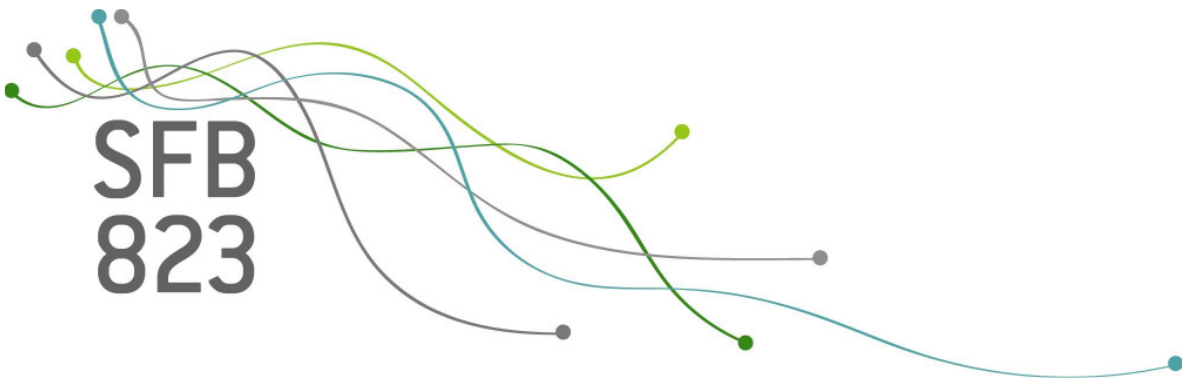
SFB 823

Combining regular and irregular histograms by penalized likelihood

Yves Rozenholc, Thoralf Mildenberger, Ursula Gather

Nr. 31/2009

Discussion Paper



Combining Regular and Irregular Histograms by Penalized Likelihood

Yves Rozenholc
UFR de Mathématiques et d'Informatique
Université Paris Descartes

Thoralf Mildenberger*
Fakultät Statistik
Technische Universität Dortmund

Ursula Gather
Fakultät Statistik
Technische Universität Dortmund

November 23, 2009

A fully automatic procedure for the construction of histograms is proposed. It consists of constructing both a regular and an irregular histogram and then choosing between the two. For the regular histogram, only the number of bins has to be chosen. Irregular histograms can be constructed using a dynamic programming algorithm if the number of bins is known. To choose the number of bins, two different penalties motivated by recent work in model selection are proposed. A complete description of the algorithm and a proper tuning of the penalties is given. Finally, different versions of the procedure are compared to other existing proposals for a wide range of densities and sample sizes. In the simulations, the squared Hellinger risk of the procedure that chooses between regular and irregular histograms is always at most twice as large as the risk of the best of the other methods. The procedure is implemented in an `R`-Package.

1. Introduction

For a sample $(X_1, X_2, \ldots, X_n)$ of a real random variable X with an unknown density f w.r.t. Lebesgue measure, we denote the realizations by $(x_1, x_2, \ldots, x_n)$ and the realizations of the order statistics by $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$. The goal in nonparametric density

*Address for correspondence: Fakultät Statistik, Technische Universität Dortmund, 44221 Dortmund, Germany. E-mail: `mildenbe@statistik.tu-dortmund.de`. Web: `http://www.statistik.tu-dortmund.de/mildenberger_en.html`

estimation is to construct an estimate $\hat{f}$ of f from the sample. In this work, we focus on estimation by histograms, which are defined as piecewise constant densities. The procedure we propose consists of constructing both a regular and an irregular histogram (both to be defined below) and then choosing between the two. Although other types of nonparametric density estimators are known to be superior to histograms according to several optimality criteria, histograms still play an important role in practice. The main reason is their simplicity and hence their interpretability (Birgé and Rozenholc, 2006). Often, the histogram is the only density estimator taught to future researchers in non-mathematical subject areas, usually introduced in an exploratory context without reference to optimality criteria.

We first introduce histograms and describe the connection to Maximum Likelihood estimation: Given $(x_1, x_2, \ldots, x_n)$ and a set of densities $\mathcal{F}$, the maximum likelihood estimate – if it exists – is given by an element $\hat{f} \in \mathcal{F}$ that maximizes the likelihood $\prod_{i=1}^n f(x_i)$ or equivalently its logarithm, the log-likelihood $L(f, x_1, \ldots, x_n)$:

$$\hat{f} := \text{argmax}_{f\in\mathcal{F}} L(f, x_1, \ldots, x_n) := \text{argmax}_{f\in\mathcal{F}} \sum_{i=1}^{n} \log(f(x_i)).$$

Without further restrictions on the class $\mathcal{F}$, the log-likelihood is unbounded, and hence, no maximum likelihood estimate exists. One possibility is to restrict $\mathcal{F}$ to a set of histograms. Consider a partition $\mathcal{I} := \{I_1, \ldots, I_D\}$ of a compact interval $K \subset \mathbb{R}$ into D intervals $I_1, \ldots, I_D$, such that $I_i \cap I_j = \emptyset$ for $i \neq j$ and $\bigcup I_i = K$. Now consider the set $\mathcal{F}_\mathcal{I}$ of all histograms that are piecewise constant on $\mathcal{I}$ and zero outside $\mathcal{I}$:

$$\mathcal{F}_\mathcal{I} := \left\{ f \,\middle|\, f = \sum_{j=1}^{D} h_j \mathbf{1}_{I_j}, h_j \geq 0, j = 1, \ldots, D \text{ and } \sum_{j=1}^{D} h_j |I_j| = 1 \right\},$$

where $\mathbf{1}_A$ denotes the indicator function of a set A and $|I|$ the length of the interval I. If K contains $[x_{(1)}, x_{(n)}]$, the *Maximum Likelihood Histogram* (ML histogram) is defined as the maximizer of the log-likelihood in $\mathcal{F}_\mathcal{I}$ and is given by

$$\hat{f}_\mathcal{I} := \text{argmax}_{f\in\mathcal{F}_\mathcal{I}} L(f, x_1, \ldots, x_n) = \frac{1}{n} \sum_{j=1}^{D} \frac{N_j}{|I_j|} \mathbf{1}_{I_j}, \tag{1}$$

with $N_j = \sum_{i=1}^n \mathbf{1}_{I_j}(x_i)$. Its log-likelihood is

$$L(\hat{f}_\mathcal{I}, x_1, \ldots, x_n) = \sum_{j=1}^{D} N_j \log \frac{N_j}{n|I_j|}. \tag{2}$$

In the following, we consider partitions $\mathcal{I} := \mathcal{I}_D := (I_1, \ldots, I_D)$ of the interval $I := [x_{(1)}, x_{(n)}]$, consisting of D intervals of the form

$$I_j := \begin{cases} [t_0, t_1] & j = 1 \\ (t_{j-1}, t_j] & j = 2, \ldots, D \end{cases},$$

with breakpoints $x_{(1)} =: t_0 < t_1 < \cdots < t_D := x_{(n)}$. A histogram is called *regular* if all intervals have the same length and *irregular* otherwise. The intervals are also referred to as *bins*.

We will only consider ML histograms in this work, and use the term "histogram" synonymously with "ML histogram" unless explicitly stated otherwise. We focus on finding a data-driven construction of a histogram with good risk behavior. Given a distance measure d between densities, the risk is defined as the expected distance between the true and the estimated density:

$$R_n(f, \hat{f}_{\mathcal{I}}, d) := E_f[d(f, \hat{f}_{\mathcal{I}}(X_1, \ldots, X_n))].$$

We consider the risks with respect to the following loss functions:

- Squared Hellinger distance

$$d_H(f, g) = \frac{1}{2} \int (\sqrt{f(t)} - \sqrt{g(t)})^2 dt, \tag{3}$$

  which has been normalized such that its maximum value is 1.

- Powers of the L_p-norms (for $p = 1$ and 2) defined by

$$d_p := \|f - g\|_p^p = \int |f(t) - g(t)|^p dt. \tag{4}$$

The L_2 distance is widely used mainly for its mathematical tractability. It is often possible to derive explicit expressions for the L_2 risk at least asymptotically, cf. Kogure (1987). However, as argued by Devroye and Györfi (1985, ch. 1), the L_1 distance can be considered more natural in the context of density estimation because – unlike the L_2 distance – it is defined for all densities and it has desirable invariance properties. We focus mainly on the Hellinger distance for several reasons: it is also defined for any two densities, it has important invariance properties and the results of Castellan (1999, 2000) are derived for the corresponding risk. Another widely used loss function is the Kullback-Leibler distance

$$d_{KL}(f, g) := \int \log\left(\frac{f(t)}{g(t)}\right) f(t) dt$$

which is not suitable in histogram density estimation since it is infinite whenever the estimated density is zero on an interval where the true distribution has positive mass. Hence it is excluded from consideration. For a detailed discussion on the choice of loss functions in histogram density estimation, see section 2.2. in Birgé and Rozenholc (2006) and the references given there.

Given the sample, the histogram $\hat{f}_{\mathcal{I}}$ depends only on the chosen partition $\mathcal{I} = (I_1, \ldots, I_D)$. The values on the intervals of the partition are fixed, namely equal to the relative frequencies divided by the bin widths. In order to achieve good performance in terms of risk, the crucial point is thus choosing the partition. A naïve comparison of the likelihood of histograms for partitions with different numbers of bins is misleading since partitions with too many bins will result in a large likelihood without yielding a sensible estimate of f. But also without any further restrictions on the allowed partitions the likelihood can be made arbitrarily large for a fixed number of bins.

Many approaches exist for the special case of regular histograms where I is divided into D equal sized bins; the problem is then reduced to the choice of D, cf. Birgé and Rozenholc (2006) and Davies et al. (2008) and the references given there. Using irregular partitions can reduce bias and therefore can improve performance for spatially inhomogenous densities, but

the increased difficulty of choosing a good partition may lead to an increase in risk for more well-behaved densities. The idea of constructing both a regular and an irregular histogram and then choosing between the two is briefly discussed in Birgé and Rozenholc (2006). To our knowledge, this approach has not yet been put into practice.

Our recommendation is to construct a regular histogram that maximizes the penalized log-likelihood

$$L(\hat{f}_{\mathcal{I}}, x_1, \ldots, x_n) - (D-1) - \log^{2.5} D \tag{5}$$

among all regular partitions of $[x_{(1)}, x_{(n)}]$ with $D = 1, \ldots, \lfloor n/\log n \rfloor$ bins (where $\lfloor x \rfloor$ denotes the largest integer not larger than x) and an irregular histogram that maximizes the penalized log-likelihood

$$L(\hat{f}_{\mathcal{I}}, x_1, \ldots, x_n) - \log \binom{n-1}{D-1} - (D-1) - \log^{2.5} D \tag{6}$$

among a set of partitions of $[x_{(1)}, x_{(n)}]$ with breakpoints equal to the sample points, where again D is the number of bins in a given partition. The final estimate is then the one with the larger penalized log-likelihood. The penalty in (5) for the regular case was proposed in Birgé and Rozenholc (2006), while the motivation for (6) is developed later in this paper, where we consider different penalty forms for the irregular case.

Several methods have been developed previously to choose a good irregular histogram. Kogure (1987) gives asymptotic results for the optimal choice of bins. His approach is based on using blocks of equisized bins, and the dependence on tuning parameters is explored via simulations in his PhD thesis (Kogure, 1986). It does not result in a fully automatic procedure. Kanazawa (1988) proposes to control the Hellinger distance between the unknown true density and the estimated histogram and introduces a dynamic programming algorithm to find the best partition with a given number of bins. Kanazawa (1992) derives the asymptotically optimal choice of the number of bins. Unfortunately this result involves the first and second derivatives of the unknown density, which leads to a construction that cannot be applied from a practical point of view. Celisse and Robin (2008) give explicit formulas for L_2 leave-p-out cross-validation for regular and irregular histograms. They only briefly comment on the case of irregular histograms and only show simulations with ad-hoc choices of the set of partitions. In our simulations, we use their explicit formula to compare risk behavior of cross-validation and our penalized likelihood approach when both are used to choose an irregular histogram from the same set of partitions. The multiresolution histogram by Engel (1997) is based on a tree of dyadic partitions to control the L_2-error. The performance crucially depends on the finest resolution level, for which no universally usable recommendation is given. Some other tree-based procedures have been suggested for the multivariate case. They can be used for the univariate case, but they either perform a complete search over a restricted set of partitions (Blanchard et al., 2007; Klemelä, 2009) or a greedy search over a full set of partitions (Klemelä, 2007) to deal with computational problems that do not occur in the univariate case. Theoretical results on conditions for consistency of histogram estimates with data-driven and possibly irregular partitions are derived in Chen and Zhao (1987); Zhao et al. (1988) and Lugosi and Nobel (1996). Devroye and Lugosi (2004) give a construction of histograms where bin widths are allowed to vary according to a pre-specified function.

Hartigan (1996) considers regular and irregular histogram construction from a Bayesian point of view. However, we are not aware of any fully tuned automatic Bayesian procedure for irregular histogram construction. Rissanen et al. (1992) give a construction based on the Minimum Description Length (MDL) paradigm, which leads to a penalized likelihood estima-

tor. A choice of several discretization parameters is needed, and the recommendation given by the authors is to perform an exhaustive search over all possible combinations of values, which is computationally expensive. A more recent proposal by Kontkanen and Myllymäki (2007) is also based on the MDL principle; it involves a discretization which results in the estimate not being a proper density. Catoni (2002) suggests a multi-stage procedure based on coding ideas that computes a density estimate by aggregating histograms.

The taut string procedure introduced by Davies and Kovac (2004) can also be used to generate an irregular histogram as described in Davies et al. (2008). Regularization is achieved not by controlling the number of bins but by controlling the modality of the estimate. The stated aim of the authors is not to minimize some risk but to find an estimate of the density that has minimum number of modes that could have generated the data, where the latter is formalized by a criterion based on differences of Kuiper metrics between the empirical and the estimated distribution. The main idea is to construct a piecewise linear spline of minimal length (the taut string) in a tube around the empirical cdf and then take its derivative, which is piecewise constant. The histogram is then constructed using the knots of the string as the boundaries of the bins. This coincides with the derivative of the string except on intervals where the string switches from the upper to the lower boundary of the tube or vice versa. Let us emphasize that although the partition is chosen without reference to maximum likelihood, the histogram constructed in this way fulfils definition (1). The main tuning parameter is the tube width, and an automatic choice is suggested by the authors. The procedure has shown a particularly good behavior also w.r.t. classical loss functions (Davies et al., 2008), and therefore is compared with our method in our simulations.

For our construction of irregular histograms, we will focus on penalized likelihood maximization techniques. For a good data-driven histogram estimate one needs an appropriate penalization to provide an automatic choice of D as well as of the partition $\mathcal{I} = (I_1, \ldots, I_D)$. Since Akaike's Information Criterion (AIC) introduced by Akaike (1973), penalized likelihood has been used with many different penalty terms. AIC aims at ensuring a good risk behavior of the resulting estimate. Another widely used criterion is the Bayesian Information Criterion (BIC) introduced by Schwarz (1978). It is constructed in such a way as to consistently estimate the smallest true model order, which in histogram density estimation would lead to very large models unless the true density is piecewise constant. In practice, criteria like AIC and BIC are routinely applied in many different statistical models, often without reference to their different conceptual backgrounds and without appropriate modifications for the model under consideration. In their original forms, both AIC and BIC do not account for multiple partitions with the same number of bins. See Chapter 7.3 of Massart (2007) for a critique of the use of AIC in histogram density estimation. Since both are widely used, we include them in our comparisons. Our penalties are motivated by recent model selection works by Barron et al. (1999), Castellan (1999, 2000) and Massart (2007). The regular histogram construction proposed in Birgé and Rozenholc (2006) with which we combine our irregular histogram is based on the same ideas.

Our paper is structured as follows: In Section 2, we review the problem of constructing an irregular histogram using penalized likelihood. Section 3 gives a description of the choice of the penalty. Section 4 gives a detailed description of the proposed procedure for irregular histograms. In Section 5, we comment on the empirical evaluation of the risks under consideration. Section 6 gives the results of a simulation study and conclusions.

2. Penalized likelihood construction of irregular histograms

Constructing an irregular histogram by penalized likelihood means maximizing

$$L(\hat{f}_{\mathcal{I}}, x_1, \ldots, x_n) - \mathrm{pen}_n(\mathcal{I}) \tag{7}$$

w.r.t. partitions $\mathcal{I} = (I_1, \ldots, I_{|\mathcal{I}|})$ of $[x_{(1)}, x_{(n)}]$, where $\mathrm{pen}_n(\mathcal{I})$ is a penalty term depending only on the partition $\mathcal{I}$ and possibly on the sample (data-driven). We will introduce a new choice here motivated by work of Barron et al. (1999), Castellan (1999, 2000) and Massart (2007).

Optimizing w.r.t. the partition $\mathcal{I}$ with $|\mathcal{I}|$ fixed in (7) leaves us with a continuous optimization problem. Without further restrictions, for $|\mathcal{I}| \geq 2$ the likelihood is unbounded. The partition

$$\{[x_{(1)}, x_{(1)} + \eta), [x_{(1)} + \eta, x_{(n)}]\}$$

for small η (such that $N_1 = 1$ and $N_2 = n - 1$) leads to a log-likelihood equal to

$$N_1 \log \frac{N_1}{n|I_1|} + N_2 \log \frac{N_2}{n|I_2|} = \log \frac{1}{n\eta} + (n-1) \log \frac{n-1}{n(x_{(n)} - x_{(1)} - \eta)}$$

which can be arbitrarily large when η goes to 0.

One possibility is to restrict to all partitions that are built with endpoints on the observations; the optimization problem (7) can then be solved using a dynamic programming algorithm first used for histogram construction by Kanazawa (1988). More details are given in Section 4.

With $D = |\mathcal{I}|$, we propose the following families of penalties parametrized by two constants c and α:

$$\mathrm{pen}_n^A(\mathcal{I}) = c \log \binom{n-1}{D-1} + \alpha(D-1) + \varepsilon_{c,\alpha}^{(1)}(D), \tag{8}$$

$$\mathrm{pen}_n^B(\mathcal{I}) = c \log \binom{n-1}{D-1} + \alpha(D-1) + \varepsilon^{(2)}(D) \tag{9}$$

$$\tag{10}$$

and

$$\mathrm{pen}_n^R(\mathcal{I}) = c \log \binom{n-1}{D-1} + \frac{\alpha}{n} \sum_{j=1}^{D} \frac{N_j}{|I_j|} + \varepsilon^{(2)}(D). \tag{11}$$

where

$$\varepsilon_{c,\alpha}^{(1)}(D) = c\,k \log D + 2\sqrt{c\alpha(D-1)(\log \binom{n-1}{D-1} + k \log D)} \tag{12}$$

and

$$\varepsilon^{(2)}(D) = \log^{2.5} D. \tag{13}$$

The precise choices for c and α obtained by simulations are described in Section 3. Note that, while the penalties given in (8) and (9) depend only on the number of bins of the partition, the penalty in formula (11) is a *random penalty* in the sense that it also depends on the data.

