because the asymmetric kernel already answers the boundary bias issue. An estimation technique closely related to H&G is the multiplicative bias correction approach developed by Jones, Linton and Nielsen (1995) (JLN)[6]. The analogue of their estimator for asymmetric kernels is

$$\bar{f}(x) = \hat{f}_b(x)\hat{\alpha}(x) = \frac{1}{n}\sum_{i=1}^{n} K\left(X_i;x,b\right)\frac{\hat{f}_b(x)}{\hat{f}_b(X_i)},$$

where $\hat{f}_b(x)$ is the usual asymmetric kernel estimator given in Equation (3.1). For symmetric kernels, this estimator has generally a smaller bias than the standard kernel estimator at the cost of a slightly larger variance. We do not further pursue this idea here since the JLN bias correction procedure is fully nonparametric. It should be possible to derive properties of this estimator in the asymmetric case combining techniques given in JLN and Chen (2000).

4.4 Examples

After developing the general framework, we now show how this framework can be applied to the estimation of densities with support on the nonnegative real line. We focus especially on examples which are relevant for the estimation of loss and income distributions. We give examples for the asymmetric LMBC and the asymmetric version of the H&G estimator, and also explore the case where the local model is chosen to fit the density directly as in H&J and Loader (1996).

4.4.1 A gamma start

A parametric start which is sufficiently flexible and can be expected to be appropriate in applications for unimodal and right skewed distributions is given by the gamma density[7]. This parametric start can

[6] This estimator can be obtained by choosing a nonparametric start estimator and choosing $1/\hat{f}_b\left(x\right)$ instead of $1/f_b\left(x,\hat{\theta}_1\right)$ as the weighting function. However, this time the first estimation step does influence the variance of the final estimator, and therefore we can not embed this estimator in the LMBC framework.

[7] An example based on the gamma (and also log-normal) density using symmetric kernels can be found in H&G. Their estimator suffers, however, from boundary bias like standard kernel estimators.

be combined with a local polynomial model for the correction factor: $r(t,\theta_2)=\theta_{21}+\theta_{22}(t-x)+...+\theta_{2(q+1)}(t-x)^q$. The estimator is $\tilde{f}_b(x)=f\left(x,\hat{\theta}_1\right)\hat{\theta}_{21}(x)$. The gamma start density in combination with the gamma kernel yields easy to implement estimators, which is a further advantage of our method and obviously of considerable importance for practical empirical investigations.

Example 1 Using the above model for the correction factor, choosing the weight functions $(t-x)^j$ for $j=0,...,q$ and using Equation (4.2), one can easily establish that the semiparametric density estimator with a gamma start $f_G\left(x,\hat{\alpha},\hat{\beta}\right)$ for a general order q is

$$\tilde{f}_b(x,q)=f_G\left(x,\hat{\alpha},\hat{\beta}\right)e_1^t\begin{bmatrix}\delta_0 & ... & \delta_q\\ ... & ... & ..\\ \delta_q & ... & \delta_{2q}\end{bmatrix}^{-1}\begin{bmatrix}\hat{f}_b^0(x)\\ ...\\ \hat{f}_b^q(x)\end{bmatrix},$$

where

$$\begin{aligned}\delta_j &= \int_0^\infty K_G\left(t;x/b+\hat{\alpha},b\frac{\hat{\beta}}{\hat{\beta}+b}\right)(t-x)^j\,dt,\\ c &= \frac{\Gamma\left(\frac{x}{b}+\hat{\alpha}\right)\left(b\frac{\hat{\beta}}{\hat{\beta}+b}\right)^{x/b+\hat{\alpha}}}{\Gamma(\hat{\alpha})\,\hat{\beta}^{\hat{\alpha}}\,\Gamma(x/b+1)\,b^{x/b+1}},\end{aligned}\tag{4.7}$$

and $\hat{f}_b^j(x)$ is the sample average of $(X_i-x)^j K_G\left(X_i;\frac{x}{b}+1,b\right)$. Choosing the local model as a constant is not particularly attractive since the bias of this estimator contains also the first order derivative of the correction function and the local model. These first derivative terms vanish if we choose a local polynomial model for the correction factor with $q\geq 2$. In particular, the local linear version of this estimator has the same bias as the asymmetric H&G estimator if K is chosen as the RIG or IG kernel:

$$\tilde{f}_b(x)=\frac{1}{n}\sum_{i=1}^n K(X_i;x,b)\left(\frac{x}{X_i}\right)^{\hat{\alpha}-1}\exp\left\{-\hat{\beta}(x-X_i)\right\}.\tag{4.8}$$

Both estimators are attractive when the true density is close to the gamma family. Otherwise a local model for the correction factor is desirable to capture curvature and further diminishes the bias. This can be attained by choosing $q\geq 3$. We will consider the gamma kernel version of the estimator in Equation (4.8) in our simulation study and will refer to it as the AHGG estimator.

Example 2 An alternative to local polynomial modelling of the correction factor is fitting a polynomial to the logarithm of the correction factor, choosing $r(t,\theta_2)=\theta_{21}\exp\left(\theta_{22}(t-x)+...+\theta_{2(q+1)}(t-x)^q\right)$. Compared to direct polynomial fitting as described above, this ensures a positive estimator and promises a better performance than the H&G estimator if the true density is not given by the parametric start[8].

[8] Note from (4.4) that the leading term of the bias of this estimator is given by $\frac{1}{2}f_0(x)\{r_0^{(2)}(x)-r^{(2)}(x,\theta_2^0)\}xb$ compared to the H&G estimator which has $\frac{1}{2}f_0(x)\,r_0^{(2)}(x)$ as the leading term.

We work out below the local log linear version of this estimator with a gamma start, again using the gamma kernel. Using Equation (4.2) and the score as the weighting function, the equation system to be solved is

$$\begin{aligned}\hat{f}_b(x) &= c\theta_{21}\exp\left(-\theta_{22}x\right)\psi\left(\theta_{22}\right), && (4.9)\\ \hat{f}_b^1(x) &= c\theta_{21}\exp\left(-\theta_{22}x\right)\left[\psi^{(1)}\left(\theta_{22}\right) - x\psi\left(\theta_{22}\right)\right], && (4.10)\end{aligned}$$

where $\psi\left(\theta_{22}\right)$ is the moment generating function associated with $K_G\left(t; x/b + \hat{\alpha}, b\frac{\hat{\beta}}{\hat{\beta}+b}\right)$ and c is as defined in (4.7). This function is[9]

$$\psi\left(\theta_{22}\right) = \left(1 - \beta^*\theta_{22}\right)^{-\alpha^*} \qquad \text{for } \theta_{22} \le \beta^{*-1},$$

where $\alpha^* = x/b + \hat{\alpha}$, $\beta^* = b\frac{\hat{\beta}}{\hat{\beta}+b}$. Since $\tilde{f}_b(x) = f_G\left(x, \hat{\alpha}, \hat{\beta}\right)\hat{\theta}_{21}$, $\hat{\theta}_{22}$ is only somewhat "silently" present in the local parameterization. Using this and Equations (4.9) and (4.10) one obtains

$$\hat{\theta}_{22} = \frac{(q+x) - \alpha^*\beta^*}{\beta^*\left(q+x\right)}, \tag{4.11}$$

where $q = \hat{f}_b^1(x)/\hat{f}_b(x)$. From (4.9), we can then obtain a closed form expression for the LMBC estimator with a gamma start and the log linear correction factor

$$\tilde{f}_b(x) = f_G\left(x, \hat{\alpha}, \hat{\beta}\right)\frac{\hat{f}_b(x)}{c\exp\left(-\hat{\theta}_{22}x\right)\psi\left(\hat{\theta}_{22}\right)}. \tag{4.12}$$

So with a gamma start this estimator is clearly simple to implement and should reveal appropriate in many circumstances because of the shape flexibility of the gamma distribution. We will refer in the Monte Carlo Section to the estimator of Equation (4.12) as the ALMBC estimator. Unfortunately this simplicity does not extend to other parametric starts than the gamma density. There the integrals corresponding to those in Equation (4.2) have to be calculated numerically.

4.4.2 Lognormal and Weibull start

Whereas the integral in Equation (4.2) can be analytically evaluated when we use a gamma kernel in combination with a gamma start, this is no longer true for other popular densities which have support on the nonnegative real line. There, numerical integration techniques are required. It is therefore convenient to choose the H&G weight function which automatically solves this problem. This simplicity of the H&G estimator makes this approach particularly attractive. We develop here two examples based on parametric start densities which are used in the literature for income and loss distribution modelling. We follow the notation of Klugman et al. (1998).

[9] It can easily be checked that the solution $\hat{\theta}_{22}$ always satisfies the restriction to be smaller than $1/\beta^*$.

Example 3 One popular parametric model for loss and income distributions is given by the lognormal probability law $LN(\mu, \sigma)$. This parametric model is usually thought as the best overall choice to fit loss data. The asymmetric H&G estimator is

$$\tilde{f}_b(x) = \frac{1}{n} \sum_{i=1}^{n} K(X_i; x, b) \frac{\exp\left\{-\frac{1}{2} (\log x - \hat{\mu})^2 / \hat{\sigma}^2\right\}}{\exp\left\{-\frac{1}{2} (\log X_i - \hat{\mu})^2 / \hat{\sigma}^2\right\}} \frac{X_i}{x}.$$

Example 4 Another useful parametric start is the Weibull $W(\theta, \tau)$ probability law. The exponential density results if $\tau = 1$. We have then:

$$\tilde{f}_b(x) = \frac{1}{n} \sum_{i=1}^{n} K(X_i; x, b) \left(\frac{X_i}{x}\right)^{1-\hat{\tau}} \exp\left(\left(\frac{1}{\hat{\theta}}\right)^{\tau} \left(X_i^{\hat{\tau}} - x^{\hat{\tau}}\right)\right).$$

Klugman et al. (1998) provide many suitable parametric densities to model loss distributions. Any of them or a mixture of densities yielding multimodal features can be used as a parametric start. We show later in this paper how suitable parametric start densities can be selected and also propose a semiparametric specification test to determine whether a density belongs to a particular parametric family.

4.4.3 Direct density modelling

Whereas there exists a huge literature concerning parametric estimation of income and loss distributions, in other areas of research it is often not clear what an appropriate parametric start for the density of interest would be. In this case, direct local modelling of the density is an alternative option. Loader (1996) concentrates on local polynomial fitting to the logarithm of the density under consideration and H&J argue that this is more attractive in semiparametric terms than direct local polynomial fitting. We refer to Hagmann and Scaillet (2003) for user friendly estimators based on the local polynomial model. Below we concentrate on an asymmetric kernel version of the popular local log linear estimator which is very simple to implement and yields, unlike the local linear estimator, always nonnegative density estimates.

Example 5 We choose directly for the density a local model given by $f(t, \theta_2) = \theta_{21} \exp(\theta_{22}(t - x))$. This choice of a local log-linear density is attempting to get the right local slope. If we take the score as the weighting function and use the same procedure as in Example 2, the semiparametric density estimator is

$$\tilde{f}_b(x) = \frac{\hat{f}_b(x)}{\exp\left(-\hat{\theta}_{22} x\right) \psi\left(\hat{\theta}_{22}\right)}, \tag{4.13}$$

where $\psi(\theta_{22})$ is the m.g.f. induced by $K_G(t; x/b + 1, b)$ and $\hat{\theta}_{22}$ is given in (4.11) for $\alpha^* = x/b + 1$ and $\beta^* = b$. This estimator is, therefore, straightforward to implement and always nonnegative. In our simulation study, we will refer to this estimator as the ALLL estimator. Unfortunately this simplicity does not extend to higher order approximations. There the integrals corresponding to those in Equation (4.2) have to be calculated numerically.

We remark that one can also derive a semiparametric density estimator which fits locally a probability density function. This could be of interest when the parameters of the density have some economic interpretation, as in the case of measures of inequality for example. We refer to Hagmann and Scaillet (2003) for a semiparametric estimator based on a local gamma model.

4.5 Choice of bandwidth

The mean square error optimal smoothing parameter at point x for the asymmetric LMBC estimator using the gamma kernel is in the interior

$$b_G^*(x) = \left(\frac{1}{2\sqrt{\pi}}\right)^{2/5} x^{-1} \left(\frac{f(x)}{\left\{f_0(x)\left[r_0^{(2)}(x) - r^{(2)}(x,\theta_2^0)\right]\right\}^2}\right)^{2/5} n^{-2/5}.$$

Note that the optimal smoothing parameter is large if the parametric guess is close to the true model. The optimal mean squared error is

$$\mathrm{MSE}_G^*(x) = \frac{5}{4}\left(\frac{f(x)}{2\sqrt{\pi}}\right)^{4/5} \left\{f_0(x)\left[r_0^{(2)}(x) - r^{(2)}(x,\theta_2^0)\right]\right\}^{2/5} n^{-4/5},$$

and does not depend on x. The optimal $MSE_G^*(x)$ in the boundary is of a less desirable order. Chen (2000) shows, however, that the impact on the $MISE$ is asymptotically negligible. Therefore, regarding global properties, the optimal bandwidth and mean integrated squared error for the gamma kernel are:

$$\begin{aligned} b_G^{**}(x) &= \left(\frac{\int_0^\infty \frac{1}{2\sqrt{\pi}} x^{-1/2} f(x)\,dx}{\int_0^\infty x^2\left[f_0(x)\left[r_0^{(2)}(x) - r^{(2)}(x,\theta_2^0)\right]\right]^2 dx}\right)^{2/5} n^{-2/5}, \\ \mathrm{MISE}_G^{**}(x) &= \frac{5}{4}\frac{\left(\int_0^\infty \frac{1}{2\sqrt{\pi}} x^{-1/2} f(x)\,dx\right)^{4/5}}{\left(\int_0^\infty x^2\left[f_0(x)\left[r_0^{(2)}(x) - r^{(2)}(x,\theta_2^0)\right]\right]^2 dx\right)^{-1/5}} n^{-4/5}. \end{aligned}$$

Hence these estimators achieve the optimal rate of convergence for the MISE within the class of non-negative kernel density estimators. Corresponding expressions for the RIG kernel and the IG kernel can be derived similarly.

A popular bandwidth selection method for symmetric kernels is unbiased least squares cross validation (LSCV). The idea of this method is to estimate the MISE of the multiplicatively corrected asymmetric kernel estimator and then minimize this expression with respect to the smoothing parameter. A nearly unbiased[10] estimator of MISE$-\int f(x)^2\,dx$ is

$$\mathrm{LSCV}(b) = \int_0^{\infty} \tilde{f}_b(x)^2\,dx - \frac{2}{n}\sum_{i=1}^{n} \tilde{f}_{b(i)}(X_i), \tag{4.14}$$

where $\tilde{f}_{b(i)}$ is the estimator constructed from the reduced data set that excludes X_i. For the asymmetric H&G estimator, Equation (4.14) can be shown to evaluate as

$$\frac{1}{n^2}\sum_{i,j}^{n} \frac{1}{f(X_i,\hat{\theta})f(X_j,\hat{\theta})}\int_0^{\infty} f(x,\hat{\theta})^2 K(X_i;x,b)\,K(X_j;x,b)\,dx$$
$$-\frac{2}{n(n-1)}\sum_{i}\sum_{j\neq i} K(X_j;X_i,b)\,\frac{f(X_i,\hat{\theta}_{(i)})}{f(X_j,\hat{\theta}_{(i)})},$$

where $\hat{\theta}_{(i)}$ is computed without X_i. One could also consider a varying smoothing parameter. We do not pursue this idea here since asymmetric kernels already vary the amount of smoothing through their changing shape. Furthermore, second generation bandwidth selection methods such as the smoothed bootstrap for symmetric kernels could be extended to the asymmetric kernel case. For a survey, see Jones, Marron and Sheather (1996).

4.6 Model diagnostics

The estimated correction factor delivers useful information for model diagnostics. The correction factor should equal one if the parametric start density coincides with the true density. We restrict our analysis in this subsection to the H&G estimator. This is because this estimator is already unbiased under true model conditions. Also, the specification test we propose below based on a parametric bootstrap procedure requires fast computation of the estimator.

H&G propose to check model adequacy by looking at a plot of the correction factor for various potential models with pointwise confidence bands to see if $r(x) = 1$ is reasonable. This plot allows to spot easily where misspecification is locally the largest. For the gamma kernel estimator the bias and variance of the correction factor are

$$E(\hat{r}(x)) = r(x) + b\left[r^{(1)}(x) + \frac{1}{2}xr^{(2)}(x)\right] + o(b), \tag{4.15}$$

$$\mathrm{Var}(\hat{r}(x)) = \begin{cases} \frac{1}{2\sqrt{\pi}} n^{-1} b^{-1/2} x^{-1/2} \frac{r(x)}{f_0(x)} & \text{if } x/b \to \infty, \\ \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)} n^{-1} b^{-1} \frac{r(x)}{f_0(x)} & \text{if } x/b \to \kappa. \end{cases} \tag{4.16}$$

[10] H&G show that in the symmetric case this estimator is nearly unbiased already for small samples. Since the arguments are not different in our case, we refer to their paper.

Another possibility, in the symmetric case also proposed by H&G, is to plot the log correction factor $\log \hat{r}(x)$ to see how far it is from zero. Hagmann and Scaillet (2003) derive bias and variance of this curve. From the results given there, a simple graphical goodness-of-fit emerges: plot x against

$$Z(x) = \begin{cases} \frac{\log \hat{r}(x) + \left(4\sqrt{\pi} n\right)^{-1} (bx)^{-1/2} f\left(x,\hat{\theta}\right)^{-1}}{\left\{\left(2\sqrt{\pi} n\right)^{-1} (bx)^{-1/2} f\left(x,\hat{\theta}\right)^{-1}\right\}^{1/2}} & \text{if } x/b \to \infty, \\ \frac{\log \hat{r}(x) + \Gamma(2\kappa+1)\left(2^{2(1+\kappa)}\Gamma^2(\kappa+1) nb\right)^{-1} f\left(x,\hat{\theta}\right)^{-1}}{\left\{\Gamma(2\kappa+1)(2^{1+2\kappa}\Gamma^2(\kappa+1) nb)^{-1} f\left(x,\hat{\theta}\right)^{-1}\right\}^{1/2}} & \text{if } x/b \to \kappa. \end{cases} \tag{4.17}$$

When the parametric start coincides with the true density, this is approximately distributed as standard normal for each x, meaning that the curve should move within ± 1.96 about 95% of the time.

Figure 2 provides an example. 500 random values from a Gamma $G(1.5, 1)$ were drawn and the density was estimated by the asymmetric HG estimator with a gamma and a Weibull start. Figure 2 shows that both estimators perform well. As expected, the correction factor for the gamma start estimator is close to one, whereas some nonparametric correction is done for the Weibull start estimator, especially in the tail of the density. Figure 3 plots the Z-statistic given in (4.17) for both estimators. Whereas the Z-statistic for the gamma start estimator is always within the confidence bands, the Weibull start estimator is outside at some of the points. The violation is not large. This is because the Weibull can capture the above gamma specification fairly well. Figure 4 shows the same procedure when the true density is LN$(0, 1)$. The correction factors for both estimators indicate that neither a gamma nor a Weibull can capture the tail of lognormal data. Also the close fit shows that although the parametric start is clearly wrong, the density is fitted quite well due to the nonparametric correction for misspecification.

The present framework can also be used to test if the data was generated by a particular parametric model $f(x, \theta)$ where $\theta \in \Theta$ (H_0). We propose the global test statistic

$$T_n = nb^{1/2} \int_0^\infty \varphi(x) \left[\hat{r}(x) - 1\right]^2 dx, \tag{4.18}$$

where $\varphi(x)$ is some appropriately chosen weighting function. Asymptotic normality of this statistic could be shown using results given in Fernandes and Monteiro (2005). It is, however, well known that similar tests based on symmetric kernel estimators are very sensitive to the choice of the smoothing parameter. Fan (1995, 1998) reports that for a wide range of values of the smoothing parameter the test statistics can have large skewness and kurtosis exhibiting behaviour more like χ^2 tests than normal tests. Size distortions can therefore be quite large. Fan shows that the parametric bootstrap can solve these problems, and we therefore propose the following standard procedure to determine the critical value of the test:

Step 1. Draw a random sample of size n, $\left\{X_j^*\right\}_{j=1}^n$, from the distribution with density function $f\left(x, \hat{\theta}\right)$

where $\hat{\theta}$ is estimated by maximum likelihood from the original data. This is the bootstrap sample. Hence conditional on the random sample $\{X_j\}_{j=1}^n$, the boostrap sample satisfies H_0 with $\theta = \theta_0$.

Step 2. Use the bootstrap sample $\left\{X_j^*\right\}_{j=1}^n$ in place of the original data to compute T_n. Call it T_n^*.

Step 3. Repeat Step 1 and Step 2 for a large number of times, say B, and obtain the empirical distribution function of $\left\{T_{n,r}^*\right\}_{r=1}^B$, called the bootstrap distribution.

Let C_α be the upper $\alpha-$percentile of the calculated bootstrap distribution. Then reject the null hypothesis at significance level α if $T_n > C_\alpha$. T_n is small under the null hypothesis for two reasons. First because the parametric model is correct and $\hat{r}(x)$ should be close to one. Second because the estimator is unbiased and should, therefore, be more precisely measured under the null than under the alternative hypothesis. We therefore expect the power of this semiparametric test to be greater than that of pure nonparametric versions of this kind of specification tests.

Furthermore the statistic given in (4.18) can be used for an adequate choice of a parametric start. The density under consideration can be estimated with different parametric starts. Then one can choose that parametric start density for which the value of the above statistic is the smallest.

4.7 Extensions

Before turning to the Monte Carlo results, we finally would like to mention that our approach can easily be extended to the estimation of densities which have support on the interval $[0, 1]$. An application in credit risk is the estimation of the density of recovery rates at default, see Renault and Scaillet (2004), Hagmann, Renault and Scaillet (2005). To accommodate two known boundaries, Chen (1999) introduced asymmetric kernels based on the beta distribution K_B with parameters $(x/b + 1, (1-x)/b + 1)$ given by

$$K_B\left(t; \frac{x}{b} + 1, \frac{(1-x)}{b} + 1\right) = \frac{t^{x/b}(1-t)^{(1-x)/b} I(0 \leq t \leq 1)}{B(x/b + 1, (1-x)/b + 1)},$$

where $B(\cdot)$ denotes the beta function. The support of the kernel again matches the support of the density and the resulting estimates are free of boundary bias. An obvious parametric start is given by the beta family of densities. Writing Equation (4.2) using a beta kernel and performing analogous calculations as before, one can establish the bias and variance of the beta kernel version of the LMBC estimator.

Proposition 2 *The bias and variance expressions of the asymmetric LMBC estimator in the case where*

K is the beta kernel is given for $q \geq 2$ *by:*

$$
\begin{aligned}
Bias\left(\tilde{f}_b^B(x)\right) &= \frac{1}{2}x(1-x)f_0(x)\left[r^{(2)}(x) - r^{(2)}(x,\theta_2^0)\right]b + o(b),\\
Var\left(\tilde{f}_b^B(x)\right) &= \begin{cases} \frac{1}{2\sqrt{\pi}}n^{-1}b^{-1/2}\{x(1-x)\}^{-1/2}f(x) & \text{if } x/b \to \infty,\\ \frac{\Gamma(2\kappa+1)}{2^{1+2\kappa}\Gamma^2(\kappa+1)}n^{-1}b^{-1}f(x) & \text{if } x/b \to \kappa. \end{cases}
\end{aligned}
$$

The variance of this semiparametric density estimator coincides with that of the pure nonparametric beta kernel estimator. We refer to Chen (1999). Compared to the bias of the nonparametric beta kernel estimator, first order derivative terms of the true density vanish (as $q \geq 2$) in the bias expression of the semiparametric estimator. Also, $f^{(2)}(x)$ is replaced by $f_0(x)\left[r^{(2)}(x) - r^{(2)}(x,\theta_2^0)\right]$. The same remarks apply for the comparison of these biases as before.