We now give arguments to explain the origins of these penalties. The penalty defined by (8) is derived from Theorem 3.2 in Castellan (1999), which is also stated as Theorem 7.9 in Massart (2007, p. 232) and from eq. (7.32) in Theorem 7.7 in Massart (2007, p.219). From the penalty form in Theorem 7.9 in Massart (2007) we derive $\varepsilon^{(1)}$:

$$\text{pen}_n(\mathcal{I}) = c_1(\sqrt{D-1} + \sqrt{c_2 x_{\mathcal{I}}})^2, \tag{14}$$

where the weights $x_{\mathcal{I}}$ are chosen such that

$$\sum_D \sum_{|\mathcal{I}|=D} e^{-x_{\mathcal{I}}} \leq \Sigma \tag{15}$$

for an absolute constant Σ. Because the endpoints of our partitions are fixed, there are $\binom{n-1}{D-1}$ different partitions with cardinality D, and we assign equal weights x_D to every partition $\mathcal{I}$ with $|\mathcal{I}| = D$ such that

$$\sum_D \binom{n-1}{D-1} e^{-x_D} \leq \Sigma.$$

To achieve this, we set

$$x_D = \log \binom{n-1}{D-1} + \varepsilon(D).$$

Then (15) becomes

$$\sum_D e^{-\varepsilon(D)} \leq \Sigma.$$

Choosing $\varepsilon(D)$ of the form $k \log D$ with $k > 1$ ensures that the sum is converging and that Σ is finite. Finally for $k > 1$ we have

$$x_D = \log \binom{n-1}{D-1} + k \log D.$$

Substitution into (14) gives

$$\begin{aligned} \text{pen}_n(\mathcal{I}) \quad = \quad & c_1 \left(D - 1 + c_2 \left(\log \binom{n-1}{D-1} + k \log D \right) \right. \\ & \left. + 2\sqrt{c_2(D-1)\left(\log \binom{n-1}{D-1} + k \log D \right)} \right). \end{aligned} \tag{16}$$

Let us emphasize that Theorem 7.9 in Massart (2007, p. 232) requires $c_1 > 1/2$ and $c_2 = 2(1 + 1/c_1)$. Coming back to our notations, with $\alpha = c_1$, $c = c_1 c_2$ we obtain Equation (12).

We now use Theorem 7.7 in Massart (2007, p. 219) to justify the random penalty in (11). The orthonormal basis considered in this theorem for a given partition $\mathcal{I}$ consists of all $\mathbf{1}_I / \sqrt{|I|}$ for all I in $\mathcal{I}$. The least squares contrast used in this theorem in our framework is $-n^{-2} \sum_{I \in \mathcal{I}} N_I^2 / |I|$. To link the minimization of the least squares contrast and the maximization of the log-likelihood, we consider the following approximation:

$$L(\hat{f}_{\mathcal{I}}, x_1, \ldots, x_n) = \sum_{j=1}^{D} N_j \log \left(\frac{N_j}{n|I_j|} \right) \approx \sum_{j=1}^{D} N_j \left(\frac{N_j}{n|I_j|} - 1 \right) = \frac{1}{n} \sum_{j=1}^{D} \frac{N_j^2}{|I_j|} - n.$$

From the penalty form (7.32) and the use of $M = 1$ and $\varepsilon = 0$ in Theorem 7.7 in Massart (2007, p. 219), following the same derivation as for $\varepsilon^{(1)}$, we find the penalty in (16) with $c_1 = 1$ and $c_2 = 2$.

Using the least squares approximation, we can use the random penalty (7.33) in Theorem 7.7 in Massart (2007). Let us emphasize that $\widehat{V}_m$ defined by Massart is in our framework $\sum_{I\in\mathcal{I}} N_I/n|I|$ with $m = \mathcal{I}$. To derive $\varepsilon^{(2)}$ in (11) we start from the penalty defined in (7.33) in Massart (2007):

$$\mathrm{pen}_n(\mathcal{I}) = (1+\varepsilon)^5 \left(\sqrt{\widehat{V}_\mathcal{I}} + \sqrt{2ML_\mathcal{I}D}\right)^2.$$

Following the same derivations as for the penalty (14), setting $M = 1$, $\varepsilon = 0$ and $L_\mathcal{I} = D^{-1}(\log\binom{n-1}{D-1} + k\log D)$ we obtain:

$$\begin{aligned}\mathrm{pen}_n(\mathcal{I}) &= \widehat{V}_\mathcal{I} + 2\log\binom{n-1}{D-1} + 2k\log D \\ &\quad +2\sqrt{2\widehat{V}_\mathcal{I}\left(\log\binom{n-1}{D-1} + k\log D\right)}.\end{aligned}$$

Let us emphasize that, because of terms of the form $\varphi(D)\widehat{V}_\mathcal{I}$, the expression in the square root above prevents the use of dynamic programming to compute the maximum of the penalized log-likelihood defined in (7). To avoid this problem we propose, following penalty forms proposed in Birgé and Rozenholc (2006) and Comte and Rozenholc (2004), to replace the remainder expression

$$2k\log D + 2\sqrt{2\widehat{V}_\mathcal{I}\left(\log\binom{n-1}{D-1} + k\log D\right)}$$

by a power of $\log D$. We have tried several values of the power and found that Formula (13) leads to a good choice. Finally, we also replaced $\varepsilon^{(1)}_{c,\alpha}$ in formula (8) by $\varepsilon^{(2)}$, leading to the penalty given in (9).

3. Choice of the Penalty

Using histograms with the endpoints of the partition placed on the observations as described later in Section 4, we ran empirical risk estimation in order to fix our penalty using the losses defined by (3) and (4) for $p = 1$ and 2. We used the same densities for calibration as in the simulations described in Section 6 but different samples and a smaller number of replications. We focused on the Hellinger risk to obtain good choices of the penalties, but the behavior w.r.t. L_1 loss is very similar. For minimizing L_2 risk, other choices may be preferable. Since no single penalty is best in all cases, the calibration of a penalty always leads to some compromise. We describe in the following what we consider to be a good proposal.

In formula (8) we tried:

- $c = 2(\alpha+1)$ and $\alpha \in \{0.5, 1\}$ following Theorem 7.9 in Massart (2007).
- $c = 2$ and $\alpha = 1$ following Theorem 7.7 eq. (7.32) in Massart (2007) with $M = 1$ and $\varepsilon = 0$.

- $c = 1$ and $\alpha \in \{0.5, 1\}$.

We always set $k = 2$. From these experiments, the most satisfactory choice is $c = 1$ and $\alpha = 0.5$. We also ran experiments replacing $\varepsilon^{(1)}_{c,\alpha}$ by $\varepsilon^{(2)}$, leading to the penalty given in (9). In this case, we have found that the most satisfactory choice is $c = 1$ and $\alpha = 1$, and this choice is even better than $\varepsilon^{(1)}_{2,1}$. Note that the resulting penalty, given in (6), exactly corresponds to the penalty in (5) proposed in Birgé and Rozenholc (2006) for the regular case, except for the additional term $\log \binom{n-1}{D-1}$ that is needed to account for multiple partitions with the same number of bins. This term is zero for a histogram with just one bin, so the penalized likelihoods for regular and irregular histograms can directly be compared in this case. Because (8) and (9) are very similar, we only use this version in our simulations in Section 6.

For the random penalty in formula (11) we ran risk evaluation experiments using all combinations of $c \in \{0.5, 1, 2\}$ and $\alpha \in \{0.5, 1\}$. Let us emphasize that $c = 2$ and $\alpha = 1$ correspond to formula (7.33) in Massart (2007) up to our choice of $\varepsilon^{(2)}$ defined in (13). From our point of view, the most satisfactory choice is $c = 1$ and $\alpha = 0.5$. When comparing the log-likelihood penalized in this way to the one in (5), care has to be taken to ensure that both give the same value for a histogram with just one bin.

To conclude this section, we remark that the results are very close. Only for the trimodal uniform density have we found differences in favor of the penalty (9). For all other densities, the absolute values of the relative differences $\left|\frac{\widehat{R}^{\mathrm{R}}_n - \widehat{R}^{\mathrm{D}}_n}{\widehat{R}^{\mathrm{D}}_n}\right|$ of the risks are less than 0.163.

4. Algorithm for constructing irregular histograms

We maximize (7) w.r.t. partitions $\mathcal{I}$ built with endpoints on the observations:

$$\mathcal{I} = ([x_{(1)}, x_{(k_1)}], (x_{(k_1)}, x_{(k_2)}], (x_{(k_2)}, x_{(k_3)}], \ldots, (x_{(k_{D-2})}, x_{(k_{D-1})}], (x_{(k_{D-1})}, x_{(n)}]),$$

where $1 < k_1 < \ldots < k_{D-1} < n$. We start from a "finest" partition $\mathcal{I}_{\max}$ defined by $D_{\max} < n$ and the choice $1 < k_1 < \ldots < k_{D_{\max}-1} < n$. Let us write this partition as

$$\mathcal{I}_{\max} = (I^0_1, \ldots, I^0_{D_{\max}}),$$

where $I^0_d = (t_{d-1}, t_d]$ for $d = 1$ to $D_{\max}$ and where $t_0 = x_{(1)} - \mathrm{eps}$, $t_{D_{\max}} = x_{(n)}$ and $t_d = x_{(k_d)}$ for $0 < d < D_{\max}$. Here eps represents the machine precision and is used only to allow for the use of left-open, right-closed intervals. Our aim is to build a sub-partition $\mathcal{I}$ of $\mathcal{I}_{\max}$ which maximizes (7). This problem is solved in polynomial time by a dynamic programming (DP) algorithm as used in Kanazawa (1988) and Comte and Rozenholc (2004). We briefly describe the algorithm in our context of penalized histograms. Let us assume that, given the sample, (7) can be rewritten as $\Phi^0(\mathcal{I}) + \Psi(D, n)$, where Φ^0 is an additive function with respect to the partition in the sense that

$$\Phi^0(\mathcal{I}) = \Phi(I_1) + \ldots + \Phi(I_D) \text{ if } \mathcal{I} = (I_1, \ldots, I_D).$$

In our case, $\Phi(I)$ depends only on the number N_I of observations in interval I and on its length $|I|$. More precisely, for a penalty of the form (8) or (9), we have

$$\Phi(I) = N_I \log \frac{N_I}{n|I|}, \tag{17}$$

and $\Psi(D,n) = \mathrm{pen}_n^A(\mathcal{I})$ or $\Psi(D,n) = \mathrm{pen}_n^B(\mathcal{I})$, respectively. For a penalty of the form (11) we have

$$\Phi(I) = N_I \log \frac{N_I}{n|I|} - \frac{\alpha}{n}\frac{N_I}{|I|},$$

and $\Psi(D,n) = c \log \binom{n-1}{D-1} + \varepsilon^{(2)}(D)$.

We denote $p_1(i,j) = \Phi((t_i,t_j])$ and $p_1(j) := p_1(0,j)$. Finally, let us define $i_1(j) = 0$. Assume that we have already computed all $p_1(i,j)$ for $0 \le i < j \le D_{max}$ (which needs $O(D_{max}^2)$ operations). The dynamic programming algorithm works as follows. First, the maxima of Φ_0 for partitions with $D = 1, \ldots, D_{max}$ bins are calculated:

- `For` $D = 2 \ldots D_{\max}$
  - `For` $j = D \ldots D_{\max}$,
    - $i_D(j) = \arg_i \max_{D-1 \le i < j}[p_{D-1}(i) + p_1(i,j)]$;
    - $p_D(j) = p_{D-1}(i_D(j)) + p_1(i_D(j), j)$

For each D, $p_D(D_{\max})$ is the maximum of $\Phi^0(\mathcal{I})$ for all sub-partitions $\mathcal{I}$ of our finest partition $\mathcal{I}_{\max}$ with D bins. The partition which achieves the maximum of $\Phi^0(\mathcal{I}) + \Psi(D,n)$ can be constructed in the following way:

- `Compute` $\widehat{D} = \arg_D \max_{1 \le D \le D_{\max}} p_D(D_{\max}) + \Psi(D,n)$.
- `Initialize` $L = D_{max}$
- `For` $j = \widehat{D}, \ldots, 1$: $\quad L = c(i_j(L[1]), L)$

The vector L defines the indices of the t_j's which are the endpoints of the best partition in the sense of (7). The notation $L[1]$ denotes the first coordinate of the vector L and $c(u.L)$ denotes concatenation of the vector u with L.

The computation of $i_D(j) = \arg_i \max_{D-1 \le i < j}[p_{D-1}(i) + p_1(i,j)]$ requires $O(j - D + 1)$ operations and the total complexity of this algorithm is of order $D_{\max}^3$. Hence the total number of operations may be of order n if we start from a finest partition with $D_{\max}$ of order $n^{1/3}$. We propose to use a greedy algorithm in order to build a finest partition with $\lfloor \max\{n^{1/3}, 100\} \rfloor$ bins if this number is smaller than n. Let us call $\mathcal{E}(\mathcal{I})$ the set of endpoints of partition $\mathcal{I}$. Starting with the partition $\mathcal{I}_0 = ([x_{(1)}, x_{(n)}])$, we produce a sequence of partitions $\mathcal{I}_D$ satisfying :

$$\mathcal{I}_{D+1} = \arg\max \Phi^0(\mathcal{I}),$$

where the maximum is taken over all partitions $\mathcal{I}$ with $\mathcal{E}(\mathcal{I}) = \mathcal{E}(\mathcal{I}_D) \cup \{t\}$ with t in $\{x_1, \ldots, x_n\} \setminus \mathcal{E}(\mathcal{I}_D)$. For all three penalty forms, we use a greedy maximization of the likelihood to obtain this partition, i.e. we always use Φ as in (17).

Let us remark that the theoretical results by Castellan (1999, 2000) and Massart (2007, ch. 7), are derived for the case of a finest regular grid with bin sizes not smaller than a constant times $\log^2(n)/n$. In particular, the set of partitions is fixed beforehand and may depend on n but not on the sample. This also means that that no bins are possible that are shorter than a constant times $\log^2(n)/n$. However, we found that, in practice, we can improve performance drastically for some densities by using a data-dependent finest grid imposing no restrictions on the smallest bins without losing much at other densities. More comments on this are given in Section 6.

5. Risk evaluation

The risks of the procedures are evaluated empirically by means of simulations. For each density f and each sample size n, N samples $x^{(j)} := (x_1^{(j)}, \ldots, x_n^{(j)})$, $j = 1, \ldots, N$ are generated and the loss functions $d = d_H, d_1, d_2$ are evaluated for every histogram procedure $\hat{f}$. We estimate the risks $R_n(f, \hat{f}, d)$ by

$$\widehat{R}_n(f, \hat{f}, d) := \frac{1}{N} \sum_{j=1}^{N} d(f, \hat{f}(x_1^{(j)}, \ldots, x_n^{(j)})).$$

We now describe how we computed our loss functions (3) and (4) to obtain empirical risk evaluation. To estimate the risks, we evaluate the losses $d(f, \hat{f}(x_1^{(j)}, \ldots, x_n^{(j)}))$ for every simulation run j by numerical integration. First note that the integrals appearing in (3) and (4) are all of the form

$$\int \delta(t)dt := \int \tilde{\delta}(f(t), g(t))dt$$

for continuous functions $\tilde{\delta}$. Care has to be taken of discontinuities in both the true densities f and the histogram estimates $\hat{f}$ and furthermore the bilogarithmic peak density has infinite peaks. For given f and $\hat{f}$, let $\tau_1 < \cdots < \tau_{L-1}$ denote the points where f or $\hat{f}$ is discontinuous or infinite. Defining the intervals $J_0 := (-\infty, \tau_1)$, $J_l := (\tau_l, \tau_{l+1})$, $l = 1, \ldots, L-1$, $J_L := (\tau_L, \infty)$, we split up the integrals into sums of integrals over open intervals where δ is continuous:

$$\int_{\mathbb{R}} \delta(t)dt := \sum_{l=0}^{L} \int_{J_l} \delta(t)dt.$$

Note that we use open intervals to allow both f and $\hat{f}$ to take any (possibly infinite) value in the point $\tau_1, \ldots, \tau_l$. To evaluate the integrals on $J = J_1, \ldots, J_{L-1}$ we use the trapeze rule

$$\int_{J_l} \delta(t)dt \approx (\kappa_K^l - \kappa_1^l) \left(\frac{1}{2}\delta(\kappa_1^l) + \delta(\kappa_2^l) + \cdots + \delta(\kappa_{K-1}^l) + \frac{1}{2}\delta(\kappa_K^l) \right)$$

for equispaced grid points $\kappa_1^l = \tau_l + \varepsilon$, $\kappa_K^l = \tau_{l+1} - \varepsilon$ and $\kappa_\nu^l = \kappa_1^l + (\nu - 1)h$ for $\nu = 2, \ldots, K-1$ with $h = \frac{\tau_{l+1} - \tau_l - 2\varepsilon}{K-1}$. We set $\varepsilon = 10^{-11}$ to integrate over open intervals.

Note that on the unbounded intervals J_0 and J_L for $d = d_H$ and $d = d_1$ we have $\int_{J_l} \delta(t)dt = \int_{J_l} f(t)dt$ since $\hat{f}$ is zero. For d_2 we replace $\pm\infty$ in the definition of J_0 and J_L by the upper and lower 10^{-10}-quantiles of f and integrate numerically as on the other intervals, in case the support of f is unbounded. Otherwise, the integrals over J_0 and J_L are zero.

6. Simulation Study and Conclusions

In order to choose the constants in the penalties given in Section 3 and to assess the performance of the penalized likelihood histogram defined as the maximizer of (7) with penalty defined by (8)-(11), we conduct a simulation study involving empirical risk estimation with respect to the losses (3) and (4) (for p=1,2). The choices we arrive at are given in Section 3. Then we compare our choices for the penalized maximum likelihood to other available methods in a separate simulation study using the same densities.