Finally the LMBC approach could be extended to a multivariate setting through use of product kernels without any particular difficulties .

5 Monte Carlo study

In this section we evaluate the finite sample performance of most of the estimators considered in the previous section. For estimators involving asymmetric kernels, we focus on the gamma kernel. This because, as demonstrated earlier on, the use of the gamma kernel allows us to obtain semiparametric estimators in closed form. This attractive property of the asymmetric gamma kernel does not transfer to the RIG kernel and the IG kernel, where numerical integration and optimization has to be used to obtain density estimates. This makes the RIG kernel and the IG kernel somewhat less attractive for a large scale simulation study as well as empirical work[11]. To the best of our knowledge, it is the first time that various semiparametric density estimators are compared on a finite sample basis.

5.1 Semiparametric estimators and test densities

We run a Monte Carlo simulation for the following semiparametric density estimators:

- the pure nonparametric gamma kernel estimator (G1) ,

[11] We examined the performance of the RIG kernel and the IG kernel in case of the HG estimator, where a closed form solution is available. Results for the RIG were similar as for the HG estimator relying on the gamma kernel, whereas the IG version performed significantly worse. This is in line with results reported in Scaillet (2004), who examines the performance of those asymmetric kernels in a pure nonparametric setting.

- the uncorrected semiparametric HG estimator with a gamma start, using the Epanechnikov kernel (SHGG),
- the local linear HG estimator with a gamma start given in Equation (2.6), using the Epanechnikov kernel (SHGGC),
- the semiparametric HG estimator with a gamma start given in Equation (4.8), using the gamma kernel (AHGG),
- the LMBC estimator with a gamma start and a log linear correction factor, given in Equation (4.12) (ALMBC),
- the local log linear estimator using the gamma kernel given in Equation (4.13) (ALLL).

We compare these estimators on three different test densities: a Gamma $G(1.5, 1)$, a Weibull $W(1, 1.5)$ and a lognormal $LN(0, 1)$. The G1 estimator takes the role of the benchmark for the other estimators. A useful semiparametric estimator should at least in some cases outperform its pure non-parametric competitor. The SHGG, SHGGC and AHGG are all HG type estimators which use the gamma as a start density. The only source of bias reduction achieved by these estimators is provided by the global parametric start. Note that the SHGGC is a direct competitor to AHGG. Indeed they are both free of boundary bias. We will see that the HG estimator with a symmetric kernel without any boundary correction (SHGG) yields in fact very unsatisfactory results. All the considered HG type density estimators should perform well for the gamma test density since the correction factor can be estimated without bias. We also expect them to perform well for the Weibull test density. This because the gamma start can come close to a Weibull density, implying that the correction factor exhibits only small curvature and is therefore simple to estimate. This is demonstrated in the left panel of Figure 5[12], where the Weibull and corresponding pseudo gamma density are plotted. The right panel shows the pseudo gamma when the true data is drawn from the lognormal test density. In this case, the gamma does not provide a reasonable start. The correction factor exhibits high curvature and is therefore more difficult to estimate. Also recall our example in Figure 4. The ALMBC estimator should perform better in this situation, since the additional local model for the correction factor theoretically leads to an improvement over the HG type estimators. Finally, the performance of the ALLL estimator is of interest because it does not need any global parametric start but provides a pure local bias correction[13].

[12] The pseudo gamma parameter values are calculated via Monte Carlo integration based on a sample of one million Weibull or lognormal random values.

[13] For comparison purposes, we also tried to implement a local log linear estimator with a symmetric kernel. However, this estimator was not suitable for a large scale simulation study, since computation in the boundary of the density requires numerical search procedures in each single step.

5.2 Design of the Monte Carlo study

The performance measures we consider are the integrated squared error (ISE) and the weighted integrated squared error (WISE) of the various estimators:

$$\begin{aligned} \text{ISE} &= \int_{-\infty}^{+\infty} \left\{ \tilde{f}(x) - f(x) \right\}^2 dx, \\ \text{WISE} &= \int_{-\infty}^{+\infty} \left\{ \tilde{f}(x) - f(x) \right\}^2 x^2 dx. \end{aligned}$$

The WISE allows us to capture the tail performance of our estimators. The experiments are based on 1,000 random samples of length $n = 100$, $n = 200$, $n = 500$ and $n = 1,000$. We provide a "best case" analysis, meaning that for each simulated sample the ISE was computed over a grid of bandwidths and the minimum value was chosen. The WISE is computed in each simulation step with the same bandwidth as the ISE[14]. Numerical integration was performed by Gauss Legendre quadrature with 96 knots.

5.3 Numerical issues

In a first simulation step the SHGGC estimator surprisingly performed much worse than the uncorrected estimator SHGG. The reason was that the correction factor can sometimes, especially in the boundary, become too influential. Following H&G, we implemented a trimmed version of the estimator given in Equation (2.6) using

$$\begin{aligned} f\left(x, \hat{\theta}_1\right) \hat{r}(x) &= \frac{1}{n} \sum_{i=1}^{n} \mathcal{K}_h(X_i - x) \min\left(\frac{f\left(x, \hat{\theta}_1\right)}{f\left(X_i, \hat{\theta}_1\right)}, a \right), \\ f\left(x, \hat{\theta}_1\right) \hat{g}(x) &= \frac{1}{n} \sum_{i=1}^{n} \mathcal{K}_h(X_i - x)(X_i - x) \min\left(\frac{f\left(x, \hat{\theta}_1\right)}{f\left(X_i, \hat{\theta}_1\right)}, a \right), \end{aligned}$$

where $a > 0$. Similar "clipping" precautions in the density estimation setting are recommended by Abramson (1982) and Terrell and Scott (1992). This trimming procedure successfully solved the numerical problems for the symmetric kernel based estimator. From our experience, $a \in [10, 50]$ is a satisfactory choice. In fact within that range, the Monte Carlo results for this estimator were only insignificantly influenced. The semiparametric asymmetric kernel density estimators did not suffer from

[14] Another procedure would be to compute the WISE as well over a grid of bandwidths and choose the minimum value in each simulation step. We do not follow this because we want to evaluate the tail performance of our estimators given that they fit the whole density well. This is achieved by computing the WISE with the ISE-minimizing bandwidth in each simulation step.

the same numerical problem and were implemented without any trimming in the form described in Section 4.

5.4 Results

Table 1 shows the simulation results for the MISE criterion with standard errors of each simulation experiment reported in brackets. The AHGG estimator brings large improvements over the G1 estimator when the parametric start is true or close to the true density. Even for the lognormal example, its performance is still slightly better. This is because the uniform distribution, which is the implicit start for the G1 estimator, is quite a conservative start. The performance of the ALMBC is especially interesting when the parametric start is a poor specification for the true density. Whereas the ALMBC shows as expected a similar performance as the AHGG estimator for the gamma and Weibull test densities, the additional local model for the correction factor brings another 20% improvement for the lognormal test density. The ALMBC estimator also performs uniformly better than the ALLL estimator, which yields compared to the G1 estimator an improvement between 15-25% across all test densities and sample sizes. The ALLL however does not rely on a parametric start and may perform better when misspecification is stronger than the one considered here.

The poor performance of the SHGG estimator demonstrates how important the boundary bias feature is in the semiparametric framework considered in this paper. Even when the parametric start is correct, SHGG performs worse than the pure nonparametric G1 estimator. Partly, this is because the boundary bias prevents an enlarged bandwidth. Although the trimmed local linear version of this estimator brings a large improvement compared to the uncorrected symmetric estimator, its performance is considerably lower than that of the AHGG estimator. In the slightly misspecified Weibull case with a sample size of 1,000, the MISE of the AHGG estimator shrinks to 46% of the MISE of the SHGGC estimator. Chen (2000) has already reported that the asymmetric kernel estimator performs better than its symmetric local linear competitor. The outperformance in our case is however much larger, since the boundary bias problem magnifies in our semiparametric framework as mentioned earlier.

Table 2 shows the same information but for the WISE criterion and makes the power of the asymmetric estimators obvious. Those estimators perform much better in the tail of the density than estimators based on symmetric kernels. For the lognormal density which has the largest tail among the test densities, the WISE of the AHGG estimator is just one third of the WISE of its symmetric kernel based competitor SHGGC. We expect this relative advantage to increase for densities that have heavier tails than the lognormal density, e.g. Pareto distributions. This tail advantage of the asymmetric kernel is due to its changing shape as one moves away from the boundary. ALMBC exhibits excellent performance also with respect to the WISE criterion.

6 Empirical applications

In this section we illustrate the usefulness of our estimation approach with two empirical applications. The first one deals with health insurance data provided by a large Swiss health insurer. The second application deals with Brazilian income data.

6.1 Application to health insurance data

The Swiss health insurance system is heavily regulated by law. All residents in Switzerland have a compulsory base insurance which covers general health expenses. In the year 2002, approximately one third of total health expenses of 45 billion Swiss Francs were covered by this base insurance, which is offered by different private insurers.

Since the cost structure among different cantons in Switzerland is very different, we focus here on claims generated by residents of the canton of Zurich. The data considered is the net payment per client in the year 2002, covering claims for the base insurance only. We show how our approach can be used to compare the shape of the loss distribution for different subpopulations to better assess each underlying risk.

Table 3 shows the descriptive statistics of our dataset. It is evident that this dataset is highly skewed and exhibits large kurtosis. Also, the average payment varies significantly with the gender and age of the subpopulations. Obviously, the claim structure is also very different depending on whether the client lives in Zurich City or in rural area. We note that the "thought of solidarity" in the Swiss health system implies that clients with an age above 26 years pay all the same premium for their base insurance, independent of age and gender.

Figure 6 shows the loss distribution for the whole sample estimated for comparison purposes by the SHGGC estimator and our ALMBC estimator, both using a gamma start. These two estimators were the best symmetric and asymmetric estimators emerging from our Monte Carlo Study. To avoid numerical overflow, the original dataset was divided by 2,000. Then the resulting density has been back transformed. The bandwidth was calculated using the LSCV procedure described in Section 4.5. We report the bandwidth chosen for the transformed data for all considered estimators in Table 4. We also tried the ALLL estimator, but the resulting density cannot be distinguished by eye from that of the ALMBC estimator and is therefore not plotted. The shape of the loss distribution is very typical; we have a peak for the small claim sizes and then a very long tail. At first glance, the estimates for the symmetric and asymmetric estimator seem quite similar. The correction factor in the right panel of Figure 6 shows, however, that the asymmetric estimator produces a smooth tail, whereas the symmetric estimator features a bumpy behaviour. The picture shows that we have to correct around the mode and also in the tails. The correction factor further left to the picture (not shown) is increasing up to a factor of 10. This is because the gamma start cannot fully capture the heavy tail of the underlying density. The imperfect start and also the bandwidth selected for the ALMBC and ALLL estimator implies that the performance of those estimators is quite similar in terms of precision. Both estimators use a larger bandwidth than the gamma kernel estimator (not plotted), indicating that they both reduce successfully the bias. This allows us to choose a larger bandwidth compared to the pure nonparametric gamma kernel estimator, which reduces the variance of the estimate. Other parametric starts could be used, e.g. a Pareto distribution to capture better the tail of the density.

Figures 7 and 8 show the loss distributions for different subpopulations, they all seem to be very reasonable and as one would expect a priori. In particular, younger people have smaller claims than older people and are less risky. Our estimators capture very well the heavy tail of clients with an age above 55. Also, the loss distributions for the young people subpopulation does not have a mode, but could be unbounded at zero. This because the majority of their claims are very small.

Although from a social point of view it may be human to charge gender and age independent premiums for health insurance, it is hard to understand, why the Swiss system does not allow the charging of location dependent premiums inside cantons. Figure 7 shows that clients living in Zurich City have a completely different risk structure than people living in the close rural neighbourhood. However, premiums within Canton of Zurich are legally restricted to be the same. Of course, to investigate that point more closely, one would have to condition on the age and gender structure more carefully, but the overall picture would hardly change dramatically.

We conclude that our proposed estimators, which are very simple to apply, seem to be a very useful estimation device for risk managers in insurance companies, and should help to design more differentiated premiums whenever allowed.

6.2 Application to Brazilian income data

Our second application concerns the analysis of the income distribution of Brazil in the year 1990. We analyse a large micro data set (n=71,523), which has been collected by the PNAD annual national household survey. The data set is interesting because Brazil is a major world economy (ninth largest GDP) and faces a strong inequality in terms of percentage shares of income accruing to the richest and to the poorest of its population. The evolution of the Brazilian income distribution in the 1980's has been examined by Cowell, Ferreira and Litchfield (1998). The data considered is monthly household income per capita denominated in 1990 cruzeiros. The strong distributional inequality is revealed by the high skewness of the income distribution, we refer to Table 5 for the descriptive statistics.

We start our analysis with the ALBMC estimator, featuring an implicit gamma start. The parameters of the gamma start density, evaluated by maximum likelihood, are given by $\left(\hat{\alpha};\hat{\beta}\right) = (0.89; 58,861)$, which would imply that the Brazilian income density is unbounded at zero. Figure 9 shows however that the ALMBC estimator does not confirm this gesture. The correction factor in the right panel approaches zero to diminish this effect. Also the correction factor indicates that the gamma model underestimates the mode of the true density as well as its tail. The ALLL estimator cannot be distinguished by eye from the ALMBC and is therefore not plotted. The original dataset was divided by 10,000 and the resulting density estimate back transformed. Again, the bandwidths for the different estimators were chosen according to LSCV and are reported in Table 6 for the transformed data. It is interesting to see that the bandwidths chosen for the ALMBC and ALLL estimators are much larger than that for the gamma kernel estimator (not plotted). This is because both semiparametric estimators can successfully reduce the bias, which allows us to increase the bandwidth and therefore reduces the variance of the estimates.

Cowell et. al. (1998) mention that the Brazilian income distribution is well approximated by a lognormal model. The above results indicate that this is not very likely at the boundary since $f(0)$ seems not to be zero. Apart from this, as can be seen in Figure 9, the lognormal start for an HG type estimator seems to be very appropriate. The relatively large bandwidth chosen by the LSCV procedure also confirms that the lognormal start contains valuable information.

At this stage we also provide a formal test for lognormality of the underlying income distribution using the test statistic given in Equation (4.18). The estimated parameters for the lognormal model are given by $(\hat{\mu}, \hat{\sigma}) = (10.20, 1.13)$. We use $\varphi(x) = f\left(x, \hat{\theta}_1\right)$ as a weighting function. This implies that we impose a heavy penalty if the difference between the semiparametric and parametric density estimate is large at those locations where the parametric model puts a lot of weight. We use the bootstrap procedure described in Section 4.6 with $B = 1,000$ to approximate the finite sample distribution of the test statistic, using the empirical bandwidth chosen for this data set. We plot the bootstrap density, estimated by the ALLL estimator, of the test statistic under the null hypothesis of lognormality in Figure 10. Although the sample is very large, a Jarque-Bera statistic of 726 rejects normality of this bootstrap density at any conventional significance level. This confirms that it is better to use a parametric bootstrap rather than relying on asymptotic results for computing critical values. The sample test statistic is given by 29.96 which compares to a critical value of 4.03 at the 1% level. So we reject the null hypothesis of lognormality of the Brazilian income distribution. This is not surprising since we work with a large sample size and it is unlikely that the underlying density can be described by just two parameters. However as we demonstrated above, the lognormal start contains very valuable information for our semiparametric modelling.

7 Concluding remarks

In this paper we have presented a semiparametric estimation framework based on asymmetric kernels for the estimation of densities on the nonnegative real line. This framework allows us to use popular parametric models from the field of actuarial science and income distribution estimation in a nonparametric fashion. Although the approach may look cumbersome at first glance, it reduces in many important cases to estimators that take closed forms and are thus very easy to implement. Our simulation results show that our estimators, especially the ALMBC estimator with a parametric start and a local model for the correction factor, exhibit excellent performance. They should therefore be useful in applied work in economics, finance and actuarial science involving non- and semiparametric techniques. This point has already been demonstrated with two empirical applications to health insurance data and Brazilian income data. The results developed here could also be exploited with straightforward modifications in regression curve and hazard rate estimation.

8 Appendix

8.1 Bias and variance of the LMBC estimator

Along the same lines as H&J, we start with the asymptotic analysis assuming a fixed smoothing parameter. The estimator $\hat{\theta}_2(x)$ we consider is the solution to Equation (4.2). For convenience we partly suppress the fixed x in the notation now. A Taylor expansion around θ_2^0 yields

$$(nb^{1/2})^{1/2}(\hat{\theta}_2-\theta_2^0) \simeq -V_n^*(\hat{\theta}_1,\theta_2^0)^{-1}(nb^{1/2})^{1/2}V_n\left(\hat{\theta}_1,\theta_2^0\right),$$

where $V_n^*(\hat{\theta}_1,\theta_2^0)$ is the $p\times p$ matrix of partial derivatives of the $V_{n,j}(\hat{\theta}_1,\theta_2)$ functions. Paralleling arguments in the supplementary section of H&J[15], we can establish asymptotic normality of the local parameter estimator $\hat{\theta}_2(x)$:

$$\left(nb^{1/2}\right)^{1/2}\left(\hat{\theta}_2-\theta_2^0\right) \xrightarrow{d} N\left(0, J_b^{-1}M_b\left(J_b^{-1}\right)^t\right),$$

where

$$J_b(x)=\int_0^\infty K(t;x,b)\left[v(t,\theta_2^0)u(t,\theta_2^0)^t m(t,\theta_1^0,\theta_2^0)+V^*\left(t,\theta_2^0\right)\left\{f(t)-m(t,\theta_1^0,\theta_2^0)\right\}\right]dt, \tag{8.1}$$

$$\begin{aligned} M_b(x) &= b^{1/2}Var_f\left(K(X_i;x,b)\,v(x,X_i,\theta_2^0)\right)\\ &= b^{1/2}\int_0^\infty K(t;x,b)^2\,v(t,\theta_2^0)v(t,\theta_2^0)^t f(t)dt-b^{1/2}\xi_b\xi_b^t, \end{aligned} \tag{8.2}$$

and $\xi_b=\int_0^\infty K(t;x,b)\,v(t,\theta_2^0)f(t)dt$. By the delta method the asymptotic distribution of $\tilde{f}_b(x)$ for a fixed smoothing parameter is

$$\left(nb^{1/2}\right)^{1/2}\left(\tilde{f}_b(x)-m(x,\theta_1^0,\theta_2^0)\right) \xrightarrow{d} N\left(0, m(x,\theta_1^0,\theta_2^0)^2u(x,\theta_2^0)^tJ_b^{-1}M_b\left(J_b^{-1}\right)^t u(x,\theta_2^0)\right). \tag{8.3}$$

In a next step we evaluate the above expressions when the smoothing parameter $b\to 0$ as $n\to\infty$. Note that

$$V(x,\theta_2^0)=\int_0^\infty K(t;x,b)\,v(x,t,\theta_2^0)f_0(x)\{r_0(t)-r(t,\theta_2^0)\}dt=E\left[q(\gamma_x,\theta_2^0)\right],$$

where $q(\gamma_x,\theta_2^0)=v(\gamma_x,\theta_2^0)f_0(\gamma_x)\{r_0(\gamma_x)-r(\gamma_x,\theta_2^0)\}$ is a $p\times 1$ vector and γ_x is a random variable whose distribution is determined by the choice of the asymmetric kernel. We develop below the bias

[15] We have the additional problem that $V_n\left(\hat{\theta}_1,\theta_2^0\right)$ depends on the first step estimator $\hat{\theta}_1$. A Taylor expansion of $V_n\left(\hat{\theta}_1,\theta_2^0\right)$ around θ_1^0 and the fact that $\hat{\theta}_1$ is $\sqrt{n}$ convergent shows however immediately, that this is not an issue. To ease notation, we partly suppress θ_1^0 in the following.

expression for the gamma kernel. In this case, γ_x is a $G(\frac{x}{b}+1,b)$ random variable[16]. Concentrating on component j of the vector q and noting that $\mu_x = E(\gamma_x) = x+b$ and $Var(\gamma_x) = xb+b^2$, we perform a Taylor expansion around μ_x and obtain

$$E\left[q_j(\gamma_x,\theta_2^0)\right] = q_j(x,\theta_2^0) + \left[q_j^{(1)}(x,\theta_2^0) + \frac{1}{2}xq_j^{(2)}(x,\theta_2^0)\right]b + o(b), \tag{8.4}$$

which equals zero at θ_2^0 and therefore

$$v_{j,0}(x)\, f_0(x)\left\{r_0(x) - r(x,\theta_2^0)\right\} = \left[q^{(1)}(x,\theta_2^0) + \frac{1}{2}xq^{(2)}(x,\theta_2^0)\right]b + o(b). \tag{8.5}$$

From (8.5) it follows that $r_0(x) - r(x,\theta_2^0) = O(b)$. This together with the fact that $E\left[\tilde{f}_b^G(x)\right] = f_0(x)\, r(x,\theta_2^0) + O\left(\frac{1}{nb^{1/2}}\right)$ can be used to obtain the bias of the LMBC gamma kernel estimator:

$$\begin{aligned}
\text{Bias}\left(\tilde{f}_b^G(x)\right) &= \frac{1}{v_{j,0}(x)}\left\{\left[q^{(1)}(x,\theta_2^0) + \frac{1}{2}xq^{(2)}(x,\theta_2^0)\right]b + o(b)\right\} + O\left(\frac{1}{nb^{1/2}}\right) \\
&= f_0(x)\left[\{r_0^{(1)}(x) - r^{(1)}(x,\theta_2^0)\} + \frac{1}{2}x\{r_0^{(2)}(x) - r^{(2)}(x,\theta_2^0)\}\right]b \\
&\quad + \left(\frac{v_{j,0}^{(1)}(x)}{v_{j,0}(x)}f_0(x) + f_0^{(1)}(x)\right)x\{r_0^{(1)}(x) - r^{(1)}(x,\theta_2^0)\}b + o(b) + O\left(\frac{1}{nb^{1/2}}\right),
\end{aligned}$$

which is the result given in the text.

Concerning the variance of the LMBC estimator, we demonstrate here results in the one parameter case. For full details in the multiple parameter case, we refer to Hagmann and Scaillet (2003).

We have to evaluate the expressions $J_b(x)$ and $M_b(x)$ as b approaches zero. In a first step

$$\begin{aligned}
J_b(x) &= \int_0^\infty K(t;x,b)\, v(t,\theta_2^0)u(t,\theta_2^0)m(t,\theta_1^0,\theta_2^0)dt + O(b) \\
&= E\left[c\left(\gamma_x,\theta_2^0\right)\right],
\end{aligned}$$

where $c\left(\gamma_x,\theta_2^0\right) = v(\gamma_x,\theta_2^0)u(\gamma_x,\theta_2^0)m(\gamma_x,\theta_1^0,\theta_2^0)$ and γ_x is random variable whose distribution depends on the choice of the asymmetric kernel. We again demonstrate the result for the gamma kernel, where γ_x follows a $G(\frac{x}{b}+1,b)$[17]. Proceeding as in (8.4)

$$E\left[c(\gamma_x,\theta_2^0)\right] = c(x,\theta_2^0) + \left[c^{(1)}(x,\theta_2^0) + \frac{1}{2}xc^{(2)}(x,\theta_2^0)\right]b + o(b).$$

[16] For the inverse gaussian or reciprocal inverse gaussian kernel, γ_x is an $IG(x,1/b)$ or $RIG(1/(x-b),1/b)$ random variable respectively. The bias derivation follows the same lines as demonstrated in this appendix for the gamma kernel, but using different expressions in the taylor expansions as outlined in Scaillet (2004).

[17] The variance expression in the IG and RIG case can be developed in the same steps as for the gamma kernel, using results described in Scaillet (2004).