The performance of the methods is compared for 12 of the 28 test-bed densities introduced by Berlinet and Devroye (1994). We used densities 1 (uniform), 4 (double exponential), 11 (normal), 12 (lognormal), 21-24 (mixtures of normals) and 25-28 (various other multimodal densities). We denote these by $f_1, \ldots, f_{12}$. We also added 4 histogram densities:

- 5 bin regular histogram:

$$\begin{aligned} f_{13}(x) &:= 0.15u_{[0,0.2]}(x) + 0.35u_{(0.2,0.4]}(x) + 0.2u_{(0.4,0.6]}(x) \\ &\quad +0.1u_{(0.6,0.8]}(x) + 0.2u_{(0.8,1.0]}(x) \end{aligned}$$

- 5 bin irregular histogram:

$$\begin{aligned} f_{14}(x) &:= 0.15u_{[0,0.13]}(x) + 0.35u_{(0.13,0.34]}(x) + 0.2u_{(0.34,0.61]}(x) \\ &\quad +0.1u_{(0.61,0.65]}(x) + 0.2u_{(0.65,1.0]}(x) \end{aligned}$$

- 10 bin regular histogram:

$$\begin{aligned} f_{15}(x) &:= 0.01u_{[0,0.1]}(x) + 0.18u_{(0.1,0.2]}(x) + 0.16u_{(0.2,0.3]}(x) \\ &\quad +0.07u_{(0.3,0.4]}(x) + 0.06u_{(0.4,0.5]}(x) + 0.01u_{(0.5,0.6]}(x) \\ &\quad +0.06u_{(0.6,0.7]}(x) + 0.37u_{(0.7,0.8]}(x) + 0.06u_{(0.8,0.9]}(x) \\ &\quad +0.02u_{(0.9,1.0]}(x) \end{aligned}$$

- 10 bin irregular histogram:

$$\begin{aligned} f_{16}(x) &:= 0.01u_{[0,0.02]}(x) + 0.18u_{(0.02,0.07]}(x) + 0.16u_{(0.07,0.14]}(x) \\ &\quad +0.07u_{(0.14,0.44]}(x) + 0.06u_{(0.44,0.53]}(x) + 0.01u_{(0.53,0.56]}(x) \\ &\quad +0.06u_{(0.56,0.67]}(x) + 0.37u_{(0.67,0.77]}(x) + 0.06u_{(0.77,0.91]}(x) \\ &\quad +0.02u_{(0.91,1.0]}(x) \end{aligned}$$

where $u_I := |I|^{-1}\mathbf{1}_I$ denotes the uniform density on an interval I. All densities are implemented in the `R`-package `benchden` (Mildenberger et al., 2009b) and are depicted in Figure 1. Note that Castellan's main theorem 3.2 in Castellan (1999) does not apply to all densities considered here, since she assumes for instance that the density is bounded away from zero. We include a wide range of densities in order to explore the behavior of the procedure in cases not covered by theory. The sample sizes are 50,100,500,1000,5000 and 10000. We use 500 replications for each scenario and estimate the resulting risks as described in Section 5 using $\kappa = 5000$.

The methods compared in the simulations are:

- **B** Penalized maximum likelihood using penalty (9) with $c = 1$ and $\alpha = 1$.

  R Penalized maximum likelihood using random penalty (11) with $c = 1$ and $\alpha = 0.5$.

  CV Leave-one-out cross-validation using formula (11) given in Celisse and Robin (2008). We also tried formula (12) of Celisse and Robin (2008) for different values of p without finding a big difference.

  Maximization is performed over a data-driven finest grid as described in Section 3 without restrictions on the minimum bin width.

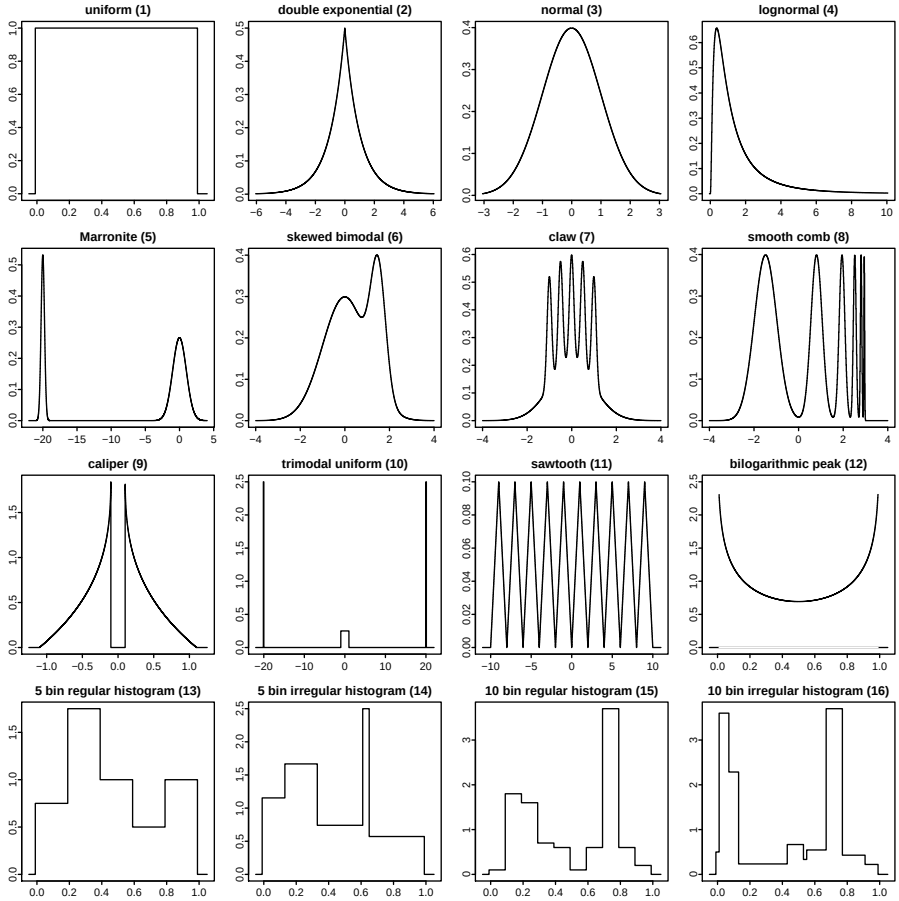


Figure 1: The densities used in the simulation study.

- Methods $\mathbf{B}, \mathbf{R}, \mathbf{CV}$ using the same data-driven grid but with the additional constraint that the minimum bin length allowed is $(x_{(n)} - x_{(1)}) \log^{1.5}(n)/n$. These are denoted by $\mathbf{B}'$, $\mathbf{R}'$ and $\mathbf{CV}'$.

- Methods $\mathbf{B}, \mathbf{R}, \mathbf{CV}$ but using a full optimization over a finest regular partition with bin width $(x_{(n)} - x_{(1)}) \lfloor n/\log^{1.5}(n) \rfloor^{-1}$. This is the grid considered in Castellan (1999), except that we slightly relax $\log^2(n)$ to $\log^{1.5}(n)$. These are denoted $\mathbf{B}''$, $\mathbf{R}''$ and $\mathbf{CV}''$.

- Penalized maximum likelihood using Information Criteria:

  AIC Akaike's Information Criterion (Akaike, 1973). The penalty is $\text{pen}_n^{\text{AIC}}(D) = (D - 1)$.

  BIC Bayesian Information Criterion (Schwarz, 1978). The penalty is $\text{pen}_n^{\text{BIC}}(D) = 0.5 \log(n)(D - 1)$.

- The taut-string method **TS** introduced by Davies and Kovac (2004). We use the function `pmden()` implemented in the R-package `ftnonpar` (Davies and Kovac, 2008) with the default values except that we set `localsq=FALSE` as local squeezing of the tube does not give a ML histogram. The histogram is then constructed using the knots of the string

as the boundaries of the bins. This coincides with the derivative of the string except on intervals where the string switches from the upper to the lower boundary of the tube or vice versa.

- Regular histogram construction **BR** due to Birgé and Rozenholc (2006). The penalty is $\text{pen}_n^{\text{BR}}(D) = D + \log(D)^{2.5}$, where the log-likelihood is maximized over all regular partitions with $1, \ldots, \lfloor n/\log n \rfloor$ bins.

- Combined approach: A regular histogram and an irregular histogram are constructed. After adding a constant such that all penalties are 0 for a histogram with 1 bin, the penalized log-likelihoods are compared. The histogram that has the larger value is chosen as the final estimate:

  B* Chooses between a regular histogram using **BR** and an irregular histogram using **B**

  R* Chooses between a regular histogram using **BR** and an irregular histogram using **R**.

Except for **TS**, all methods have been implemented in the `R`-package `histogram` (Mildenberger et al., 2009a). The program code to reproduce all simulation results will be available from the corresponding author's website. For the discussion of the results, we focus on squared Hellinger risk. Table 1 shows the binary logarithms of relative squared Hellinger risks w.r.t. the best method for any given n and density: $\log_2(\widehat{R}_n^{\text{method}}/\widehat{R}_n^{\text{best}})$ for all 15 methods in the simulation study. Thus, a value of 0 means that the method was best in this particular setting and a value of 1 means that the risk of the method is twice as large as the risk of the best method. A table giving the absolute risk values as well as tables for the other risks can be found in the appendix. Generally, the L_1 and Hellinger risks behave similarly, while the results for L_2 are much less conclusive and no clear recommendation for using one of the methods under consideration can be given when minimizing the L_2 risk is the main aim. Table 8 in the appendix shows the modes of the number of bins chosen for all methods as well as the corresponding frequencies with which this number was chosen.

As was to be expected, the table shows no clear overall best method for all scenarios. The relative risk w.r.t. the best method is always smaller than 2 for our proposal **B***. In this sense, it can be seen as the best method. Let us remark that other methods could be better in particular situations. In many cases, **TS** or one of our proposals **B** and **R** is either the best or the binary logarithm of relative risk w.r.t. the best method is close to zero. These three methods are – except for **B*** – also the only ones in the simulation study for which this quantity is always strictly smaller than $\log_2 3 \approx 1.58$, meaning that the empirical risk is never greater than three times the risk achieved by the best method. The random penalty **R** seems to be slightly better than **B** in many cases, the most notable exception being the trimodal uniform density f_{10} for $n = 50$.

Cross-validation using the same set of partitions (i.e. a data-driven finest grid without further restrictions on minimum bin width) performs rather poorly, especially when the underlying density is a histogram (f_1, f_{10}, f_{13}-f_{16}) or when it has infinite peaks (f_{12}). Note that it is particularly bad for the uniform, which can be a major problem in many applications like grey level estimation of image differences. Relative performance of **CV** w.r.t. the best method becomes generally worse when sample size increases. If we compare the random and deterministic penalties and cross-validation for the case of full dynamic programming

	n	**B**	**R**	**CV**	**B′**	**R′**	**CV′**	**B″**	**R″**	**CV″**	**AIC**	**BIC**	**TS**	**BR**	**B***	**R***
f_1	50	0.13	0.05	1.53	0.03	0.03	0.76	0	0	0.44	2.07	1.27	0.07	0.12	0.18	0.13
	100	0.08	0.02	2.31	0.02	0.02	1.25	0	0.01	0.72	2.91	1.56	0.1	0.11	0.17	0.12
	500	0.06	0.03	4.2	0.01	0.02	2.56	0	0.01	1.69	4.64	2.04	0.19	0.13	0.17	0.14
	1000	0.04	0.02	4.64	0.01	0.01	3.05	0	0.01	2.3	4.98	2.1	0.12	0.11	0.14	0.12
	5000	0.01	0	5.14	0	0	3.68	0	0	3.83	5.38	1.89	0.12	0.12	0.13	0.12
	10000	0	0	5.39	0	0	3.98	0	0	4.71	5.59	1.85	0.2	0.18	0.19	0.18
f_2	50	0.42	0.34	0.5	0.38	0.34	0.09	0.52	0.47	0	0.91	0.53	0.13	0.14	0.17	0.16
	100	0.57	0.51	0.81	0.56	0.51	0.05	0.56	0.54	0	1.34	0.61	0.02	0.2	0.24	0.25
	500	0.71	0.67	1.79	0.7	0.66	0.04	0.69	0.67	0.07	2.2	0.75	0	0.2	0.2	0.2
	1000	0.74	0.7	1.94	0.72	0.7	0.08	0.72	0.7	0.15	2.25	0.73	0	0.18	0.18	0.18
	5000	0.74	0.72	1.78	0.73	0.71	0.23	0.71	0.7	0.47	1.94	0.62	0	0.11	0.11	0.11
	10000	0.69	0.71	1.59	0.69	0.71	0.21	0.67	0.68	0.68	1.7	0.56	0	0.07	0.07	0.07
f_3	50	0.51	0.42	0.68	0.49	0.43	0.14	0.51	0.48	0	1.15	0.64	0.51	0.23	0.24	0.23
	100	0.47	0.4	1.07	0.45	0.39	0.17	0.48	0.43	0	1.62	0.72	0.25	0.15	0.16	0.17
	500	0.65	0.6	2.08	0.64	0.6	0.36	0.65	0.59	0.1	2.49	0.81	0.13	0	0	0
	1000	0.73	0.69	2.28	0.73	0.69	0.54	0.73	0.66	0.25	2.59	0.84	0.19	0	0	0
	5000	1.01	0.95	2.31	1	0.95	0.93	0.97	0.92	0.98	2.49	0.93	0.47	0	0	0
	10000	1.03	0.99	2.19	1.02	0.99	0.94	1	0.96	1.34	2.32	0.93	0.57	0	0	0
f_4	50	0.56	0.5	0.68	0.74	0.72	0.45	0.68	0.65	0.41	1.12	0.67	0	0.55	0.56	0.52
	100	0.63	0.6	1.08	0.72	0.72	0.38	0.7	0.69	0.38	1.62	0.79	0	0.61	0.62	0.59
	500	0.56	0.61	1.79	0.71	0.72	0.36	0.66	0.69	0.41	2.2	0.62	0	0.67	0.55	0.62
	1000	0.31	0.42	1.74	0.72	0.74	0.4	0.7	0.73	0.46	2.05	0.39	0	0.75	0.34	0.52
	5000	0.62	0.82	1.84	1.56	1.63	1.35	1.34	1.43	1.26	2	0.55	0	1.99	0.62	0.82
	10000	0.55	0.77	1.62	1.48	1.56	1.27	1.27	1.37	1.25	1.73	0.45	0	2.05	0.55	0.77
f_5	50	0.25	0.19	0.5	2.18	2.18	2.18	1.6	1.59	1.57	0.64	0.33	0	1.12	0.25	0.19
	100	0.47	0.44	0.76	2.71	2.71	2.73	1.81	1.82	1.77	1.07	0.48	0	0.41	0.47	0.44
	500	0.79	0.77	1.73	1.69	1.69	1.92	1.89	1.9	1.68	2.05	0.81	0	1.18	0.79	0.77
	1000	0.89	0.91	1.92	1.56	1.56	1.78	1.29	1.3	0.98	2.16	0.85	0	0.95	0.89	0.91
	5000	0.97	1.01	1.8	1.04	1.06	0.66	0.94	0.99	0.22	1.93	0.85	0	0.76	0.97	1.01
	10000	0.98	1	1.63	0.98	1	0.47	0.96	1	0.22	1.71	0.82	0	0.68	0.98	1
f_6	50	0.4	0.31	0.75	0.38	0.32	0.24	0.33	0.3	0	1.22	0.66	0.51	0.28	0.3	0.27
	100	0.36	0.29	1.14	0.35	0.3	0.31	0.45	0.39	0	1.7	0.8	0.44	0.31	0.31	0.29
	500	0.5	0.45	1.98	0.5	0.45	0.31	0.48	0.42	0	2.38	0.71	0.1	0.09	0.09	0.1
	1000	0.55	0.49	2.11	0.54	0.49	0.42	0.52	0.47	0.11	2.42	0.67	0.06	0	0	0.01
	5000	0.88	0.82	2.24	0.88	0.82	0.84	0.83	0.78	0.92	2.43	0.84	0	0.15	0.15	0.15
	10000	0.92	0.87	2.1	0.91	0.87	0.85	0.91	0.86	1.26	2.23	0.83	0	0.11	0.11	0.11
f_7	50	0.28	0.16	0.48	0.25	0.17	0.04	0.39	0.35	0	0.86	0.45	0.39	0.19	0.2	0.15
	100	0.05	0.01	0.66	0.02	0	0.08	0.2	0.18	0.05	1.15	0.53	0.06	0.14	0.12	0.08
	500	0.68	0.64	1.44	0.66	0.63	0.14	0.69	0.66	0.15	1.84	0.68	0	0.27	0.28	0.3
	1000	0.89	0.87	1.8	0.88	0.86	0.44	1.15	1.14	0.29	2.1	0.8	0	0.45	0.45	0.45
	5000	1.16	1.13	1.93	1.15	1.13	0.68	1.13	1.1	0.82	2.07	1.07	0	0.74	0.74	0.74
	10000	1.25	1.24	1.85	1.25	1.24	0.76	1.26	1.26	1.11	1.94	1.11	0	0.77	0.77	0.77
f_8	50	0.3	0.24	0.08	0.29	0.26	0.05	0.34	0.32	0	0.4	0.21	0.22	0.14	0.16	0.14
	100	0.37	0.28	0.27	0.37	0.3	0	0.63	0.55	0.09	0.67	0.27	0.17	0.09	0.09	0.09
	500	0.44	0.36	0.9	0.51	0.46	0.24	0.66	0.59	0	1.25	0.35	0.03	0.1	0.1	0.1
	1000	0.47	0.39	1.05	0.54	0.49	0.27	0.64	0.58	0.09	1.33	0.37	0	0.19	0.19	0.19
	5000	0.81	0.75	1.18	0.87	0.82	0.52	0.86	0.8	0.41	1.31	0.69	0	0.46	0.46	0.46
	10000	0.99	0.93	1.17	1.01	0.96	0.61	0.99	0.93	0.66	1.25	0.82	0	0.7	0.7	0.7
f_9	50	0.44	0.42	0	0.45	0.45	0.16	0.44	0.44	0.36	0.34	0.03	0.26	0.36	0.38	0.37
	100	0.46	0.28	0.72	0.49	0.28	0	1.37	1.34	1.09	1.13	0.47	0.17	0.6	0.51	0.37
	500	0.55	0.5	1.81	0.52	0.48	0.36	1.64	1.62	1.36	2.17	0.68	0	0.85	0.63	0.55
	1000	0.61	0.56	2.02	0.59	0.55	0.5	1.65	1.62	1.37	2.3	0.72	0	1.08	0.66	0.61
	5000	0.74	0.7	2.02	0.73	0.7	0.76	1.66	1.64	1.58	2.17	0.72	0	1.23	0.77	0.74
	10000	0.73	0.69	1.87	0.72	0.69	0.74	1.61	1.59	1.69	1.98	0.65	0	1.31	0.77	0.74
f_{10}	50	0	1.31	0.55	3.15	3.15	3.14	3.12	3.12	3.12	0.68	0.31	0.15	2.68	0	2.68
	100	0.06	0	1.02	4.2	4.2	4.2	4.02	4.02	4.02	1.32	0.53	0.32	3.29	0.06	0
	500	0.02	0	2.4	5.82	5.82	5.82	5.64	5.64	5.64	2.74	0.79	0.51	5.08	0.02	0
	1000	0.01	0	2.8	6.8	6.8	6.8	6.34	6.34	6.34	3.07	0.82	0.56	5.63	0.01	0
	5000	0	0	3.33	9.09	9.09	9.09	7.57	7.57	7.57	3.49	0.74	0.73	0.17	0	0
	10000	0	0	3.48	10.1	10.1	10.1	6.46	6.46	6.46	3.59	0.65	0.81	0.17	0	0
f_{11}	50	0.06	0.02	0.54	0.01	0.01	0.22	0	0	0.13	0.81	0.51	0.19	0.03	0.08	0.04
	100	0.02	0.01	0.39	0	0	0.24	0	0	0.13	0.74	0.43	0.25	0.03	0.05	0.03
	500	0.86	0.61	0.8	0.89	0.7	0.27	1.06	1.06	0.15	1.14	0.48	0.11	0	0	0
	1000	0.63	0.58	0.98	0.62	0.59	0.4	0.86	0.69	0.08	1.24	0.63	0.09	0	0	0
	5000	0.97	0.9	1	0.96	0.9	0.45	0.98	0.9	0.24	1.08	0.84	0.13	0	0	0
	10000	1.06	0.99	0.93	1.05	0.98	0.57	1.09	1.01	0.37	0.97	0.91	0.17	0	0	0
f_{12}	50	0.21	0.11	0.88	0.01	0.02	0.25	0	0.01	0.14	1.36	0.79	0.03	0.15	0.27	0.18
	100	0.23	0.17	1.17	0.13	0.14	0.27	0.12	0.13	0	1.75	0.73	0.15	0.16	0.24	0.18
	500	0.37	0.33	2.41	0.3	0.3	0.88	0.29	0.28	0.28	2.83	0.9	0	0.29	0.37	0.34
	1000	0.52	0.49	2.63	0.48	0.47	1.09	0.52	0.51	0.5	2.95	0.89	0	0.36	0.45	0.42
	5000	0.65	0.62	2.61	0.64	0.62	1.24	0.63	0.62	1.22	2.8	0.8	0	0.52	0.56	0.56
	10000	0.68	0.66	2.48	0.67	0.66	1.18	0.67	0.65	1.61	2.62	0.74	0	0.57	0.6	0.59
f_{13}	50	0.1	0.07	0.69	0.04	0.05	0.02	0.01	0.02	0	1.16	0.65	0.06	0.06	0.09	0.06
	100	0.43	0.4	1.21	0.39	0.38	0.25	0.38	0.38	0.04	1.77	0.87	0.4	0	0.01	0.01
	500	1.41	1.34	3.43	1.39	1.33	1.86	1.53	1.5	1.41	3.85	1.88	1.1	0	0	0
	1000	1.06	0.98	3.98	1.02	0.98	2.44	0.83	0.78	1.7	4.29	1.97	1.11	0	0	0
	5000	0.91	0.9	4.67	0.9	0.9	3.32	1.2	1.2	3.35	4.87	1.96	1.13	0	0	0
	10000	0.92	0.92	4.83	0.9	0.91	3.57	1.39	1.39	4.1	5	1.93	1.25	0	0	0
f_{14}	50	0.18	0.13	0.65	0.12	0.11	0.14	0.12	0.13	0	1.1	0.59	0.19	0.02	0.08	0.04
	100	0.26	0.23	1	0.24	0.23	0.3	0.36	0.33	0.19	1.54	0.65	0.29	0	0.03	0.02
	500	0.08	0.03	2.41	0.02	0	0.85	0.58	0.56	0.67	2.83	0.83	0.09	0.88	0.35	0.28
	1000	0.14	0.11	2.98	0.12	0.11	1.45	0.83	0.82	1.25	3.3	0.98	0	0.94	0.21	0.15
	5000	0.01	0	3.76	0	0	2.37	1.42	1.42	2.68	3.97	1.07	0.39	1.9	0.01	0
	10000	0.01	0.01	3.96	0	0	2.67	0.92	0.93	3.31	4.13	1.04	0.57	2.64	0.01	0.01
f_{15}	50	0.1	0	0.06	0.35	0.29	0.06	0.46	0.42	0.06	0.43	0.09	0.11	0.03	0.04	0.01
	100	0.27	0.18	0.65	0.24	0.21	0	0.69	0.65	0.35	1.11	0.46	0.1	0.27	0.25	0.2
	500	0.15	0	1.79	0.15	0.01	0.31	1.15	1.09	0.87	2.17	0.48	0.21	0.48	0.27	0.13
	1000	0.06	0.01	2.39	0.03	0	0.87	1.26	1.25	1.32	2.69	0.67	0.34	0.87	0.26	0.16
	5000	0.01	0	2.82	0	0	1.57	1.34	1.34	2.17	3.01	0.64	0.16	0.99	0.18	0.12
	10000	0.01	0	2.8	0.01	0	1.69	1.5	1.48	2.59	2.95	0.51	0.1	0.85	0.14	0.1
f_{16}	50	0.11	0	0.29	0.96	0.93	0.74	0.74	0.71	0.48	0.65	0.29	0	0.34	0.26	0.2
	100	0.1	0.03	0.72	0	0.01	0	0.9	0.91	0.78	1.2	0.46	0.02	0.55	0.25	0.21
	500	0.07	0.02	1.8	0.58	0.57	0.79	1.53	1.52	1.32	2.19	0.5	0	1.21	0.09	0.03
	1000	0	0	2.18	0.27	0.28	0.84	1.21	1.22	1.26	2.49	0.55	0.03	1.24	0	0
	5000	0.06	0	2.64	0.04	0	1.36	1.12	1.1	1.92	2.83	0.48	0.45	1.57	0.06	0
	10000	0.04	0	2.97	0.04	0	1.83	1.17	1.15	2.58	3.12	0.59	0.8	1.65	0.04	0