In the one parameter case the first integral term in Equation (8.2) is

$$\int_0^\infty K\left(t;x,b\right)^2 v(t,\theta_2^0)^2 f(t)dt = B_b\left(x\right)E\left[\eta\left(\zeta_x,\theta_2^0\right)\right],$$

where ζ_x follows a $G(\frac{2x}{b}+1,\frac{b}{2})$ random variable, $\eta\left(\zeta_x,\theta_2^0\right)=v(\zeta_x,\theta_2^0)^2 f(\zeta_x)$ and

$$B_b(x)=\frac{\Gamma\left(2x/b+1\right)/b}{2^{2x/b+1}\Gamma^2\left(x/b+1\right)}.$$

Applying the same trick as in (8.4), one can show that the first term in $M_b\left(x\right)$ is $b^{1/2}B_b\left(x\right)v(x,\theta_2^0)^2 f(x)+O(b^{3/2})$. Proceeding similarly with the second term in Equation (8.2) we get that

$$\xi_b = w(x,\theta_2^0)+O(b),$$

where $w(x,\theta_2^0)=v(x,\theta_2^0)f(x)$. This yields

$$b^{1/2}\xi_b^2 = b^{1/2}w(x,\theta_2^0)^2+O(b^{3/2}).$$

Collecting terms,

$$M_b\left(x\right)=b^{1/2}\left[B_b\left(x\right)v(x,\theta_2^0)^2 f(x)-w(x,\theta_2^0)^2\right]+O(b^{3/2}).$$

Having derived these preliminary results and using (8.3), we can now tackle the variance of the asymmetric LMBC gamma kernel estimator in the one parameter case.

$$\begin{aligned}\operatorname{Var}\left(\tilde{f}_b^G(x)\right) &= \frac{1}{nb^{1/2}}\left[m(x,\theta_1^0,\theta_2^0)^2 u(t,\theta_2^0)^2 J_b^{-2}M_b\right] \\ &= \frac{1}{nb^{1/2}}m(x,\theta_1^0,\theta_2^0)^2 u(t,\theta_2^0)^2\left(\frac{1}{c(x,\theta_2^0)+O\left(b\right)}\right)^2 \\ &\quad\times\left[b^{1/2}\left[B_b\left(x\right)v(x,\theta_2^0)^2 f(x)-w(x,\theta_2^0)^2\right]+O(b^{3/2})\right] \\ &= \frac{1}{n}B_b\left(x\right)f\left(x\right)-\frac{f(x)^2}{n}+O\left(b/n\right).\end{aligned}$$

Using the approximation result for $B_b\left(x\right)$ given in Chen (2000, p. 474) proves Equation (4.5) in the main text.

Acknowledgements Both authors gratefully acknowledge financial support from the Swiss National Science Foundation through the National Centre of Competence in Research: Financial Valuation and Risk Management (NCCR-FINRISK). We would like to thank the editors, and the two referees for constructive criticism. We also would like to thank R. Abul Naga, N. Arnold, F. Cowell, C. Gouriéroux, O. Linton, A. Monfort, P. Funk, M. Rockinger, R. Stulz, E. von Thadden, B. Ziemba, participants at the 2003 ESRC Econometric Study Group Meeting in Bristol, the 2003 LAMES in Panama City, the 2004

RES Meeting in Swansea, the 2004 "Semiparametrics in Rio" Meeting, the 2004 Brazilian Symposium of Probability and Statistics in Caxambu, the 2004 ESEM in Madrid, the FAME Doctoral Workshop in Lausanne, the Swiss Doctoral Workshop in Finance in Gerzensee, the CERN seminar in Geneva, and the Economics Lunch seminar in Basle for helpful comments. J. Litchfield and H. Beck kindly provided the Brazilian income and health insurance data. Part of this research was done when the first author was visiting the LSE and the second author THEMA and ECARES.

References

[1] Abadir, M. K. and S. Lawford, 2004, Optimal Asymmetric Kernels, Economics Letters 83, 61-68.

[2] Abramson, I. S., 1982, On Bandwidth Variation in Kernel Estimates - A square Root Law, The Annals of Statistics 10, 1217-1223.

[3] Aït-Sahalia, Y., 1996a, Nonparametric Pricing of Interest Rate Derivative Securities, Econometrica 64, 527-560.

[4] Aït-Sahalia, Y., 1996b, Continuous-Time Models of the Spot Interest Rate, The Review of Financial Studies 9, 385-426.

[5] Bickel, P. J. and M. Rosenblatt, 1973, On some Global Measures of the Deviations of Density Function Estimates, The Annals of Statistics 1, 1071-1095.

[6] Bolancé, C., Guillen, M. and J.P. Nielsen, 2003, Kernel Density Estimation of Actuarial Loss Functions, Insurance: Mathematics and Economics 32, 3-18.

[7] 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.

[8] Chen, S.X., 1999, A Beta Kernel Estimator for Density Functions, Computational Statistics and Data Analysis 31, 131-145.

[9] Chen, S.X., 2000, Probability Density Function Estimation Using Gamma Kernels, Annals of the Institute of Statistical Mathematics 52, 471-480.

[10] Chen, S.X., 2002, Local linear smoothers using asymmetric kernels, Annals of the Institute of Statistical Mathematics 54, 312-323.

[11] Cowell, F., 1999, Measurement of Inequality, in: Atkinson, A. and F. Bourguignon, eds., Handbook of Income Distribution, North Holland, Amsterdam.

[12] Cowell, F., Ferreira, F. and J. Litchfield, 1998, Income Distribution in Brazil 1981-1990: Parametric and Non-Parametric Approaches, Journal of Income Distribution 8, 63-76.

[13] Cox, J.C., Ingersoll, J.E. and S.A. Ross, 1985, A Theory of the Term Structure of Interest Rates, Econometrica 53, 385-408.

[14] Engle, R.F., 2000, The Econometrics of Ultra-High Frequency Data, Econometrica 68, 1-22.

[15] Fan, J. and I. Gijbels, 1992, Variable Bandwidth and Local Linear Regression Smoothers, The Annals of Statistics 20, 2008-2036.

[16] Fan, J. and I. Gijbels, 1996, Local Polynomial Modelling and its Applications, London: Chapmann & Hall.

[17] Fan, Y., 1995, Bootstrapping a Consistent Nonparametric Goodness-of-Fit Test, Econometric Reviews 14, 367-382.

[18] Fan, Y., 1998, Goodness-of-Fit Tests based on Kernel Density Estimators with Fixed Smoothing Parameters, Econometric Theory 14, 604-621.

[19] Fernandes, M. and J. Grammig, 2005, Nonparametric Specification Tests for Conditional Duration Models, Journal of Econometrics 127, 35–68.

[20] Fernandes, M. and P.K. Monteiro, 2005, Central Limit Theorem for Asymmetric Kernel Functionals, Annals of the Institute of Statistical Mathematics 57, 425-442.

[21] Gozalo, P. and O.B. Linton, 2000, Local Nonlinear Least Squares: Using Parametric Information in Nonparametric Regression, Journal of Econometrics 99, 63-106.

[22] Gouriéroux, Ch., Monfort, A. and A. Trognon, 1984, Pseudo Maximum Likelihood Methods: Theory, Econometrica 52, 680-700.

[23] Gouriéroux, Ch. and J. Jasiak, 2001, Local Likelihood Density Estimation and Value at Risk, Discussion Paper CREST and CEPREMAP.

[24] Hagmann, M., Renault, O. and O. Scaillet, 2005, Estimation of Recovery Rate Densities: Nonparametric and Semi-parametric Approaches versus Industry Practice, in Recovery Risk: The Next Challenge in Credit Risk Management, Eds: Resti, A., Sironi, A., Altman E., Risk Publications, London, 323-346

[25] Hagmann, M. and O. Scaillet, 2003, Local Multiplicative Bias Correction for Asymmetric Kernel Density Estimators, FAME Working Paper 91.

[26] Hjort, N.L. and I.K. Glad, 1995, Nonparametric Density Estimation with a Parametric Start, The Annals of Statistics 24, 882-904.

[27] Hjort, N.L. and M.C. Jones, 1996, Locally Parametric Nonparametric Density Estimation, The Annals of Statistics 24, 1619-1647.

[28] Hogg, R. and S.A. Klugman, 1984, Loss Distributions, New York: John Wiley & Sons, Inc.

[29] Jones, M.C., 1993, Simple Boundary Correction for Kernel Density Estimation, Statistics and Computing 3, 135-146.

[30] Jones, M.C. and P. Foster, 1996, A Simple Nonnegative Boundary Correction Method for Kernel Density Estimation, Statistica Sinica 6, 1005-1013.

[31] Jones, M.C., Marron, J.S. and S.J. Sheather, 1996, A Brief Survey of Bandwidth Selection for Density Estimation, Journal of the American Statistical Association 92, 1063-1073.

[32] Jones, M.C., Linton, O.B. and J.P. Nielsen, 1995, A Simple but Effective Bias Reduction Method for Density and Regression Estimation, Biometrika 82, 327-338.

[33] Klugman, S.A., Panjer H.A. and G.E. Willmot, 1998, Loss Models: From Data to Decisions, New York: John Wiley & Sons, Inc.

[34] Linton, O.B. and A. Pakes, 2001, Nonlinear Estimators in Econometrics, manuscript, London School of Economics and Political Science.

[35] Loader, C.R., 1996, Local Likelihood Density Estimation, The Annals of Statistics 24, 1602-1618.

[36] Müller, H., 1991, Smooth Optimum Kernel Estimators near Endpoints, Biometrika 78, 521-520.

[37] Renault, O. and O. Scaillet, 2004, On the Way to Recovery: A Nonparametric Bias Free Estimation of Recovery Rate Densities, Journal of Banking and Finance 28, 2915-2931.

[38] Rice, J., 1984, Boundary Modification for Kernel Regression, Communications in Statistics: Theory and Methods 13, 893-900.

[39] Rosenblatt, M., 1956, Remarks on Some Nonparametric Estimates of a Density Function, The Annals of Mathematical Statistics 27, 642-669.

[40] Scaillet, O., 2004, Density Estimation Using Inverse and Reciprocal Inverse Gaussian Kernels, Journal of Nonparametric Statistics 16, 217-226.

[41] Schuster, E., 1985, Incorporating Support Constraints into Nonparametric Estimators of Densities, Communications in Statistics: Theory and Methods, 1123-1136.

[42] Terrell, G. R. and D. W. Scott, 1992, Variable Kernel Density Estimation, The Annals of Statistics 20, 1236-1265.

[43] Tyurin, K., 2003, Competing Risks and the Order Flow Dynamics in FOREX Electronic Brokerage, Discussion Paper, Indiana University.

[44] White, H., 1982, Maximum Likelihood Estimation of Misspecified Models, Econometrica 50, 1-26.

Figure 1: The gamma kernel function for different x values.