Table 1: Binary logarithms of relative squared Hellinger Risks w.r.t best method

optimization over a finest regular partition with bin length $(x_{(n)} - x_{(1)})\lfloor n/\log^{1.5}(n)\rfloor^{-1}$ (**B″**, **R″**,**CV″**), the picture changes. Overall, the performance of all three methods is not bad, in particular **CV″** often outperforms **B″** and **R″**, which behave very similarly. An exception are again histogram densities, where cross-validation performs badly, especially for the uniform. Putting a constraint on the minimum bin size causes a problem for all three methods when

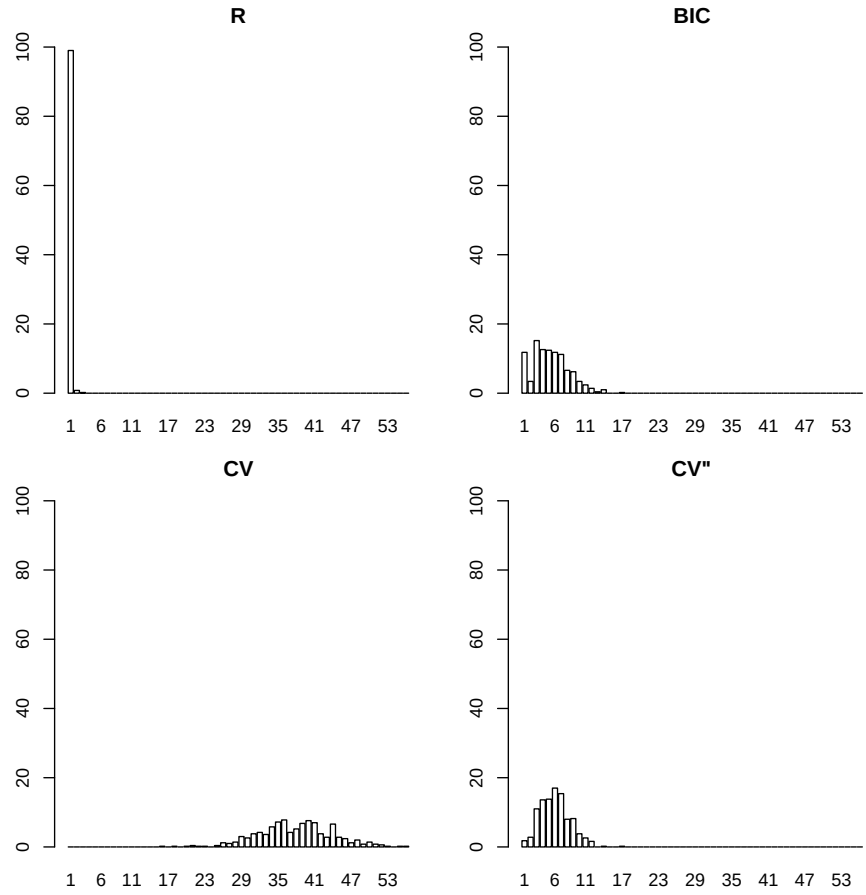


Figure 2: Barplots of number of bins chosen for the uniform density with $n = 500$ (percent of simulation runs).

the density has very sharp peaks (especially the trimodal uniform f_{10}). The intermediate case, i.e. using both penalties and cross-validation ($\mathbf{B}'$, $\mathbf{R}'$,$\mathbf{CV}'$) for a data-driven finest grid but adding the constraint that bin widths have to be at least $(x_{(n)} - x_{(1)}) \log^{1.5}(n)/n$ could be suspected to give a compromise between the finest regular grid suggested by theory and the greedy algorithm for a data-driven grid. However, $\mathbf{B}'$ and $\mathbf{R}'$ share the catastrophic behavior of $\mathbf{B}''$ and $\mathbf{R}''$ at the trimodal uniform density f_{10} without offering a real improvement over $\mathbf{B}$ and $\mathbf{R}$ at the more well-behaved densities. On the other hand, $\mathbf{CV}'$ is a good compromise between $\mathbf{CV}$ and $\mathbf{CV}''$, as it is in many cases either better than both or not far from the better of the two. It still shows bad behavior for the uniform and trimodal uniform densities. Table 8 shows that cross-validation has a pronounced tendency to choose histograms with a much larger number of bins than the penalized likelihood methods for all three sets of partitions. This is also illustrated by Fig. 2: For the uniform distribution with $n = 500$, our proposal $\mathbf{R}$ often chooses only one bin, which is the best possible for the uniform. $\mathbf{CV}$ chooses by far too many bins, resulting in a bad risk behavior. This is less extreme for $\mathbf{CV}''$, but the number of bins chosen is still too large and the risk is more than 3 times larger than the best achieved for this setting.

The taut string method $\mathbf{TS}$ shows a particularly good behavior in terms of Hellinger risk.

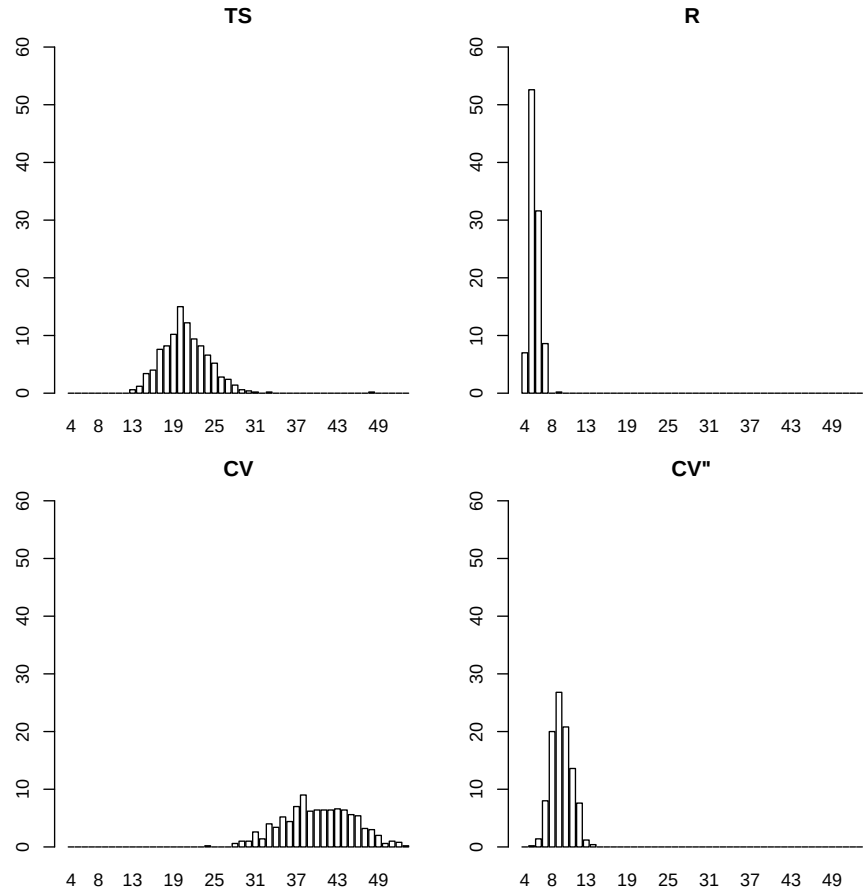


Figure 3: Barplots of number of bins chosen for the lognormal density $n = 500$ (percent of simulation runs).

One should note here that it does not control the number of bins but the modality of the estimate, thereby avoiding overfitting while still being able to chose a large number of bins to give sufficient detail. An example is given in Fig. 3, where the number of bins chosen for the lognormal distribution f_4 (with $n = 500$) is shown. Of the four methods shown, **CV** performs worst, choosing again a large number of bins. **TS** is best in this scenario and uses a larger number of bins than both **R** and **CV**$''$. One should note that the taut string method was originally derived for different aims than achieving a good behavior w.r.t. a given loss function (Davies and Kovac, 2004), and many questions regarding behavior in a more classical framework remain open.

Using **AIC** leads to very bad results. It has already been shown theoretically in Castellan (1999) and Massart (2007) and from a more practical point of view in Birgé and Rozenholc (2006) that **AIC** underpenalizes even for regular histograms. The tendency to overfit becomes even worse when using **AIC** for irregular histograms, since now there are many partitions with the same number of bins, leading to problems similar to those arising in multiple testing. As argued by Castellan (1999) and Massart (2007), in this case, an additional penalty is needed. Table 8 shows that the number of bins chosen by **AIC** on average is very often the largest among all methods considered. In many cases, the ratio of the Hellinger risk to the best risk

achieved by any method in the simulation study is at least 4, often even much larger. **BIC** is a criterion which does not aim for a good control of risk but at asymptotically identifying the "smallest true model", if it exists. It shares the problem of **AIC** of not taking into account the number of models of the same size. It shows some good behavior in particular for small sample sizes that deteriorates when samples become larger. Particularly noteworthy is the bad performance for "simple" models like the uniform (which is also shown in Fig. 2) and the 5 bin regular histogram density f_{13}.

The regular histogram method **BR**, which improves on Akaike's penalization, is the best method for f_3, f_{11} and f_{13}, at least when the sample size is not very small. This shows that the greater flexibility of an irregular histogram over a regular one may be outweighed by the greater difficulty in choosing a good partition, as was already remarked by Birgé and Rozenholc (2006). Regular histograms are inferior for spatially inhomogeneous densities like f_4 and f_{10}.

The simulations show that one can successfully combine the advantages of regular and irregular histograms ($\mathbf{B}^*$ and $\mathbf{R}^*$). The Hellinger risk for $\mathbf{B}^*$ is always within twice the risk of the best method for a given situation. While $\mathbf{R}^*$ is even better in most cases, it shares the bad behaviour of **BR** for f_{10} and $n = 50$. Table 2 shows that for larger sample sizes both $\mathbf{B}^*$ and $\mathbf{R}^*$ almost always choose an irregular histogram for densities where this is advantageous (the spatially inhomogenous densities f_4, f_5, f_{10}, f_{14} and f_{16}) and a regular partition in most of the other cases.

n	method	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}	f_{15}	f_{16}
50	$\mathbf{B}^*$	0.08	0.08	0.05	0.75	1.00	0.05	0.14	0.07	0.11	1.00	0.15	0.39	0.06	0.11	0.22	0.51
	$\mathbf{R}^*$	0.02	0.13	0.10	0.68	1.00	0.11	0.27	0.06	0.12	0.00	0.02	0.16	0.03	0.07	0.22	0.43
100	$\mathbf{B}^*$	0.04	0.07	0.04	0.74	1.00	0.11	0.23	0.02	0.19	1.00	0.12	0.43	0.01	0.12	0.24	0.65
	$\mathbf{R}^*$	0.00	0.10	0.10	0.64	1.00	0.24	0.40	0.05	0.36	1.00	0.05	0.28	0.01	0.11	0.33	0.65
500	$\mathbf{B}^*$	0.03	0.00	0.00	0.73	1.00	0.01	0.03	0.00	0.42	1.00	0.00	0.47	0.00	0.69	0.36	0.99
	$\mathbf{R}^*$	0.00	0.00	0.00	0.47	1.00	0.04	0.07	0.00	0.53	1.00	0.00	0.45	0.00	0.75	0.48	0.99
1000	$\mathbf{B}^*$	0.02	0.00	0.00	0.89	1.00	0.00	0.00	0.00	0.58	1.00	0.00	0.41	0.00	0.96	0.49	1.00
	$\mathbf{R}^*$	0.01	0.00	0.00	0.63	1.00	0.01	0.00	0.00	0.64	1.00	0.00	0.41	0.00	0.98	0.57	1.00
5000	$\mathbf{B}^*$	0.01	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.65	1.00	0.00	0.26	0.00	1.00	0.43	1.00
	$\mathbf{R}^*$	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.70	1.00	0.00	0.28	0.00	1.00	0.47	1.00
10000	$\mathbf{B}^*$	0.00	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.66	1.00	0.00	0.18	0.00	1.00	0.35	1.00
	$\mathbf{R}^*$	0.00	0.00	0.00	1.00	0.98	0.00	0.00	0.00	0.71	1.00	0.00	0.20	0.00	1.00	0.39	1.00

Table 2: Frequency of choosing an irregular partition.

To summarize, we propose a practical method of irregular histogram construction inspired by theoretical works by Barron et al. (1999), Castellan (1999, 2000) and Massart (2007). It can be easily implemented using a dynamic programming algorithm and it performs well for a wide range of different densities and sample sizes, even for some cases not covered by the underlying theory. Performance is shown to be improved when combined with the regular histogram apporach proposed in Birgé and Rozenholc (2006). All procedures proposed here are available in the R-package `histogram` (Mildenberger et al., 2009a).

Acknowledgments

This work has been supported in part by the Collaborative Research Centers "Reduction of Complexity in Multivariate Data Structures" (SFB 475) and "Statistical Modelling of Nonlinear Dynamic Processes" (SFB 823) of the German Research Foundation (DFG). The authors also wish to thank Henrike Weinert for discussions and programming in earlier stages of the work as well as two anonymous referees for giving comments that greatly improved the paper.

References

Akaike, H., 1973. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716-723.

Barron, A., Birgé, L., Massart, P., 1999. Risk bounds for model selection via penalization. Probability Theory and Related Fields 113, 301-413.

Berlinet, A., Devroye, L., 1994. A comparison of kernel density estimates. Publications de l'Institut de Statistique de l'Universite de Paris 38, 3-59.