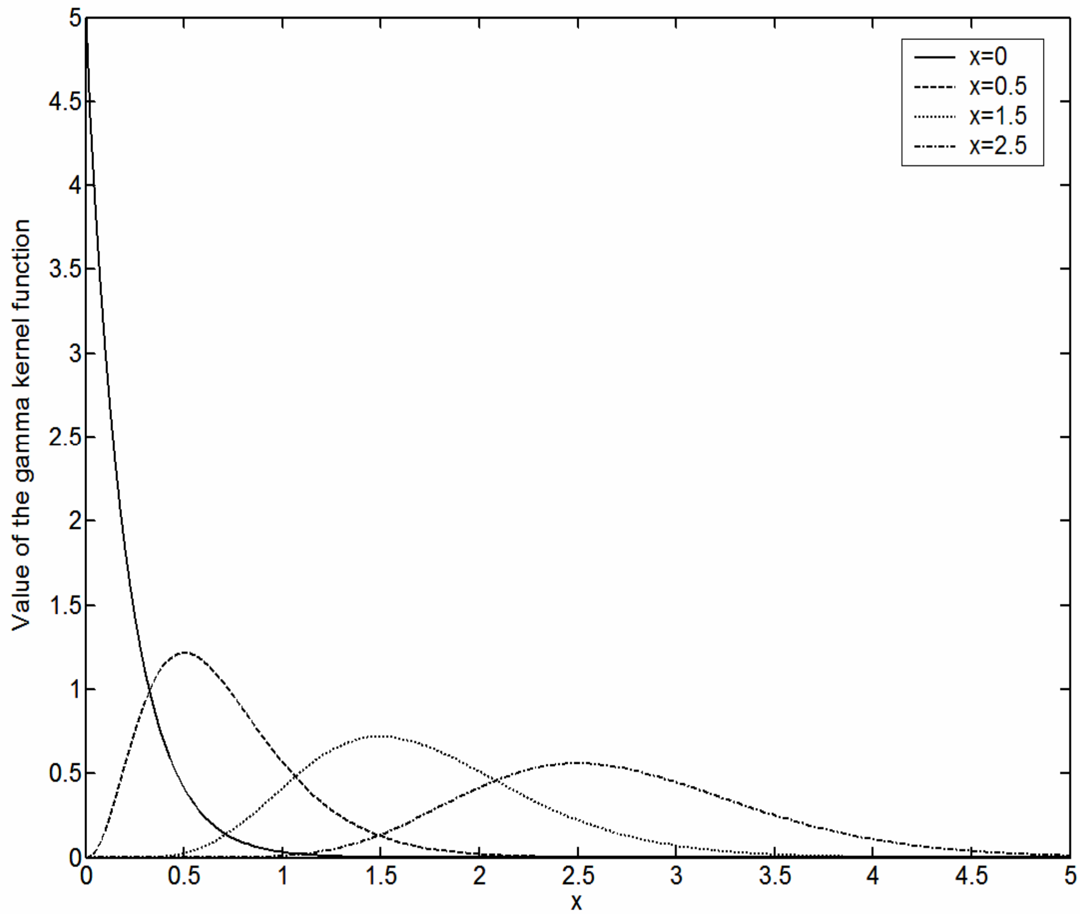

Figure 2: HG estimator with gamma and Weibull start. True density: G(1.5,1)

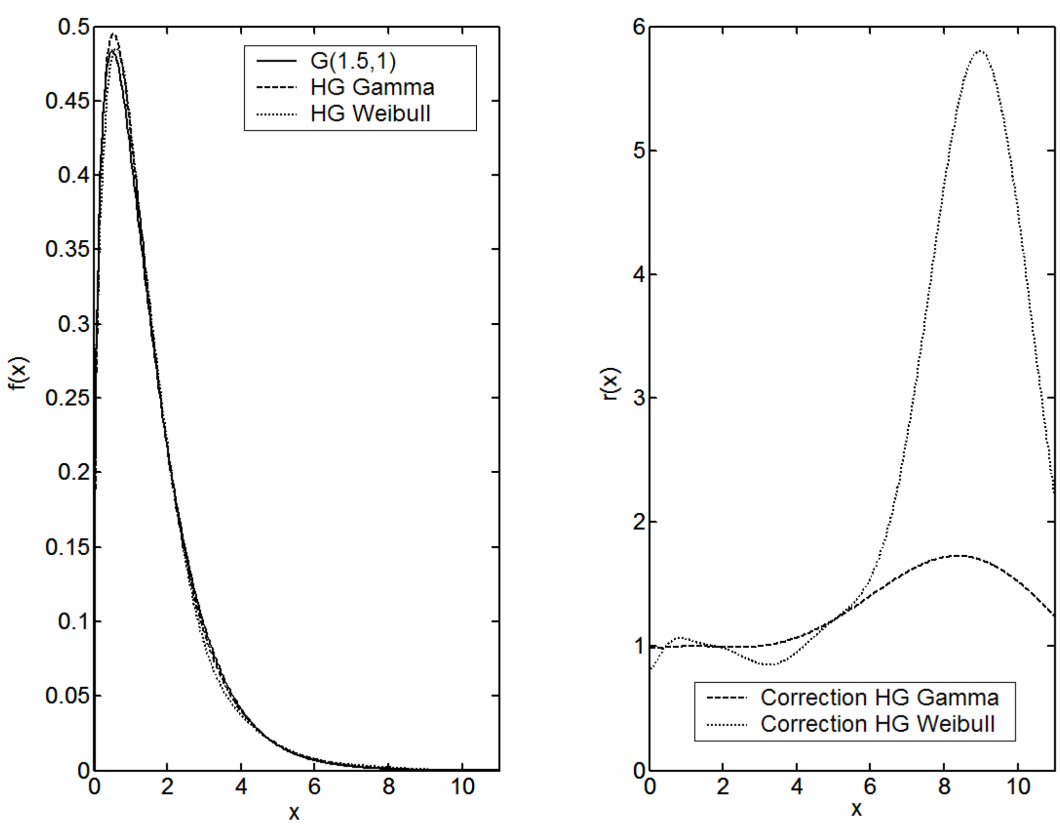

Figure 3: The Z-statistics associated with the examples in Figure 2.

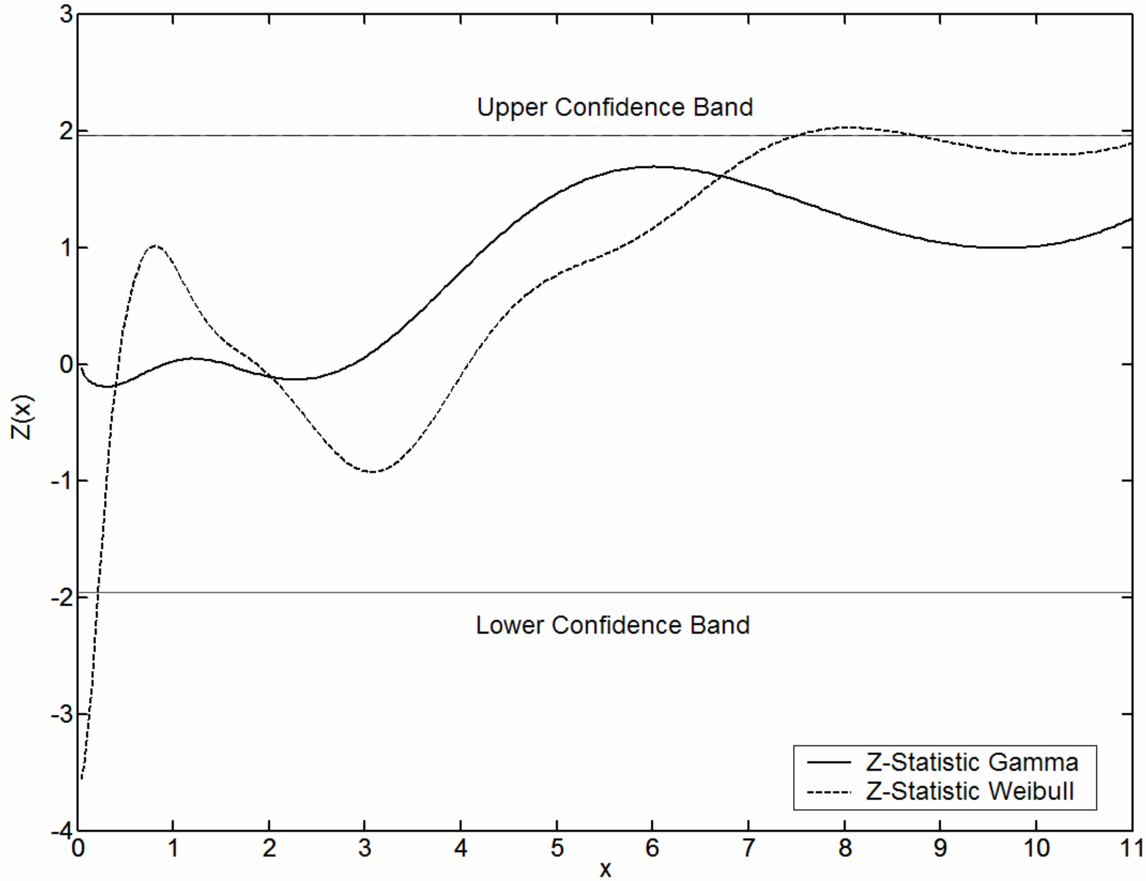

Figure 4: HG estimator with gamma and Weibull start. True density: LN(0,1)

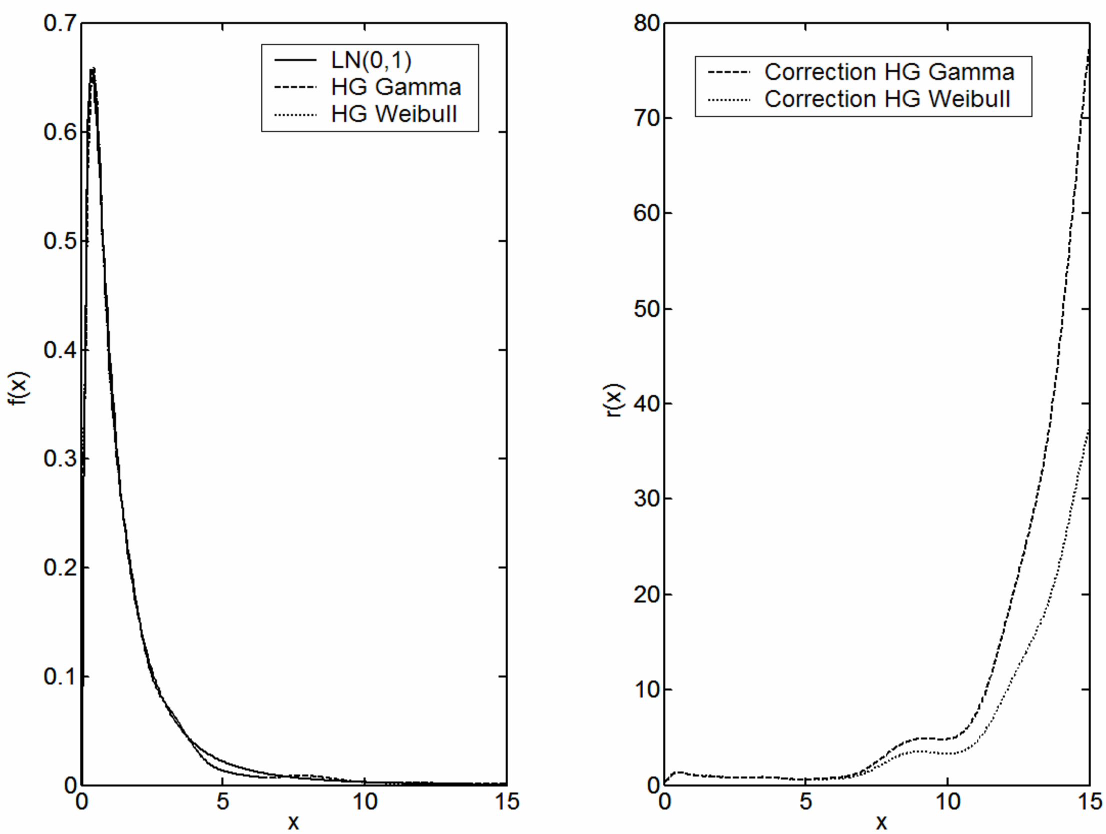

Figure 5: Pseudo gamma densities for the LN(0,1) and W(1,1.5).