Birgé, L., Rozenholc, Y., 2006. How many bins should be put in a regular histogram? ESAIM: Probability and Statistics 10, 24-45.

Blanchard, G., Schäfer, C., Rozenholc, Y., Müller, K.-R., 2007. Optimal dyadic decision trees. Machine Learning 66, 209-241.

Castellan, G., 1999. Modified Akaike's criterion for histogram density estimation. Technical Report 99.61, Université de Paris-Sud.

Castellan, G., 2000. Sélection d'histogrammes à l'aide d'un critère de type Akaike. Comptes rendus de l'Académie des sciences Paris 330, Série I, 729-732.

Catoni, O., 2002. Data compression and adaptive histograms, in: Cucker, F., Rojas J.M. (Eds.), Foundations of Computational Mathematics, Proceedings of the Smalefest 2000, World Scientific, Singapore, pp. 35-60.

Celisse, A., Robin, S., 2008. Nonparametric density estimation by exact leave-p-out cross-validation. Computational Statistics and Data Analysis 52, 2350-2368.

Chen, X.R., Zhao, L.C., 1987. Almost sure L_1-norm convergence for data-based histogram density estimators. Journal of Multivariate Analysis 21, 179-188.

Comte, F., Rozenholc, Y., 2004. A new algorithm for fixed design regression and denoising. Annals of the Institute of Statistical Mathematics 56, 449-473.

Davies, P. L., Gather, U., Nordman, D. J., Weinert, H., 2008. A comparison of automatic histogram constructions. ESAIM: Probability and Statistics 13, 181-196.

Davies, P. L., Kovac, A., 2004. Densities, spectral densities and modality. The Annals of Statistics 32, 1093-1136.

Davies, P.L., Kovac, A., 2008. ftnonpar: Features and strings for nonparametric regression. R package version 0.1-83.

Devroye, L., Györfi, L., 1985. Nonparametric density estimation: the L_1 view. Wiley, New York.

Devroye, L., Lugosi, G.. 2004. Bin width selection in multivariate histograms by the combinatorial method, Test 13, 129-145.

Engel, J., 1997. The multiresolution histogram. Metrika 46, 41-57.

Hartigan, J.A., 1996. Bayesian histograms, in: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics 5, Oxford University Press, Oxford, pp. 211-222.

Kanazawa, Y., 1988. An optimal variable cell histogram. Communications in Statistics - Theory and Methods 17, 1401-1422.

Kanazawa, Y., 1992. An optimal variable cell histogram based on the sample spacings. The Annals of Statistics 20,219-304.

Klemelä, J., 2007. Density estimation with stagewise optimization of the empirical risk. Machine Learning 67, 169-195.

Klemelä, J., 2009. Multivariate histograms with data-dependent partitions. Statistica Sinica 19, 159-176.

Kogure, A., 1986. Optimal cells for a histogram. PhD thesis, Yale University.

Kogure, A., 1987. Asymptotically optimal cells for a histogram. The Annals of Statistics 15, 1023-1030.

Kontkanen, P., Myllymäki, P., 2007. MDL histogram density estimation. In: Meila M., Shen S. (Eds.), Proc. 11th International Conference on Artificial Intelligence and Statistics (AISTATS 2007), Puerto Rico, March 2007. http://www.stat.umn.edu/∼aistat/proceedings/start.htm

Lugosi, G., Nobel, A., 1996. Consistency of data-driven histogram methods for density estimation and classification. The Annals of Statistics 24, 687-706.

Massart, P., 2007. Concentration inequalities and model selection. Lecture Notes in Mathematics Vol. 1896, Springer, New York.

Mildenberger, T., Rozenholc, Y., Zasada, D., 2009a. histogram: Construction of regular and irregular histograms with different options for automatic choice of bins. R package version 0.0-20.

Mildenberger, T., Weinert, H., Tiemeyer, S., 2009b. benchden: 28 benchmark densities from Berlinet/Devroye (1994). R package version 1.0.2.

Rissanen, J., Speed, T. P., Yu, B., 1992. Density estimation by stochastic complexity. IEEE Transactions on Information Theory 38, 315-323.

Schwarz, G., 1978. Estimating the dimension of a model. The Annals of Statistics 6, 461-464.

Zhao, L.C, Krishnaiah, P.R., Chen, X.R., 1988. Almost sure L_r-norm convergence for data-based histogram estimates. Theory of Probability and its Applications 35, 396-403.

A. Tables

	n	B	R	CV	B′	R′	CV′	B″	R″	CV″	AIC	BIC	TS	BR	B*	R*
f_1	50	0.022	0.021	0.058	0.021	0.021	0.034	0.02	0.02	0.027	0.085	0.049	0.021	0.022	0.023	0.022
	100	0.011	0.01	0.05	0.01	0.01	0.024	0.01	0.01	0.016	0.075	0.029	0.011	0.011	0.011	0.011
	500	0.002	0.002	0.036	0.002	0.002	0.012	0.002	0.002	0.006	0.049	0.008	0.002	0.002	0.002	0.002
	1000	0.001	0.001	0.025	0.001	0.001	0.008	0.001	0.001	0.005	0.032	0.004	0.001	0.001	0.001	0.001
	5000	0	0	0.008	0	0	0.003	0	0	0.003	0.009	0.001	0	0	0	0
	10000	0	0	0.004	0	0	0.002	0	0	0.003	0.005	0	0	0	0	0
f_2	50	0.066	0.063	0.07	0.065	0.063	0.053	0.071	0.069	0.05	0.093	0.072	0.054	0.055	0.056	0.056
	100	0.047	0.046	0.056	0.047	0.045	0.033	0.047	0.047	0.032	0.081	0.049	0.032	0.037	0.038	0.038
	500	0.019	0.018	0.039	0.018	0.018	0.012	0.018	0.018	0.012	0.052	0.019	0.011	0.013	0.013	0.013
	1000	0.012	0.012	0.028	0.012	0.012	0.008	0.012	0.012	0.008	0.035	0.012	0.007	0.008	0.008	0.008
	5000	0.005	0.005	0.01	0.005	0.005	0.003	0.005	0.005	0.004	0.011	0.004	0.003	0.003	0.003	0.003
	10000	0.003	0.003	0.006	0.003	0.003	0.002	0.003	0.003	0.003	0.006	0.003	0.002	0.002	0.002	0.002
f_3	50	0.058	0.054	0.065	0.057	0.054	0.045	0.057	0.056	0.04	0.09	0.063	0.057	0.047	0.048	0.047
	100	0.036	0.034	0.054	0.035	0.034	0.029	0.036	0.035	0.026	0.08	0.043	0.031	0.029	0.029	0.029
	500	0.014	0.014	0.039	0.014	0.014	0.012	0.014	0.014	0.01	0.051	0.016	0.01	0.009	0.009	0.009
	1000	0.009	0.009	0.028	0.009	0.009	0.008	0.009	0.009	0.007	0.034	0.01	0.006	0.006	0.006	0.006
	5000	0.004	0.004	0.009	0.004	0.004	0.003	0.004	0.003	0.004	0.01	0.003	0.003	0.002	0.002	0.002
	10000	0.002	0.002	0.005	0.002	0.002	0.002	0.002	0.002	0.003	0.006	0.002	0.002	0.001	0.001	0.001
f_4	50	0.061	0.059	0.067	0.07	0.069	0.057	0.067	0.065	0.056	0.091	0.066	0.042	0.061	0.062	0.06
	100	0.04	0.04	0.055	0.043	0.043	0.034	0.043	0.042	0.034	0.08	0.045	0.026	0.04	0.04	0.04
	500	0.016	0.017	0.039	0.018	0.018	0.014	0.018	0.018	0.015	0.051	0.017	0.011	0.018	0.016	0.017
	1000	0.01	0.011	0.028	0.014	0.014	0.011	0.014	0.014	0.012	0.035	0.011	0.008	0.014	0.011	0.012
	5000	0.004	0.005	0.009	0.008	0.008	0.007	0.007	0.007	0.006	0.01	0.004	0.003	0.01	0.004	0.005
	10000	0.003	0.003	0.005	0.005	0.005	0.004	0.004	0.005	0.004	0.006	0.002	0.002	0.007	0.003	0.003
f_5	50	0.087	0.084	0.104	0.333	0.333	0.334	0.223	0.221	0.219	0.114	0.092	0.073	0.16	0.087	0.084
	100	0.061	0.06	0.075	0.288	0.288	0.292	0.154	0.155	0.15	0.092	0.061	0.044	0.058	0.061	0.06
	500	0.022	0.022	0.043	0.042	0.042	0.049	0.048	0.048	0.041	0.054	0.023	0.013	0.029	0.022	0.022
	1000	0.015	0.015	0.03	0.024	0.024	0.028	0.02	0.02	0.016	0.036	0.015	0.008	0.015	0.015	0.015
	5000	0.006	0.006	0.01	0.006	0.006	0.004	0.005	0.006	0.003	0.011	0.005	0.003	0.005	0.006	0.006
	10000	0.004	0.004	0.006	0.004	0.004	0.003	0.004	0.004	0.002	0.006	0.003	0.002	0.003	0.004	0.004
f_6	50	0.05	0.047	0.064	0.05	0.047	0.045	0.048	0.047	0.038	0.089	0.06	0.054	0.046	0.047	0.046
	100	0.031	0.03	0.054	0.031	0.03	0.03	0.033	0.032	0.024	0.079	0.042	0.033	0.03	0.03	0.03
	500	0.014	0.013	0.039	0.014	0.013	0.012	0.014	0.013	0.01	0.051	0.016	0.01	0.01	0.01	0.01
	1000	0.009	0.009	0.027	0.009	0.009	0.008	0.009	0.009	0.007	0.034	0.01	0.007	0.006	0.006	0.006
	5000	0.004	0.003	0.009	0.003	0.003	0.003	0.003	0.003	0.004	0.01	0.003	0.002	0.002	0.002	0.002
	10000	0.002	0.002	0.005	0.002	0.002	0.002	0.002	0.002	0.003	0.006	0.002	0.001	0.001	0.001	0.001
f_7	50	0.066	0.061	0.076	0.065	0.062	0.056	0.072	0.07	0.055	0.099	0.075	0.072	0.063	0.063	0.061
	100	0.039	0.038	0.06	0.038	0.038	0.04	0.044	0.043	0.039	0.084	0.055	0.04	0.042	0.041	0.04
	500	0.024	0.023	0.04	0.024	0.023	0.016	0.024	0.023	0.016	0.053	0.024	0.015	0.018	0.018	0.018
	1000	0.016	0.016	0.03	0.016	0.015	0.012	0.019	0.019	0.01	0.036	0.015	0.009	0.012	0.012	0.012
	5000	0.006	0.006	0.01	0.006	0.006	0.004	0.006	0.006	0.005	0.011	0.006	0.003	0.004	0.004	0.004
	10000	0.004	0.004	0.006	0.004	0.004	0.003	0.004	0.004	0.004	0.006	0.004	0.002	0.003	0.003	0.003
f_8	50	0.111	0.106	0.096	0.111	0.108	0.093	0.114	0.113	0.09	0.119	0.105	0.105	0.1	0.101	0.1
	100	0.079	0.074	0.073	0.079	0.075	0.061	0.095	0.089	0.065	0.097	0.073	0.068	0.065	0.065	0.065
	500	0.032	0.03	0.043	0.033	0.032	0.027	0.037	0.035	0.023	0.055	0.03	0.024	0.025	0.025	0.025
	1000	0.021	0.02	0.031	0.022	0.021	0.018	0.023	0.022	0.016	0.037	0.019	0.015	0.017	0.017	0.017
	5000	0.008	0.008	0.01	0.008	0.008	0.006	0.008	0.008	0.006	0.011	0.007	0.004	0.006	0.006	0.006
	10000	0.005	0.005	0.006	0.005	0.005	0.004	0.005	0.005	0.004	0.006	0.005	0.003	0.004	0.004	0.004
f_9	50	0.119	0.117	0.087	0.119	0.119	0.098	0.119	0.119	0.112	0.111	0.089	0.105	0.112	0.114	0.113
	100	0.055	0.049	0.066	0.056	0.049	0.04	0.104	0.102	0.085	0.088	0.055	0.045	0.061	0.057	0.052
	500	0.017	0.017	0.041	0.017	0.017	0.015	0.037	0.036	0.03	0.053	0.019	0.012	0.021	0.018	0.017
	1000	0.011	0.011	0.029	0.011	0.011	0.01	0.023	0.022	0.019	0.036	0.012	0.007	0.015	0.011	0.011
	5000	0.004	0.004	0.01	0.004	0.004	0.004	0.008	0.008	0.007	0.011	0.004	0.002	0.006	0.004	0.004
	10000	0.003	0.002	0.006	0.003	0.002	0.003	0.005	0.005	0.005	0.006	0.002	0.002	0.004	0.003	0.003
f_{10}	50	0.087	0.216	0.127	0.773	0.77	0.767	0.754	0.752	0.752	0.139	0.108	0.097	0.556	0.087	0.555
	100	0.043	0.042	0.084	0.765	0.764	0.764	0.675	0.675	0.676	0.104	0.06	0.052	0.408	0.043	0.042
	500	0.008	0.008	0.044	0.467	0.467	0.468	0.414	0.414	0.414	0.055	0.014	0.012	0.281	0.008	0.008
	1000	0.004	0.004	0.029	0.465	0.465	0.466	0.338	0.338	0.338	0.035	0.007	0.006	0.207	0.004	0.004
	5000	0.001	0.001	0.009	0.464	0.464	0.464	0.161	0.161	0.162	0.01	0.001	0.001	0.001	0.001	0.001
	10000	0	0	0.005	0.464	0.464	0.464	0.037	0.037	0.037	0.005	0.001	0.001	0	0	0
f_{11}	50	0.077	0.075	0.108	0.075	0.075	0.086	0.074	0.074	0.081	0.13	0.106	0.085	0.076	0.078	0.076
	100	0.065	0.065	0.084	0.064	0.064	0.076	0.064	0.064	0.07	0.107	0.086	0.076	0.065	0.066	0.066
	500	0.049	0.042	0.047	0.05	0.044	0.033	0.057	0.056	0.03	0.06	0.038	0.029	0.027	0.027	0.027
	1000	0.026	0.025	0.034	0.026	0.026	0.022	0.031	0.028	0.018	0.04	0.026	0.018	0.017	0.017	0.017
	5000	0.011	0.011	0.011	0.011	0.011	0.008	0.011	0.011	0.007	0.012	0.01	0.006	0.006	0.006	0.006
	10000	0.008	0.007	0.007	0.008	0.007	0.005	0.008	0.007	0.005	0.007	0.007	0.004	0.004	0.004	0.004
f_{12}	50	0.041	0.038	0.065	0.035	0.036	0.042	0.035	0.035	0.039	0.09	0.061	0.036	0.039	0.042	0.04
	100	0.027	0.026	0.052	0.026	0.026	0.028	0.025	0.025	0.023	0.078	0.039	0.026	0.026	0.027	0.026
	500	0.009	0.009	0.038	0.009	0.009	0.013	0.009	0.009	0.009	0.05	0.013	0.007	0.009	0.009	0.009
	1000	0.006	0.006	0.027	0.006	0.006	0.009	0.006	0.006	0.006	0.034	0.008	0.004	0.006	0.006	0.006
	5000	0.002	0.002	0.009	0.002	0.002	0.003	0.002	0.002	0.003	0.01	0.003	0.001	0.002	0.002	0.002
	10000	0.001	0.001	0.005	0.001	0.001	0.002	0.001	0.001	0.003	0.006	0.002	0.001	0.001	0.001	0.001
f_{13}	50	0.044	0.043	0.066	0.042	0.042	0.041	0.041	0.041	0.041	0.091	0.064	0.043	0.043	0.043	0.043
	100	0.031	0.031	0.054	0.03	0.03	0.027	0.03	0.03	0.024	0.079	0.042	0.031	0.023	0.023	0.023
	500	0.01	0.009	0.038	0.009	0.009	0.013	0.01	0.01	0.009	0.052	0.013	0.008	0.004	0.004	0.004
	1000	0.004	0.003	0.027	0.004	0.003	0.009	0.003	0.003	0.006	0.034	0.007	0.004	0.002	0.002	0.002
	5000	0.001	0.001	0.009	0.001	0.001	0.003	0.001	0.001	0.003	0.01	0.001	0.001	0	0	0
	10000	0	0	0.005	0	0	0.002	0	0	0.003	0.005	0.001	0	0	0	0
f_{14}	50	0.049	0.048	0.068	0.047	0.047	0.048	0.047	0.048	0.043	0.093	0.066	0.05	0.044	0.046	0.045
	100	0.033	0.032	0.055	0.033	0.032	0.034	0.035	0.035	0.031	0.08	0.043	0.034	0.028	0.028	0.028
	500	0.008	0.007	0.039	0.007	0.007	0.013	0.011	0.011	0.012	0.051	0.013	0.008	0.013	0.009	0.009
	1000	0.004	0.004	0.027	0.004	0.004	0.009	0.006	0.006	0.008	0.034	0.007	0.003	0.007	0.004	0.004
	5000	0.001	0.001	0.008	0.001	0.001	0.003	0.002	0.002	0.004	0.01	0.001	0.001	0.002	0.001	0.001
	10000	0	0	0.005	0	0	0.002	0.001	0.001	0.003	0.005	0.001	0	0.002	0	0
f_{15}	50	0.085	0.079	0.083	0.101	0.097	0.083	0.109	0.106	0.083	0.107	0.084	0.086	0.081	0.081	0.08
	100	0.049	0.046	0.064	0.048	0.047	0.041	0.066	0.064	0.052	0.088	0.056	0.044	0.049	0.048	0.047
	500	0.013	0.012	0.04	0.013	0.012	0.014	0.026	0.025	0.021	0.053	0.016	0.014	0.016	0.014	0.013
	1000	0.006	0.005	0.028	0.005	0.005	0.01	0.013	0.013	0.013	0.035	0.009	0.007	0.01	0.006	0.006
	5000	0.001	0.001	0.009	0.001	0.001	0.004	0.003	0.003	0.006	0.01	0.002	0.001	0.002	0.001	0.001
	10000	0.001	0.001	0.005	0.001	0.001	0.002	0.002	0.002	0.004	0.005	0.001	0.001	0.001	0.001	0.001
f_{16}	50	0.072	0.067	0.082	0.129	0.127	0.111	0.111	0.109	0.092	0.104	0.082	0.067	0.084	0.079	0.077
	100	0.04	0.039	0.062	0.038	0.038	0.038	0.07	0.071	0.065	0.087	0.052	0.038	0.055	0.045	0.044
	500	0.012	0.012	0.04	0.017	0.017	0.02	0.034	0.033	0.029	0.053	0.016	0.012	0.027	0.012	0.012
	1000	0.006	0.006	0.028	0.008	0.008	0.011	0.015	0.015	0.015	0.035	0.009	0.006	0.015	0.006	0.006
	5000	0.001	0.001	0.009	0.001	0.001	0.004	0.003	0.003	0.005	0.01	0.002	0.002	0.004	0.001	0.001
	10000	0.001	0.001	0.005	0.001	0.001	0.002	0.001	0.001	0.004	0.005	0.001	0.001	0.002	0.001	0.001