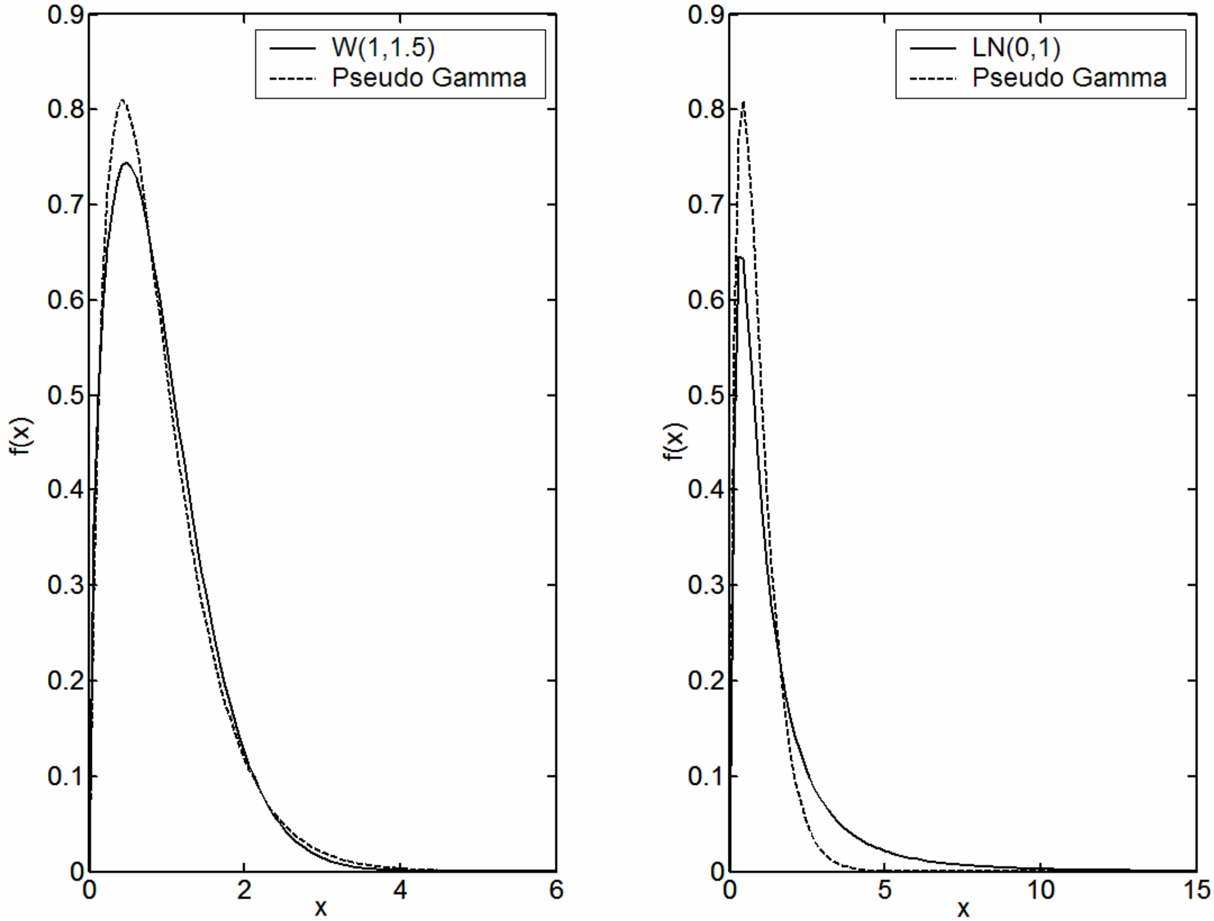

Figure 6: Loss distribution and correction factor for all clients.

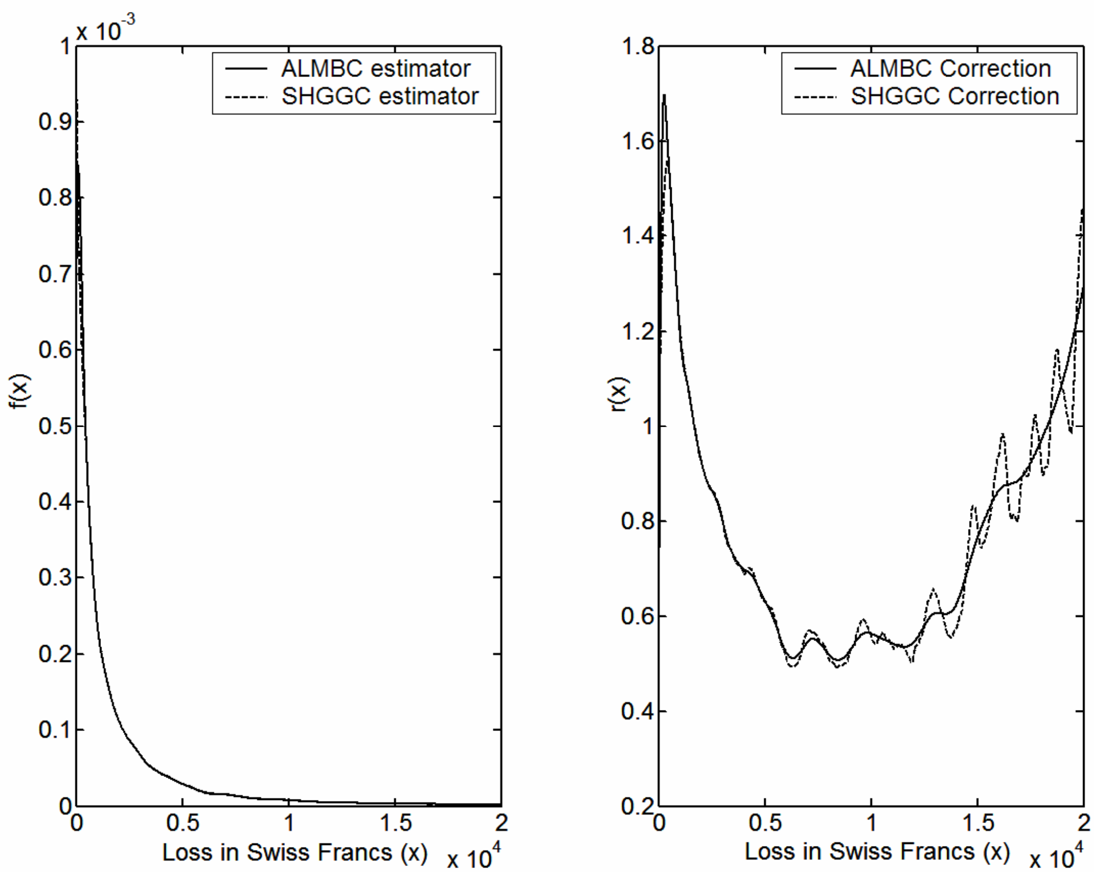

Figure 7: Loss distribution for Zurich City and countryside clients.

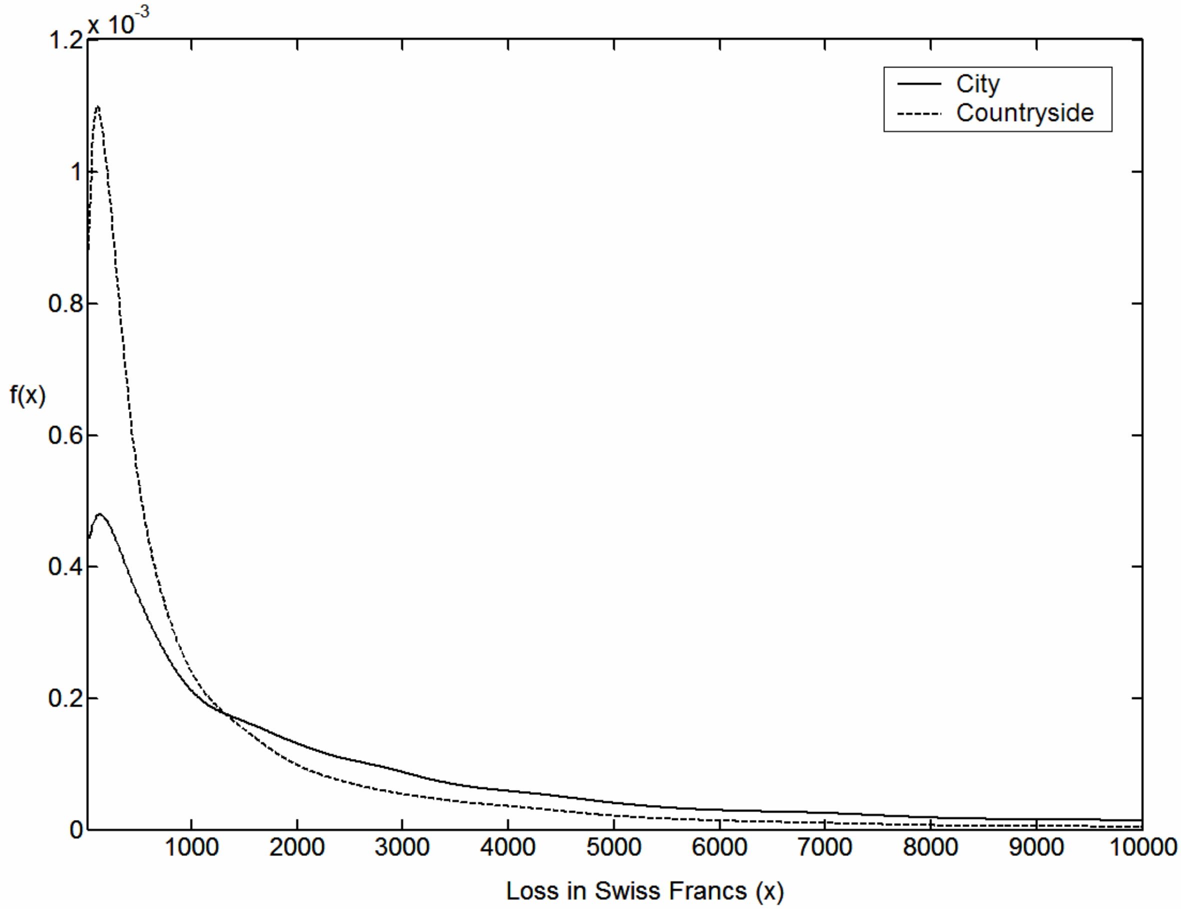

Figure 8: Loss distribution for clients with different age structure.

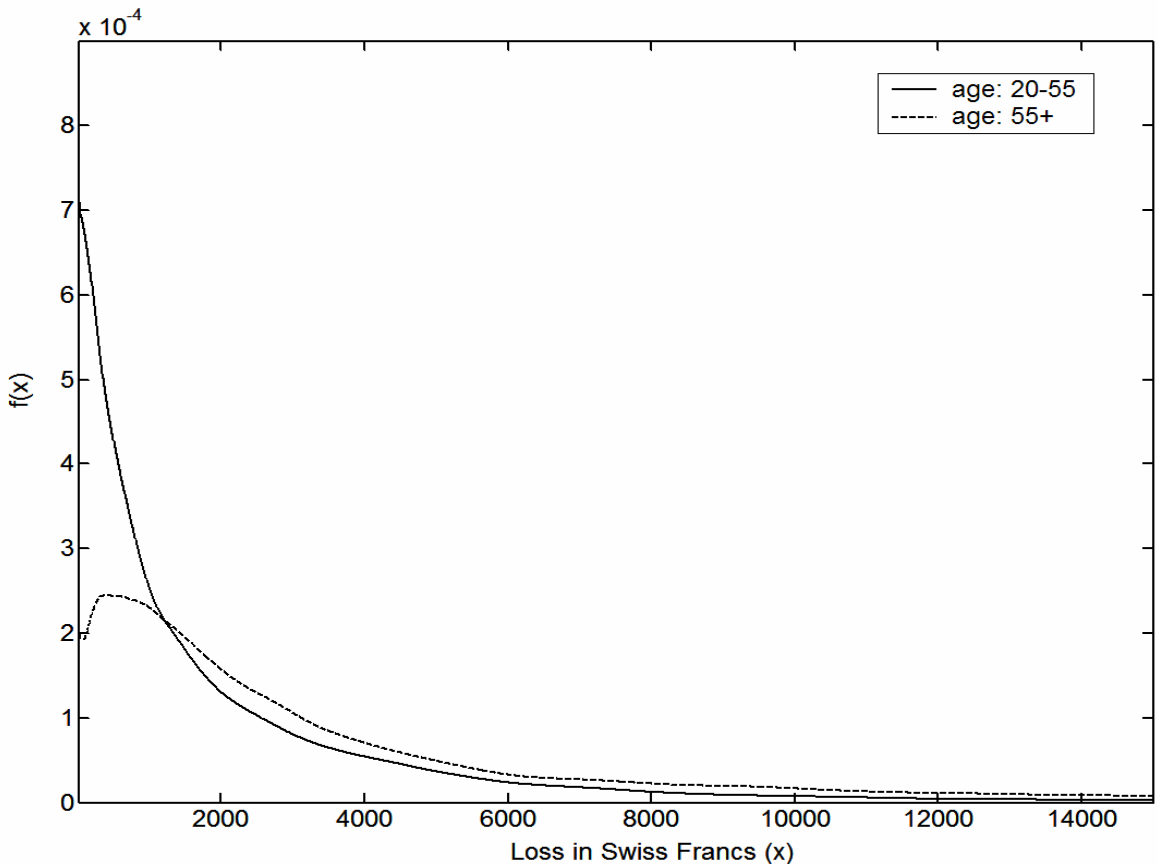

Figure 9: Brazilian income distribution and correction factors.