Table 3: Squared Hellinger Risks.

	n	B	R	CV	B′	R′	CV′	B″	R″	CV″	AIC	BIC	TS	BR	B*	R*
f_1	50	0.093	0.089	0.426	0.086	0.087	0.29	0.081	0.081	0.202	0.518	0.289	0.089	0.113	0.116	0.114
	100	0.043	0.043	0.413	0.042	0.043	0.273	0.04	0.041	0.169	0.502	0.211	0.047	0.062	0.064	0.063
	500	0.009	0.009	0.361	0.008	0.008	0.213	0.008	0.008	0.129	0.402	0.091	0.012	0.019	0.019	0.019
	1000	0.004	0.004	0.276	0.004	0.004	0.179	0.004	0.004	0.123	0.298	0.062	0.006	0.012	0.012	0.012
	5000	0.001	0.001	0.113	0.001	0.001	0.084	0.001	0.001	0.104	0.117	0.019	0.002	0.004	0.004	0.004
	10000	0	0	0.073	0	0	0.057	0	0	0.099	0.075	0.011	0.002	0.004	0.004	0.004
f_2	50	0.499	0.485	0.514	0.491	0.483	0.422	0.534	0.524	0.41	0.575	0.497	0.431	0.436	0.442	0.442
	100	0.431	0.423	0.469	0.429	0.422	0.339	0.441	0.438	0.328	0.54	0.41	0.326	0.364	0.369	0.372
	500	0.276	0.275	0.394	0.276	0.275	0.209	0.279	0.28	0.2	0.434	0.258	0.188	0.222	0.222	0.222
	1000	0.226	0.226	0.318	0.225	0.226	0.171	0.227	0.229	0.162	0.338	0.208	0.147	0.178	0.178	0.178
	5000	0.139	0.141	0.161	0.139	0.141	0.109	0.139	0.141	0.102	0.164	0.125	0.084	0.105	0.105	0.105
	10000	0.112	0.115	0.118	0.112	0.115	0.088	0.112	0.114	0.085	0.119	0.1	0.066	0.084	0.084	0.084
f_3	50	0.44	0.42	0.477	0.436	0.421	0.37	0.443	0.437	0.331	0.553	0.435	0.441	0.372	0.374	0.373
	100	0.334	0.326	0.453	0.333	0.325	0.309	0.346	0.338	0.271	0.531	0.356	0.296	0.287	0.289	0.29
	500	0.219	0.216	0.384	0.218	0.215	0.211	0.223	0.218	0.171	0.424	0.214	0.155	0.168	0.168	0.168
	1000	0.18	0.178	0.305	0.18	0.178	0.178	0.181	0.178	0.142	0.324	0.172	0.121	0.135	0.135	0.135
	5000	0.112	0.111	0.147	0.112	0.111	0.109	0.112	0.111	0.1	0.15	0.102	0.068	0.076	0.076	0.076
	10000	0.09	0.09	0.105	0.09	0.09	0.084	0.09	0.089	0.087	0.106	0.081	0.054	0.06	0.06	0.06
f_4	50	0.464	0.461	0.492	0.521	0.519	0.453	0.512	0.507	0.435	0.561	0.462	0.342	0.449	0.458	0.454
	100	0.379	0.383	0.463	0.412	0.414	0.351	0.405	0.406	0.335	0.536	0.379	0.264	0.365	0.375	0.375
	500	0.249	0.261	0.385	0.268	0.275	0.228	0.264	0.274	0.217	0.425	0.233	0.17	0.254	0.247	0.255
	1000	0.2	0.214	0.312	0.23	0.237	0.195	0.23	0.238	0.185	0.331	0.188	0.146	0.225	0.201	0.216
	5000	0.124	0.139	0.153	0.152	0.162	0.129	0.143	0.154	0.112	0.156	0.112	0.076	0.167	0.124	0.139
	10000	0.1	0.113	0.112	0.117	0.127	0.097	0.112	0.123	0.086	0.113	0.09	0.059	0.113	0.1	0.113
f_5	50	0.505	0.495	0.574	1.047	1.047	1.046	0.84	0.837	0.832	0.622	0.53	0.459	0.697	0.505	0.495
	100	0.437	0.432	0.507	0.939	0.94	0.942	0.741	0.743	0.725	0.57	0.434	0.353	0.427	0.437	0.432
	500	0.267	0.27	0.398	0.371	0.374	0.357	0.403	0.407	0.35	0.433	0.261	0.186	0.301	0.267	0.27
	1000	0.22	0.226	0.319	0.268	0.27	0.248	0.254	0.258	0.21	0.336	0.208	0.143	0.206	0.22	0.226
	5000	0.135	0.141	0.158	0.139	0.143	0.117	0.136	0.142	0.101	0.16	0.125	0.081	0.118	0.135	0.141
	10000	0.11	0.114	0.115	0.11	0.114	0.09	0.111	0.115	0.08	0.116	0.1	0.064	0.093	0.11	0.113
f_6	50	0.372	0.356	0.478	0.368	0.356	0.372	0.357	0.352	0.311	0.556	0.418	0.395	0.358	0.36	0.356
	100	0.29	0.282	0.453	0.29	0.284	0.322	0.304	0.296	0.261	0.532	0.349	0.299	0.294	0.292	0.29
	500	0.208	0.205	0.384	0.207	0.205	0.216	0.206	0.203	0.172	0.423	0.211	0.179	0.173	0.173	0.174
	1000	0.173	0.171	0.302	0.173	0.171	0.18	0.173	0.171	0.144	0.321	0.167	0.143	0.138	0.138	0.138
	5000	0.107	0.106	0.146	0.107	0.106	0.108	0.106	0.105	0.101	0.149	0.1	0.07	0.08	0.08	0.08
	10000	0.09	0.089	0.105	0.089	0.089	0.084	0.09	0.089	0.089	0.106	0.08	0.053	0.063	0.063	0.063
f_7	50	0.494	0.472	0.562	0.489	0.473	0.453	0.523	0.515	0.453	0.623	0.526	0.526	0.484	0.483	0.473
	100	0.394	0.389	0.505	0.39	0.388	0.4	0.423	0.42	0.405	0.571	0.464	0.399	0.413	0.408	0.402
	500	0.343	0.339	0.406	0.341	0.337	0.253	0.347	0.345	0.261	0.445	0.307	0.251	0.266	0.268	0.271
	1000	0.266	0.266	0.337	0.264	0.265	0.218	0.315	0.314	0.2	0.356	0.24	0.188	0.211	0.211	0.211
	5000	0.16	0.16	0.176	0.16	0.16	0.133	0.161	0.161	0.125	0.179	0.149	0.103	0.127	0.127	0.127
	10000	0.13	0.131	0.132	0.13	0.131	0.107	0.133	0.134	0.104	0.133	0.119	0.08	0.101	0.101	0.101
f_8	50	0.681	0.657	0.619	0.674	0.66	0.578	0.695	0.683	0.572	0.685	0.633	0.658	0.598	0.603	0.599
	100	0.557	0.534	0.547	0.554	0.534	0.46	0.631	0.604	0.477	0.612	0.52	0.508	0.465	0.466	0.467
	500	0.334	0.323	0.415	0.34	0.333	0.305	0.366	0.355	0.271	0.45	0.316	0.272	0.273	0.273	0.273
	1000	0.269	0.261	0.335	0.273	0.268	0.247	0.286	0.279	0.223	0.353	0.254	0.21	0.224	0.224	0.224
	5000	0.166	0.163	0.172	0.168	0.166	0.145	0.169	0.167	0.137	0.174	0.155	0.112	0.135	0.135	0.135
	10000	0.136	0.134	0.129	0.136	0.134	0.115	0.137	0.135	0.114	0.129	0.125	0.086	0.112	0.112	0.112
f_9	50	0.613	0.602	0.544	0.607	0.601	0.515	0.604	0.6	0.525	0.608	0.533	0.57	0.558	0.568	0.562
	100	0.433	0.409	0.481	0.428	0.401	0.338	0.555	0.542	0.443	0.546	0.413	0.361	0.379	0.382	0.375
	500	0.26	0.256	0.392	0.258	0.254	0.223	0.306	0.302	0.245	0.431	0.247	0.187	0.207	0.226	0.229
	1000	0.211	0.208	0.315	0.21	0.207	0.188	0.239	0.237	0.19	0.334	0.197	0.146	0.174	0.193	0.193
	5000	0.132	0.131	0.159	0.131	0.13	0.12	0.14	0.139	0.121	0.162	0.119	0.083	0.103	0.118	0.119
	10000	0.106	0.105	0.116	0.106	0.105	0.095	0.111	0.11	0.104	0.117	0.094	0.065	0.084	0.095	0.096
f_{10}	50	0.342	0.581	0.556	1.855	1.851	1.847	1.84	1.838	1.838	0.622	0.457	0.447	1.43	0.342	1.428
	100	0.205	0.196	0.479	1.846	1.846	1.846	1.727	1.727	1.727	0.557	0.309	0.314	1.022	0.205	0.196
	500	0.065	0.064	0.38	1.02	1.02	1.024	1.125	1.125	1.125	0.421	0.123	0.137	0.851	0.065	0.064
	1000	0.044	0.044	0.289	1.012	1.012	1.026	0.952	0.952	0.956	0.309	0.081	0.096	0.699	0.044	0.044
	5000	0.017	0.017	0.12	1.001	1.001	1.018	0.556	0.556	0.565	0.124	0.028	0.045	0.046	0.017	0.017
	10000	0.012	0.012	0.08	0.999	0.999	1.013	0.143	0.143	0.154	0.081	0.018	0.033	0.032	0.012	0.012
f_{11}	50	0.526	0.517	0.689	0.515	0.515	0.562	0.512	0.512	0.542	0.759	0.659	0.561	0.519	0.53	0.521
	100	0.504	0.501	0.616	0.498	0.498	0.54	0.497	0.497	0.523	0.677	0.612	0.565	0.502	0.507	0.504
	500	0.452	0.409	0.447	0.457	0.421	0.359	0.49	0.489	0.338	0.484	0.385	0.344	0.316	0.316	0.316
	1000	0.325	0.321	0.366	0.324	0.321	0.3	0.349	0.33	0.263	0.385	0.323	0.265	0.252	0.252	0.252
	5000	0.214	0.209	0.202	0.214	0.209	0.178	0.215	0.21	0.163	0.203	0.202	0.147	0.146	0.146	0.146
	10000	0.176	0.172	0.157	0.176	0.172	0.147	0.178	0.173	0.135	0.158	0.165	0.114	0.116	0.116	0.116
f_{12}	50	0.34	0.336	0.481	0.323	0.325	0.373	0.32	0.321	0.34	0.55	0.419	0.325	0.347	0.352	0.349
	100	0.304	0.303	0.441	0.301	0.303	0.315	0.301	0.302	0.267	0.521	0.332	0.304	0.304	0.305	0.303
	500	0.19	0.189	0.377	0.187	0.187	0.235	0.186	0.184	0.179	0.417	0.198	0.156	0.184	0.191	0.19
	1000	0.161	0.16	0.304	0.161	0.159	0.198	0.163	0.161	0.154	0.324	0.157	0.121	0.148	0.155	0.154
	5000	0.101	0.1	0.146	0.101	0.1	0.112	0.101	0.1	0.115	0.149	0.092	0.068	0.091	0.093	0.093
	10000	0.083	0.082	0.104	0.083	0.082	0.085	0.082	0.082	0.105	0.105	0.073	0.054	0.073	0.075	0.075
f_{13}	50	0.377	0.377	0.496	0.371	0.373	0.369	0.364	0.365	0.359	0.568	0.467	0.37	0.373	0.376	0.375
	100	0.346	0.345	0.456	0.341	0.342	0.302	0.332	0.333	0.275	0.532	0.376	0.341	0.271	0.272	0.273
	500	0.177	0.17	0.381	0.177	0.169	0.224	0.191	0.187	0.169	0.424	0.173	0.156	0.079	0.079	0.079
	1000	0.082	0.079	0.301	0.081	0.079	0.191	0.077	0.075	0.132	0.321	0.111	0.099	0.055	0.055	0.055
	5000	0.027	0.027	0.129	0.027	0.027	0.098	0.028	0.028	0.107	0.133	0.038	0.04	0.023	0.023	0.023
	10000	0.018	0.018	0.087	0.018	0.018	0.071	0.019	0.019	0.101	0.089	0.025	0.029	0.016	0.016	0.016
f_{14}	50	0.442	0.434	0.499	0.437	0.432	0.402	0.452	0.454	0.371	0.571	0.453	0.463	0.384	0.393	0.387
	100	0.346	0.342	0.458	0.349	0.346	0.353	0.396	0.386	0.329	0.535	0.361	0.367	0.296	0.299	0.299
	500	0.144	0.141	0.378	0.142	0.14	0.218	0.178	0.176	0.186	0.421	0.166	0.147	0.212	0.164	0.158
	1000	0.093	0.092	0.299	0.093	0.092	0.183	0.119	0.117	0.158	0.319	0.109	0.095	0.145	0.097	0.093
	5000	0.026	0.026	0.128	0.026	0.026	0.095	0.036	0.036	0.112	0.131	0.038	0.044	0.07	0.026	0.026
	10000	0.017	0.017	0.086	0.017	0.017	0.069	0.02	0.02	0.101	0.088	0.025	0.032	0.06	0.017	0.017
f_{15}	50	0.548	0.519	0.557	0.648	0.629	0.55	0.702	0.686	0.561	0.627	0.521	0.579	0.522	0.53	0.52
	100	0.368	0.355	0.496	0.358	0.354	0.326	0.483	0.476	0.405	0.571	0.411	0.366	0.373	0.367	0.359
	500	0.153	0.142	0.383	0.153	0.142	0.199	0.249	0.242	0.23	0.424	0.177	0.188	0.182	0.162	0.15
	1000	0.091	0.09	0.299	0.091	0.089	0.17	0.14	0.139	0.171	0.32	0.12	0.126	0.128	0.096	0.092
	5000	0.049	0.049	0.133	0.049	0.049	0.098	0.061	0.061	0.108	0.136	0.055	0.057	0.045	0.041	0.041
	10000	0.038	0.038	0.089	0.039	0.038	0.072	0.047	0.046	0.097	0.09	0.037	0.041	0.028	0.029	0.029
f_{16}	50	0.477	0.453	0.544	0.795	0.781	0.695	0.698	0.684	0.586	0.608	0.51	0.476	0.499	0.49	0.474
	100	0.336	0.327	0.486	0.323	0.323	0.31	0.511	0.514	0.478	0.562	0.39	0.341	0.398	0.356	0.35
	500	0.176	0.176	0.387	0.205	0.203	0.231	0.312	0.309	0.285	0.429	0.193	0.184	0.272	0.177	0.176
	1000	0.117	0.119	0.304	0.127	0.129	0.175	0.17	0.172	0.183	0.325	0.135	0.134	0.198	0.117	0.119
	5000	0.044	0.042	0.133	0.043	0.042	0.096	0.054	0.053	0.107	0.137	0.049	0.071	0.105	0.044	0.042
	10000	0.025	0.025	0.089	0.025	0.025	0.07	0.03	0.029	0.093	0.09	0.031	0.053	0.072	0.025	0.025

Table 4: L_1 Risks.

	n	B	R	CV	B′	R′	CV′	B″	R″	CV″	AIC	BIC	TS	BR	B*	R*
f_1	50	0.2	0.14	2.4	0.09	0.1	1.84	0	0	1.32	2.68	1.84	0.14	0.48	0.52	0.5
	100	0.1	0.09	3.36	0.07	0.09	2.76	0	0.04	2.07	3.64	2.39	0.22	0.63	0.68	0.65
	500	0.15	0.19	5.52	0.04	0.11	4.76	0	0.04	4.04	5.67	3.54	0.6	1.28	1.28	1.28
	1000	0.12	0.11	6.09	0.08	0.08	5.46	0	0.05	4.92	6.2	3.93	0.58	1.52	1.53	1.54
	5000	0	0	7.02	0	0	6.6	0	0	6.91	7.07	4.43	1.43	2.33	2.33	2.33
	10000	0	0	7.48	0	0	7.13	0	0	7.92	7.52	4.79	2.12	3.3	3.3	3.3
f_2	50	0.28	0.24	0.33	0.26	0.24	0.04	0.38	0.35	0	0.49	0.28	0.07	0.09	0.11	0.11
	100	0.4	0.38	0.53	0.4	0.37	0.06	0.44	0.43	0.01	0.73	0.33	0	0.16	0.18	0.19
	500	0.56	0.55	1.07	0.55	0.55	0.16	0.57	0.58	0.09	1.21	0.46	0	0.24	0.24	0.24
	1000	0.62	0.62	1.11	0.61	0.62	0.21	0.63	0.64	0.14	1.2	0.5	0	0.27	0.27	0.27
	5000	0.73	0.75	0.94	0.73	0.75	0.38	0.73	0.75	0.28	0.97	0.58	0	0.32	0.32	0.32
	10000	0.77	0.8	0.85	0.77	0.8	0.42	0.77	0.8	0.38	0.86	0.61	0	0.36	0.36	0.36
f_3	50	0.41	0.34	0.53	0.4	0.35	0.16	0.42	0.4	0	0.74	0.39	0.41	0.17	0.18	0.17
	100	0.3	0.26	0.74	0.3	0.26	0.19	0.35	0.32	0	0.97	0.39	0.13	0.08	0.09	0.1
	500	0.5	0.48	1.31	0.5	0.48	0.45	0.52	0.49	0.14	1.45	0.47	0	0.12	0.12	0.12
	1000	0.57	0.56	1.33	0.57	0.56	0.56	0.58	0.55	0.23	1.42	0.5	0	0.15	0.15	0.15
	5000	0.72	0.7	1.1	0.72	0.7	0.67	0.71	0.7	0.54	1.13	0.57	0	0.16	0.16	0.16
	10000	0.75	0.74	0.97	0.75	0.74	0.64	0.75	0.74	0.7	0.99	0.6	0	0.17	0.17	0.17
f_4	50	0.44	0.43	0.52	0.61	0.6	0.4	0.58	0.57	0.35	0.71	0.43	0	0.39	0.42	0.41
	100	0.52	0.54	0.81	0.64	0.65	0.41	0.62	0.62	0.34	1.02	0.52	0	0.47	0.51	0.51
	500	0.56	0.62	1.19	0.66	0.7	0.43	0.64	0.69	0.35	1.33	0.46	0	0.59	0.54	0.59
	1000	0.46	0.56	1.1	0.66	0.7	0.42	0.66	0.71	0.35	1.18	0.37	0	0.63	0.46	0.57
	5000	0.7	0.88	1.02	1	1.09	0.77	0.91	1.02	0.56	1.04	0.56	0	1.14	0.7	0.88
	10000	0.75	0.94	0.91	0.98	1.1	0.72	0.92	1.06	0.53	0.93	0.61	0	0.94	0.75	0.94
f_5	50	0.14	0.11	0.32	1.19	1.19	1.19	0.87	0.87	0.86	0.44	0.21	0	0.6	0.14	0.11
	100	0.31	0.29	0.52	1.41	1.41	1.42	1.07	1.07	1.04	0.69	0.3	0	0.28	0.31	0.29
	500	0.52	0.54	1.1	1	1.01	0.94	1.12	1.13	0.91	1.22	0.49	0	0.7	0.52	0.54
	1000	0.62	0.66	1.15	0.91	0.91	0.79	0.83	0.85	0.55	1.23	0.54	0	0.53	0.62	0.66
	5000	0.74	0.8	0.96	0.78	0.83	0.54	0.75	0.81	0.32	0.99	0.62	0	0.54	0.74	0.8
	10000	0.79	0.83	0.86	0.79	0.84	0.5	0.8	0.86	0.33	0.87	0.66	0	0.54	0.79	0.83
f_6	50	0.26	0.19	0.62	0.24	0.19	0.26	0.2	0.18	0	0.84	0.42	0.34	0.2	0.21	0.19
	100	0.15	0.11	0.79	0.15	0.12	0.3	0.22	0.18	0	1.03	0.42	0.19	0.17	0.16	0.15
	500	0.27	0.26	1.16	0.27	0.26	0.33	0.27	0.24	0	1.3	0.3	0.06	0.01	0.01	0.02
	1000	0.33	0.31	1.13	0.33	0.31	0.39	0.32	0.31	0.06	1.22	0.27	0.05	0	0	0
	5000	0.62	0.6	1.07	0.62	0.6	0.63	0.61	0.59	0.54	1.1	0.52	0	0.2	0.2	0.2
	10000	0.75	0.74	0.98	0.75	0.74	0.66	0.76	0.74	0.74	1	0.58	0	0.25	0.25	0.25
f_7	50	0.13	0.06	0.31	0.11	0.06	0	0.21	0.19	0	0.46	0.22	0.21	0.1	0.09	0.06
	100	0.02	0	0.38	0.01	0	0.05	0.13	0.11	0.06	0.56	0.26	0.04	0.09	0.07	0.05
	500	0.45	0.43	0.69	0.44	0.43	0.01	0.47	0.46	0.05	0.82	0.29	0	0.08	0.09	0.11
	1000	0.5	0.5	0.84	0.49	0.5	0.22	0.75	0.74	0.09	0.92	0.35	0	0.17	0.17	0.17
	5000	0.64	0.64	0.78	0.64	0.64	0.37	0.65	0.65	0.28	0.8	0.54	0	0.31	0.31	0.31
	10000	0.71	0.71	0.72	0.71	0.71	0.43	0.74	0.75	0.39	0.73	0.58	0	0.34	0.34	0.34
f_8	50	0.25	0.2	0.11	0.24	0.21	0.01	0.28	0.25	0	0.26	0.15	0.2	0.06	0.08	0.07
	100	0.28	0.21	0.25	0.27	0.21	0	0.46	0.39	0.05	0.41	0.18	0.14	0.01	0.02	0.02
	500	0.3	0.25	0.61	0.33	0.3	0.17	0.43	0.39	0	0.73	0.22	0.01	0.01	0.01	0.01
	1000	0.36	0.31	0.67	0.38	0.35	0.23	0.44	0.41	0.09	0.75	0.27	0	0.09	0.09	0.09
	5000	0.56	0.54	0.61	0.58	0.56	0.37	0.59	0.57	0.28	0.63	0.47	0	0.26	0.26	0.26
	10000	0.66	0.64	0.58	0.67	0.65	0.42	0.67	0.65	0.41	0.59	0.54	0	0.38	0.38	0.38
f_9	50	0.25	0.23	0.08	0.24	0.22	0	0.23	0.22	0.03	0.24	0.05	0.15	0.12	0.14	0.13
	100	0.36	0.27	0.51	0.34	0.25	0	0.72	0.68	0.39	0.69	0.29	0.09	0.16	0.18	0.15
	500	0.48	0.45	1.07	0.46	0.44	0.25	0.71	0.69	0.39	1.2	0.4	0	0.15	0.27	0.29
	1000	0.53	0.51	1.11	0.52	0.51	0.37	0.71	0.7	0.38	1.19	0.43	0	0.25	0.4	0.4
	5000	0.67	0.66	0.94	0.67	0.66	0.54	0.76	0.75	0.55	0.97	0.52	0	0.32	0.52	0.53
	10000	0.71	0.69	0.84	0.7	0.69	0.54	0.77	0.76	0.68	0.85	0.54	0	0.36	0.55	0.56
f_{10}	50	0	0.77	0.7	2.44	2.44	2.43	2.43	2.43	2.43	0.86	0.42	0.39	2.06	0	2.06
	100	0.07	0	1.29	3.24	3.24	3.24	3.14	3.14	3.14	1.51	0.66	0.68	2.38	0.07	0
	500	0.02	0	2.57	3.99	3.99	4	4.14	4.14	4.14	2.72	0.94	1.1	3.73	0.02	0
	1000	0.02	0	2.73	4.54	4.54	4.56	4.45	4.45	4.46	2.83	0.9	1.15	4	0.02	0
	5000	0.01	0	2.85	5.91	5.91	5.94	5.07	5.07	5.09	2.9	0.76	1.43	1.46	0.01	0
	10000	0	0	2.78	6.42	6.42	6.44	3.62	3.62	3.73	2.81	0.66	1.5	1.48	0	0
f_{11}	50	0.04	0.01	0.43	0.01	0.01	0.13	0	0	0.08	0.57	0.36	0.13	0.02	0.05	0.03
	100	0.02	0.01	0.31	0	0	0.12	0	0	0.07	0.45	0.3	0.18	0.01	0.03	0.02
	500	0.51	0.37	0.5	0.53	0.41	0.18	0.63	0.63	0.09	0.61	0.28	0.12	0	0	0
	1000	0.37	0.35	0.54	0.36	0.35	0.25	0.47	0.39	0.06	0.61	0.36	0.07	0	0	0
	5000	0.55	0.52	0.46	0.55	0.52	0.28	0.55	0.52	0.15	0.47	0.46	0	0	0	0
	10000	0.62	0.59	0.46	0.62	0.59	0.37	0.64	0.61	0.25	0.47	0.54	0	0.02	0.02	0.02
f_{12}	50	0.09	0.07	0.59	0.01	0.02	0.22	0	0.01	0.09	0.78	0.39	0.02	0.12	0.14	0.13
	100	0.19	0.18	0.73	0.18	0.18	0.24	0.17	0.18	0	0.97	0.31	0.19	0.19	0.2	0.19
	500	0.28	0.28	1.27	0.26	0.26	0.59	0.25	0.24	0.19	1.42	0.34	0	0.23	0.29	0.28
	1000	0.41	0.4	1.32	0.4	0.39	0.71	0.43	0.41	0.35	1.42	0.37	0	0.29	0.35	0.34
	5000	0.56	0.55	1.09	0.56	0.55	0.71	0.56	0.54	0.74	1.12	0.43	0	0.4	0.45	0.45
	10000	0.62	0.61	0.95	0.62	0.61	0.66	0.61	0.6	0.96	0.97	0.45	0	0.44	0.48	0.48
f_{13}	50	0.07	0.07	0.46	0.04	0.05	0.04	0.02	0.02	0	0.66	0.38	0.04	0.05	0.06	0.06
	100	0.35	0.35	0.75	0.33	0.33	0.15	0.29	0.29	0.02	0.97	0.47	0.33	0	0	0.01
	500	1.17	1.11	2.27	1.16	1.1	1.5	1.28	1.24	1.1	2.43	1.13	0.98	0	0	0
	1000	0.57	0.52	2.44	0.56	0.51	1.79	0.47	0.43	1.25	2.54	1	0.83	0	0	0
	5000	0.22	0.22	2.5	0.22	0.22	2.1	0.31	0.31	2.22	2.54	0.73	0.8	0	0	0
	10000	0.16	0.16	2.47	0.16	0.16	2.16	0.27	0.27	2.67	2.49	0.67	0.87	0	0	0
f_{14}	50	0.25	0.22	0.43	0.23	0.22	0.12	0.28	0.29	0	0.62	0.29	0.32	0.05	0.08	0.06
	100	0.23	0.21	0.63	0.24	0.23	0.25	0.42	0.38	0.15	0.85	0.29	0.31	0	0.02	0.01
	500	0.04	0.01	1.43	0.02	0	0.63	0.34	0.33	0.41	1.59	0.24	0.07	0.6	0.22	0.17
	1000	0.03	0	1.71	0.03	0	1	0.38	0.36	0.79	1.8	0.26	0.06	0.66	0.08	0.03
	5000	0	0	2.28	0	0	1.85	0.45	0.46	2.09	2.32	0.52	0.74	1.41	0	0
	10000	0	0	2.31	0	0	1.98	0.2	0.2	2.54	2.34	0.5	0.88	1.79	0	0
f_{15}	50	0.08	0	0.1	0.32	0.28	0.08	0.44	0.4	0.11	0.27	0.01	0.16	0.01	0.03	0
	100	0.17	0.13	0.61	0.14	0.12	0	0.57	0.55	0.31	0.81	0.34	0.17	0.19	0.17	0.14
	500	0.11	0	1.43	0.11	0	0.49	0.81	0.77	0.7	1.58	0.32	0.41	0.36	0.19	0.09
	1000	0.03	0	1.74	0.02	0	0.93	0.65	0.64	0.94	1.84	0.43	0.5	0.52	0.1	0.05
	5000	0.25	0.25	1.69	0.25	0.25	1.25	0.56	0.56	1.39	1.73	0.41	0.47	0.14	0	0.01
	10000	0.44	0.42	1.64	0.44	0.42	1.33	0.71	0.69	1.77	1.67	0.39	0.52	0	0.03	0.04
f_{16}	50	0.08	0	0.27	0.81	0.79	0.62	0.62	0.59	0.37	0.42	0.17	0.07	0.14	0.11	0.07
	100	0.12	0.08	0.65	0.06	0.06	0	0.72	0.73	0.63	0.86	0.33	0.14	0.36	0.2	0.18
	500	0.01	0	1.14	0.22	0.21	0.4	0.83	0.82	0.7	1.29	0.13	0.06	0.63	0.01	0
	1000	0	0.02	1.38	0.12	0.14	0.58	0.55	0.56	0.65	1.48	0.21	0.2	0.77	0	0.02
	5000	0.04	0	1.65	0.03	0	1.17	0.35	0.32	1.33	1.69	0.22	0.75	1.3	0.04	0
	10000	0.03	0	1.83	0.03	0	1.49	0.27	0.24	1.9	1.86	0.34	1.1	1.53	0.03	0

Table 5: Binary logarithms of relative L_1 Risks w.r.t best method.

	n	B	R	CV	B′	R′	CV′	B″	R″	CV″	AIC	BIC	TS	BR	B*	R*
f_1	50	0.637	0.05	0.956	0.047	0.047	0.154	0.043	0.043	0.101	9.587	8.104	0.062	0.056	0.639	0.059
	100	0.191	0.022	0.646	0.022	0.022	0.126	0.021	0.021	0.07	7.928	5.767	0.028	0.027	0.191	0.028
	500	0.077	0.004	0.457	0.004	0.004	0.078	0.004	0.004	0.038	3.855	1.802	0.006	0.005	0.078	0.006
	1000	0.019	0.002	0.293	0.002	0.002	0.059	0.002	0.002	0.033	1.767	0.517	0.003	0.003	0.019	0.003
	5000	0.002	0	0.083	0	0	0.02	0	0	0.023	0.328	0.029	0.001	0.001	0.002	0.001
	10000	0.001	0	0.046	0	0	0.012	0	0	0.02	0.145	0.009	0	0	0.001	0
f_2	50	0.803	0.053	0.178	0.053	0.052	0.041	0.063	0.061	0.038	2.04	1.678	0.058	0.043	0.078	0.044
	100	0.226	0.039	0.127	0.04	0.039	0.027	0.043	0.043	0.025	1.965	1.43	0.025	0.03	0.031	0.031
	500	0.034	0.017	0.115	0.017	0.017	0.011	0.018	0.018	0.009	0.779	0.336	0.009	0.012	0.012	0.012
	1000	0.017	0.012	0.07	0.011	0.012	0.007	0.012	0.012	0.006	0.343	0.093	0.005	0.008	0.008	0.008
	5000	0.005	0.005	0.02	0.005	0.005	0.003	0.004	0.005	0.002	0.047	0.009	0.002	0.003	0.003	0.003
	10000	0.003	0.003	0.011	0.003	0.003	0.002	0.003	0.003	0.002	0.02	0.005	0.001	0.002	0.002	0.002
f_3	50	0.958	0.051	0.185	0.053	0.051	0.042	0.054	0.053	0.033	2.773	2.348	0.059	0.04	0.187	0.041
	100	0.36	0.031	0.152	0.032	0.031	0.029	0.034	0.033	0.022	3.047	1.999	0.026	0.024	0.025	0.024
	500	0.026	0.013	0.117	0.014	0.013	0.015	0.014	0.014	0.009	0.941	0.437	0.007	0.008	0.008	0.008
	1000	0.012	0.009	0.08	0.009	0.009	0.011	0.009	0.009	0.006	0.445	0.152	0.004	0.005	0.005	0.005
	5000	0.004	0.004	0.021	0.004	0.004	0.004	0.004	0.004	0.003	0.063	0.011	0.001	0.002	0.002	0.002
	10000	0.002	0.002	0.012	0.002	0.002	0.003	0.002	0.002	0.003	0.026	0.004	0.001	0.001	0.001	0.001
f_4	50	1.83	0.069	0.396	0.084	0.084	0.07	0.08	0.08	0.063	3.658	3.096	0.049	0.067	1.821	0.067
	100	0.232	0.047	0.225	0.055	0.056	0.044	0.051	0.052	0.039	2.68	1.864	0.03	0.046	0.193	0.046
	500	0.025	0.025	0.151	0.026	0.028	0.023	0.026	0.028	0.021	0.932	0.293	0.017	0.027	0.025	0.025
	1000	0.016	0.017	0.103	0.023	0.024	0.02	0.022	0.024	0.018	0.412	0.081	0.014	0.023	0.017	0.019
	5000	0.006	0.007	0.028	0.013	0.015	0.012	0.01	0.012	0.008	0.052	0.01	0.003	0.019	0.006	0.007
	10000	0.004	0.005	0.015	0.008	0.009	0.007	0.006	0.007	0.005	0.021	0.005	0.002	0.011	0.004	0.005
f_5	50	0.472	0.059	0.24	0.158	0.158	0.158	0.129	0.129	0.128	1.971	1.59	0.065	0.109	0.472	0.059
	100	0.216	0.048	0.146	0.147	0.147	0.146	0.112	0.113	0.11	1.624	1.059	0.038	0.057	0.216	0.048
	500	0.035	0.02	0.11	0.045	0.046	0.041	0.053	0.054	0.048	0.666	0.261	0.01	0.032	0.035	0.02
	1000	0.019	0.014	0.071	0.023	0.024	0.019	0.019	0.02	0.016	0.312	0.076	0.006	0.016	0.019	0.014
	5000	0.005	0.006	0.02	0.006	0.006	0.004	0.005	0.006	0.003	0.043	0.009	0.002	0.006	0.005	0.006
	10000	0.004	0.004	0.01	0.003	0.004	0.002	0.003	0.004	0.002	0.018	0.005	0.001	0.003	0.004	0.004
f_6	50	0.336	0.04	0.148	0.042	0.04	0.044	0.04	0.039	0.032	2.302	1.85	0.051	0.04	0.305	0.04
	100	0.097	0.026	0.138	0.027	0.026	0.032	0.03	0.029	0.022	1.924	1.348	0.03	0.028	0.032	0.027
	500	0.021	0.013	0.111	0.013	0.013	0.015	0.013	0.013	0.01	0.811	0.352	0.009	0.01	0.01	0.01
	1000	0.012	0.009	0.075	0.009	0.009	0.011	0.009	0.009	0.007	0.409	0.136	0.006	0.006	0.006	0.006
	5000	0.004	0.003	0.021	0.003	0.003	0.004	0.003	0.003	0.004	0.065	0.01	0.001	0.002	0.002	0.002
	10000	0.002	0.002	0.012	0.002	0.002	0.003	0.002	0.002	0.003	0.027	0.004	0.001	0.001	0.001	0.001
f_7	50	0.244	0.081	0.254	0.086	0.081	0.08	0.094	0.092	0.074	2.447	1.931	0.115	0.085	0.144	0.082
	100	0.202	0.056	0.214	0.055	0.055	0.066	0.066	0.065	0.063	2.892	2.111	0.066	0.064	0.134	0.061
	500	0.068	0.044	0.162	0.044	0.043	0.029	0.045	0.044	0.029	1.182	0.535	0.033	0.031	0.031	0.031
	1000	0.042	0.03	0.114	0.03	0.03	0.022	0.039	0.039	0.017	0.592	0.208	0.018	0.02	0.02	0.02
	5000	0.012	0.011	0.033	0.011	0.011	0.008	0.011	0.011	0.007	0.081	0.021	0.005	0.007	0.007	0.007
	10000	0.008	0.008	0.018	0.007	0.008	0.005	0.008	0.008	0.005	0.035	0.01	0.003	0.005	0.005	0.005
f_8	50	0.591	0.108	0.211	0.107	0.105	0.092	0.108	0.106	0.087	2.31	1.927	0.119	0.095	0.144	0.096
	100	0.633	0.08	0.171	0.081	0.077	0.062	0.095	0.089	0.065	2.501	1.896	0.085	0.064	0.07	0.064
	500	0.076	0.033	0.12	0.036	0.035	0.031	0.04	0.038	0.024	0.834	0.386	0.028	0.026	0.026	0.026
	1000	0.041	0.022	0.082	0.024	0.023	0.021	0.026	0.025	0.017	0.419	0.161	0.017	0.018	0.018	0.018
	5000	0.01	0.009	0.022	0.009	0.009	0.008	0.009	0.009	0.007	0.06	0.019	0.005	0.007	0.007	0.007
	10000	0.006	0.006	0.012	0.006	0.006	0.005	0.006	0.006	0.005	0.023	0.008	0.003	0.005	0.005	0.005
f_9	50	2.944	0.293	0.585	0.283	0.281	0.25	0.279	0.277	0.249	7.391	6.314	0.301	0.264	1.032	0.272
	100	1.695	0.162	0.539	0.169	0.153	0.122	0.256	0.249	0.202	6.173	4.562	0.159	0.158	0.773	0.152
	500	0.228	0.065	0.351	0.064	0.063	0.054	0.108	0.107	0.087	2.823	1.417	0.045	0.062	0.13	0.06
	1000	0.06	0.042	0.245	0.042	0.042	0.039	0.071	0.07	0.056	1.301	0.45	0.028	0.047	0.056	0.041
	5000	0.017	0.017	0.075	0.017	0.017	0.017	0.027	0.027	0.024	0.214	0.044	0.009	0.019	0.017	0.016
	10000	0.011	0.011	0.041	0.011	0.011	0.011	0.017	0.017	0.017	0.09	0.019	0.006	0.013	0.011	0.011
f_{10}	50	2.154	0.626	1.814	1.349	1.349	1.348	1.343	1.342	1.342	7.77	5.867	0.514	1.281	2.154	1.279
	100	0.712	0.152	1.07	1.347	1.347	1.347	1.317	1.317	1.317	6.412	3.41	0.272	1.203	0.712	0.152
	500	0.042	0.029	0.604	1.245	1.245	1.246	1.178	1.178	1.179	2.212	0.367	0.062	1.009	0.042	0.029
	1000	0.017	0.014	0.414	1.244	1.244	1.245	1.092	1.092	1.092	1.095	0.149	0.033	0.819	0.017	0.014
	5000	0.003	0.003	0.125	1.244	1.244	1.244	0.633	0.633	0.633	0.189	0.015	0.008	0.004	0.003	0.003
	10000	0.001	0.001	0.07	1.244	1.244	1.244	0.145	0.145	0.145	0.091	0.006	0.004	0.002	0.001	0.001
f_{11}	50	0.054	0.019	0.072	0.018	0.018	0.023	0.018	0.018	0.021	0.474	0.388	0.064	0.019	0.054	0.019
	100	0.024	0.017	0.051	0.017	0.017	0.022	0.017	0.017	0.019	0.516	0.392	0.04	0.017	0.025	0.018
	500	0.032	0.013	0.031	0.015	0.013	0.01	0.016	0.016	0.009	0.209	0.107	0.012	0.008	0.008	0.008
	1000	0.016	0.008	0.021	0.008	0.008	0.007	0.009	0.008	0.005	0.103	0.044	0.007	0.005	0.005	0.005
	5000	0.004	0.003	0.006	0.003	0.003	0.003	0.003	0.003	0.002	0.014	0.007	0.002	0.002	0.002	0.002
	10000	0.003	0.002	0.003	0.002	0.002	0.002	0.002	0.002	0.002	0.005	0.003	0.001	0.001	0.001	0.001
f_{12}	50	2.336	0.319	2.019	0.26	0.261	0.293	0.256	0.257	0.273	9.862	8.235	0.273	0.274	2.327	0.316
	100	0.87	0.247	0.898	0.221	0.221	0.216	0.218	0.219	0.187	7.819	5.291	0.226	0.212	0.802	0.235
	500	0.268	0.098	0.58	0.093	0.093	0.111	0.093	0.093	0.08	4.104	2.071	0.081	0.094	0.183	0.097
	1000	0.125	0.072	0.388	0.07	0.069	0.082	0.072	0.071	0.058	2.227	0.916	0.055	0.066	0.102	0.069
	5000	0.033	0.028	0.133	0.028	0.028	0.034	0.028	0.028	0.03	0.423	0.09	0.022	0.029	0.031	0.029
	10000	0.021	0.019	0.079	0.019	0.019	0.022	0.019	0.019	0.024	0.198	0.039	0.015	0.02	0.02	0.02
f_{13}	50	0.972	0.232	1.094	0.223	0.224	0.214	0.215	0.216	0.202	12.409	10.751	0.282	0.21	0.925	0.213
	100	0.626	0.193	0.634	0.19	0.187	0.162	0.185	0.183	0.125	9.248	6.797	0.192	0.125	0.133	0.126
	500	0.144	0.057	0.534	0.059	0.056	0.09	0.066	0.064	0.061	4.844	2.575	0.047	0.015	0.015	0.015
	1000	0.066	0.021	0.353	0.021	0.021	0.069	0.017	0.017	0.037	2.171	0.766	0.023	0.008	0.008	0.008
	5000	0.007	0.004	0.1	0.004	0.004	0.025	0.005	0.005	0.025	0.34	0.037	0.005	0.001	0.001	0.001
	10000	0.003	0.002	0.056	0.002	0.002	0.015	0.003	0.003	0.022	0.159	0.015	0.003	0.001	0.001	0.001
f_{14}	50	6.292	0.305	1.433	0.295	0.293	0.297	0.293	0.294	0.257	15.627	13.875	0.341	0.269	5.101	0.276
	100	1.732	0.246	0.732	0.238	0.236	0.239	0.259	0.252	0.218	9.845	7.246	0.255	0.196	1.653	0.206
	500	0.3	0.061	0.572	0.06	0.059	0.097	0.097	0.095	0.091	5.019	2.672	0.063	0.118	0.185	0.076
	1000	0.067	0.031	0.385	0.031	0.031	0.073	0.058	0.057	0.067	2.282	0.723	0.027	0.055	0.069	0.032
	5000	0.009	0.005	0.115	0.005	0.005	0.027	0.017	0.017	0.033	0.391	0.043	0.007	0.021	0.009	0.005
	10000	0.003	0.002	0.065	0.002	0.002	0.017	0.005	0.005	0.024	0.178	0.016	0.004	0.018	0.003	0.002
f_{15}	50	4.35	0.615	1.3	0.828	0.795	0.672	0.903	0.87	0.66	14.492	11.245	0.795	0.609	2.28	0.607
	100	3.325	0.31	1.25	0.3	0.297	0.261	0.527	0.515	0.408	14.552	10.072	0.319	0.361	1.4	0.329
	500	0.174	0.064	0.914	0.069	0.064	0.114	0.215	0.212	0.189	6.307	2.613	0.089	0.124	0.108	0.083
	1000	0.04	0.031	0.599	0.032	0.031	0.081	0.108	0.109	0.108	3.163	0.84	0.048	0.077	0.049	0.039
	5000	0.009	0.008	0.158	0.008	0.008	0.032	0.028	0.028	0.043	0.458	0.049	0.012	0.02	0.011	0.01
	10000	0.005	0.005	0.079	0.005	0.005	0.019	0.018	0.018	0.032	0.189	0.018	0.007	0.01	0.006	0.006
f_{16}	50	5.351	0.63	1.701	1.219	1.188	1.026	1.052	1.024	0.829	17.715	13.894	0.781	0.652	3.98	0.637
	100	3.881	0.344	1.54	0.324	0.326	0.315	0.68	0.684	0.589	16.833	11.568	0.373	0.459	3.295	0.378
	500	0.32	0.114	1.124	0.178	0.178	0.2	0.368	0.366	0.294	7.512	3.007	0.128	0.255	0.322	0.115
	1000	0.073	0.057	0.722	0.074	0.075	0.105	0.152	0.153	0.139	3.662	0.941	0.067	0.13	0.073	0.057
	5000	0.012	0.01	0.19	0.01	0.01	0.035	0.029	0.029	0.046	0.534	0.059	0.02	0.038	0.012	0.01
	10000	0.005	0.005	0.098	0.005	0.005	0.022	0.015	0.015	0.032	0.222	0.02	0.012	0.018	0.005	0.005