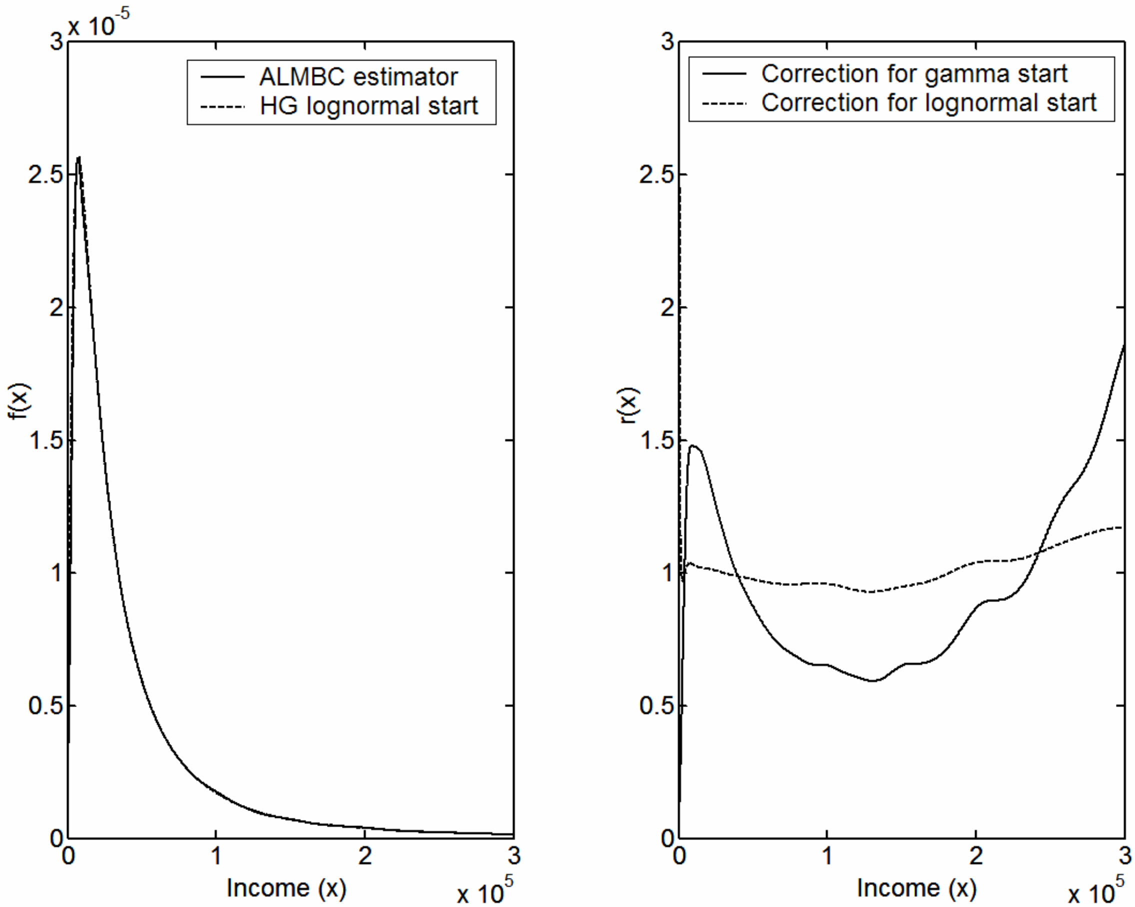

Figure 10: Density of the bootstraped test statistic values.

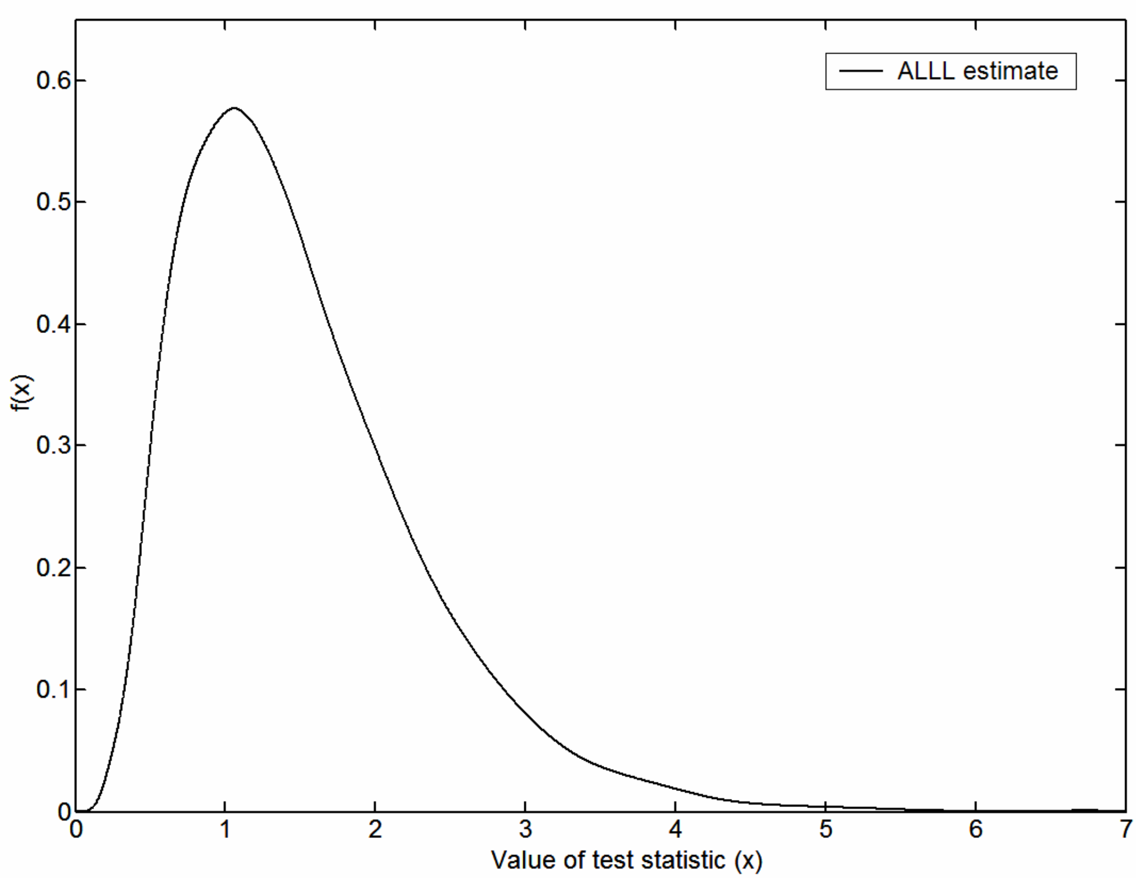

Table 1: Summary results for the monte carlo study for the MISE criterion, standard deviations for each simulation are reported in brackets.

Test density	G1	AHGG	SHGG	SHGGC	ALMBC	ALLL
G(1.5,1)						
n=100	778	388	1040	456	404	659
	(545)	(433)	(658)	(499)	(461)	(521)
n=200	517	181	658	216	190	404
	(341)	(195)	(366)	(220)	(209)	(287)
n=500	268	71	396	83	75	205
	(158)	(79)	(169)	(85)	(84)	(126)
n=1000	173	38	267	44	39	130
	(96)	(39)	(107)	(44)	(41)	(74)
W(1,1.5)						
n=100	1224	737	1210	915	735	1017
	(957)	(708)	(770)	(842)	(707)	(838)
n=200	782	412	793	561	417	625
	(554)	(354)	(418)	(466)	(335)	(432)
n=500	414	208	490	370	214	328
	(265)	(165)	(209)	(269)	(161)	(193)
n=1000	251	127	341	279	130	199
	(142)	(94)	(120)	(178)	(90)	(103)
LN(0,1)						
n=100	1289	1192	1618	1431	968	1051
	(727)	(770)	(799)	(803)	(600)	(651)
n=200	790	733	1039	924	586	631
	(433)	(435)	(450)	(470)	(350)	(384)
n=500	394	367	561	486	294	313
	(198)	(193)	(217)	(233)	(175)	(187)
n=1000	240	225	356	303	184	194
	(114)	(109)	(125)	(135)	(106)	(111)

Table 2: Summary results for the monte carlo study for the WISE criterion, standard deviations for each simulation are reported in brackets.

Test density	G1	AHGG	SHGG	SHGGC	ALMBC	ALLL
G(1.5,1)						
n=100	1422	785	3010	1165	708	1111
	(545)	(918)	(1799)	(1012)	(739)	(829)
n=200	902	380	1897	580	355	686
	(631)	(443)	(986)	(484)	(361)	(471)
n=500	481	150	1107	238	141	352
	(276)	(168)	(473)	(206)	(150)	(219)
n=1000	295	83	712	120	74	216
	(161)	(100)	(273)	(109)	(79)	(124)
W(1,1.5)						
n=100	895	492	1119	668	531	732
	(753)	(506)	(833)	(617)	(496)	(547)
n=200	548	275	704	403	303	443
	(395)	(275)	(418)	(328)	(256)	(295)
n=500	277	143	400	241	149	224
	(180)	(130)	(209)	(170)	(122)	(140)
n=1000	168	88	270	172	90	138
	(106)	(76)	(127)	(114)	(70)	(80)
LN(0,1)						
n=100	2907	2946	7851	6850	2155	2285
	(1623)	(1745)	(4085)	(3755)	(1261)	(1265)
n=200	1762	1719	4940	4472	1284	1367
	(851)	(898)	(2035)	(1892)	(669)	(680)
n=500	859	824	2563	2308	620	656
	(358)	(363)	(844)	(817)	(288)	(294)
n=1000	518	495	1604	1438	380	402
	(197)	(197)	(448)	(437)	(170)	(175)

Table 3: Sample Statistics for Health Insurance Data

	Wh. Sample	Men	Women	Age 20-55	Age 55+	Zurich City	Countryside
Sample Size	42,722.00	17,478.00	25,244.00	15,579.00	12,098.00	8,737.00	8,670.00
Mean payment	2,971.30	2,582.90	3,240.30	2,984.50	5,697.20	4,960.60	1,966.70
Stand. Dev.	6,671.70	6,488.60	6,782.70	6,542.10	9,119.70	9,221.60	5,017.40
Skewness	7.35	9.24	6.23	7.94	5.09	5.85	9.50
Kurtosis	102.94	152.85	74.74	102.53	58.44	71.41	151.12
Minimum	0.10	0.10	0.10	0.10	0.10	0.10	0.10
Maximum	202,870.00	202,870.00	169,650.00	132,890.00	202,870.00	202,870.00	126,700.00
1st quartile	320.30	278.80	359.35	403.20	1,051.80	575.40	237.25
Median	923.28	761.65	1,067.00	1,165.30	2,473.80	1,804.90	612.95
3rd quartile	2,695.80	2,193.40	3,039.20	3,014.40	6,000.10	4,879.10	1,749.50

Table 4: LSCV Bandwidth Health Data

	Whole	Age 20-55	Age 55+	City	Country
G1	0.0063	0.0116	0.0199	0.0135	0.0093
ALLL	0.0300	0.0711	0.1649	0.0873	0.0560
ALMBC	0.0443	0.1983	0.1331	0.1400	0.0975
SHGGC	0.2118	-	-	-	-

Table 5: Sample Statistics Income Data

	Income Data
Sample Size	71,523.00
Mean payment	52,183.00
Standard deviaton	90,661.00
Skewness	11.01
Kurtosis	319.32
Minimum	2.00
Maximum	5,011,000.00
1st quartile	12,165.00
Median	26,142.00
3rd quartile	57,000.00

Table 6: LSCV Bandwidth Income Data

Estimator	Bandwidth
G1	0.0233
ALLL	0.0835
ALMBC	0.0856
AHGLOGN	0.2168