Table 6: L_2 Risks.

	n	B	R	CV	B′	R′	CV′	B″	R″	CV″	AIC	BIC	TS	BR	B*	R*
f_1	50	3.88	0.22	4.46	0.11	0.12	1.83	0	0	1.22	7.79	7.55	0.53	0.38	3.88	0.44
	100	3.2	0.09	4.96	0.07	0.09	2.6	0	0.03	1.76	8.58	8.12	0.44	0.38	3.2	0.41
	500	4.28	0.15	6.85	0.03	0.06	4.29	0	0.02	3.26	9.93	8.83	0.68	0.46	4.31	0.5
	1000	3.19	0.13	7.17	0.05	0.05	4.86	0	0.02	4.01	9.76	7.99	0.44	0.39	3.23	0.47
	5000	1.98	0	7.58	0	0	5.56	0	0	5.71	9.56	6.04	0.44	0.43	1.89	0.43
	10000	1.31	0	7.8	0	0	5.88	0	0	6.63	9.47	5.39	0.66	0.62	1.6	0.62
f_2	50	4.41	0.49	2.24	0.5	0.46	0.11	0.73	0.68	0	5.76	5.47	0.62	0.17	1.04	0.22
	100	3.2	0.69	2.37	0.7	0.67	0.13	0.8	0.81	0	6.32	5.86	0.02	0.3	0.33	0.35
	500	1.99	1.02	3.74	0.99	1.01	0.32	1.05	1.1	0.13	6.5	5.29	0	0.52	0.53	0.52
	1000	1.69	1.17	3.73	1.12	1.16	0.43	1.15	1.22	0.2	6.02	4.13	0	0.61	0.61	0.61
	5000	1.39	1.39	3.46	1.31	1.39	0.71	1.3	1.39	0.41	4.71	2.34	0	0.69	0.69	0.69
	10000	1.43	1.51	3.28	1.38	1.51	0.81	1.37	1.49	0.6	4.16	2.03	0	0.78	0.78	0.78
f_3	50	4.87	0.63	2.5	0.7	0.63	0.35	0.72	0.69	0	6.4	6.16	0.84	0.29	2.51	0.31
	100	4.04	0.51	2.8	0.57	0.5	0.43	0.65	0.58	0	7.12	6.51	0.27	0.13	0.18	0.16
	500	1.99	1.02	4.15	1.05	1.02	1.16	1.1	1.04	0.44	7.16	6.05	0	0.32	0.32	0.32
	1000	1.58	1.22	4.32	1.24	1.22	1.46	1.22	1.19	0.69	6.81	5.25	0	0.42	0.42	0.42
	5000	1.74	1.5	4.08	1.52	1.5	1.8	1.51	1.5	1.45	5.65	3.12	0	0.45	0.45	0.45
	10000	1.65	1.6	3.9	1.6	1.6	1.81	1.58	1.57	1.83	5.1	2.51	0	0.49	0.49	0.49
f_4	50	5.23	0.5	3.02	0.78	0.79	0.53	0.71	0.72	0.37	6.23	5.99	0	0.45	5.22	0.46
	100	2.97	0.65	2.92	0.88	0.91	0.58	0.77	0.81	0.4	6.5	5.97	0	0.64	2.7	0.64
	500	0.61	0.56	3.18	0.67	0.75	0.45	0.63	0.74	0.32	5.81	4.14	0	0.67	0.61	0.58
	1000	0.19	0.26	2.87	0.69	0.76	0.5	0.67	0.76	0.37	4.88	2.53	0	0.74	0.24	0.4
	5000	1.01	1.38	3.29	2.26	2.39	2.13	1.89	2.09	1.59	4.22	1.8	0	2.77	1.01	1.38
	10000	1.07	1.53	3.1	2.17	2.34	2	1.82	2.08	1.42	3.66	1.55	0	2.69	1.07	1.53
f_5	50	3	0	2.03	1.43	1.43	1.43	1.13	1.13	1.12	5.06	4.75	0.15	0.89	3	0
	100	2.5	0.34	1.94	1.94	1.94	1.94	1.56	1.56	1.53	5.41	4.8	0	0.57	2.5	0.34
	500	1.79	0.99	3.44	2.15	2.17	2.01	2.38	2.4	2.23	6.03	4.68	0	1.64	1.79	0.99
	1000	1.66	1.27	3.58	1.97	1.98	1.69	1.69	1.73	1.43	5.71	3.66	0	1.4	1.66	1.27
	5000	1.54	1.61	3.39	1.57	1.68	1.21	1.46	1.65	0.65	4.53	2.23	0	1.56	1.54	1.61
	10000	1.61	1.68	3.17	1.55	1.68	1.08	1.56	1.72	0.64	3.94	1.98	0	1.55	1.61	1.68
f_6	50	3.41	0.33	2.23	0.4	0.32	0.47	0.33	0.29	0	6.19	5.87	0.69	0.33	3.27	0.33
	100	2.15	0.25	2.66	0.31	0.25	0.56	0.46	0.39	0	6.45	5.94	0.47	0.33	0.55	0.31
	500	1.15	0.51	3.59	0.54	0.51	0.72	0.55	0.5	0.05	6.46	5.25	0	0.1	0.1	0.11
	1000	1	0.57	3.6	0.6	0.57	0.85	0.62	0.59	0.14	6.05	4.46	0	0.06	0.06	0.07
	5000	1.28	1.19	3.84	1.21	1.19	1.58	1.17	1.15	1.28	5.45	2.83	0	0.55	0.55	0.55
	10000	1.48	1.42	3.71	1.44	1.42	1.67	1.44	1.42	1.72	4.94	2.35	0	0.62	0.62	0.62
f_7	50	1.72	0.13	1.78	0.21	0.12	0.11	0.34	0.31	0	5.04	4.7	0.64	0.19	0.96	0.14
	100	1.89	0.03	1.97	0.02	0	0.28	0.26	0.24	0.2	5.72	5.27	0.26	0.23	1.3	0.15
	500	1.21	0.58	2.46	0.57	0.55	0	0.61	0.6	0	5.33	4.19	0.16	0.06	0.08	0.09
	1000	1.29	0.81	2.72	0.78	0.8	0.32	1.18	1.18	0	5.1	3.59	0.05	0.18	0.18	0.18
	5000	1.23	1.18	2.71	1.17	1.18	0.65	1.15	1.15	0.41	4.01	2.03	0	0.51	0.51	0.51
	10000	1.37	1.34	2.59	1.31	1.34	0.79	1.36	1.39	0.62	3.53	1.74	0	0.59	0.59	0.59
f_8	50	2.77	0.31	1.28	0.3	0.27	0.08	0.32	0.29	0	4.73	4.47	0.45	0.13	0.73	0.14
	100	3.35	0.36	1.47	0.39	0.31	0	0.61	0.53	0.07	5.34	4.94	0.46	0.04	0.17	0.06
	500	1.65	0.45	2.3	0.55	0.51	0.33	0.72	0.65	0	5.1	3.99	0.21	0.07	0.07	0.07
	1000	1.25	0.37	2.25	0.47	0.42	0.28	0.57	0.52	0	4.6	3.22	0	0.06	0.06	0.06
	5000	1.02	0.85	2.13	0.95	0.92	0.67	0.94	0.91	0.42	3.61	1.95	0	0.49	0.49	0.49
	10000	1.22	1.12	2.06	1.17	1.14	0.82	1.16	1.12	0.74	3.09	1.59	0	0.82	0.82	0.82
f_9	50	3.56	0.23	1.23	0.18	0.17	0.01	0.16	0.15	0	4.89	4.66	0.28	0.09	2.05	0.13
	100	3.8	0.41	2.14	0.47	0.32	0	1.07	1.03	0.73	5.66	5.22	0.39	0.37	2.66	0.31
	500	2.35	0.53	2.97	0.52	0.49	0.29	1.28	1.26	0.96	5.98	4.99	0	0.48	1.54	0.43
	1000	1.11	0.61	3.14	0.6	0.58	0.48	1.35	1.34	1.01	5.55	4.02	0	0.76	1.01	0.57
	5000	0.89	0.82	3	0.81	0.81	0.82	1.49	1.49	1.33	4.5	2.2	0	1	0.81	0.78
	10000	0.92	0.85	2.78	0.85	0.85	0.85	1.51	1.5	1.5	3.92	1.65	0	1.12	0.86	0.82
f_{10}	50	2.07	0.28	1.82	1.39	1.39	1.39	1.39	1.39	1.39	3.92	3.51	0	1.32	2.07	1.32
	100	2.23	0	2.82	3.15	3.15	3.15	3.12	3.12	3.12	5.4	4.49	0.84	2.99	2.23	0
	500	0.55	0	4.4	5.44	5.44	5.45	5.37	5.37	5.37	6.27	3.68	1.13	5.14	0.55	0
	1000	0.21	0	4.84	6.43	6.43	6.43	6.24	6.24	6.24	6.24	3.37	1.2	5.82	0.21	0
	5000	0.03	0	5.46	8.78	8.78	8.78	7.8	7.8	7.8	6.06	2.39	1.54	0.4	0.03	0
	10000	0.01	0	5.65	9.79	9.79	9.79	6.69	6.69	6.69	6.02	2.14	1.68	0.43	0.01	0
f_{11}	50	1.6	0.06	2	0.02	0.03	0.38	0	0	0.23	4.72	4.43	1.84	0.06	1.59	0.08
	100	0.54	0.04	1.61	0	0.01	0.39	0	0	0.21	4.95	4.56	1.27	0.04	0.56	0.07
	500	2.01	0.68	1.95	0.9	0.73	0.34	1.03	1.03	0.14	4.72	3.76	0.62	0	0	0
	1000	1.64	0.66	2.03	0.67	0.65	0.48	0.88	0.72	0.1	4.35	3.14	0.39	0	0	0
	5000	1.24	0.97	1.68	1.01	0.96	0.56	1.01	0.96	0.32	3.04	2	0.08	0	0	0
	10000	1.22	1.08	1.46	1.12	1.07	0.69	1.14	1.08	0.51	2.32	1.49	0	0.01	0.01	0.01
f_{12}	50	3.19	0.32	2.98	0.02	0.03	0.19	0	0	0.09	5.27	5.01	0.09	0.1	3.18	0.3
	100	2.22	0.4	2.26	0.24	0.24	0.21	0.22	0.23	0	5.39	4.82	0.27	0.18	2.1	0.33
	500	1.75	0.29	2.87	0.23	0.23	0.48	0.22	0.22	0	5.69	4.7	0.02	0.24	1.2	0.29
	1000	1.19	0.38	2.82	0.34	0.34	0.58	0.38	0.37	0.08	5.34	4.06	0	0.26	0.89	0.33
	5000	0.56	0.34	2.58	0.34	0.34	0.59	0.34	0.34	0.44	4.25	2.01	0	0.36	0.45	0.37
	10000	0.45	0.31	2.38	0.3	0.3	0.5	0.3	0.3	0.68	3.7	1.34	0	0.38	0.42	0.38
f_{13}	50	2.27	0.2	2.44	0.14	0.15	0.08	0.09	0.09	0	5.94	5.73	0.48	0.06	2.19	0.08
	100	2.33	0.63	2.35	0.61	0.59	0.37	0.57	0.56	0.01	6.21	5.77	0.62	0	0.09	0.02
	500	3.23	1.9	5.13	1.94	1.88	2.56	2.11	2.07	2	8.31	7.39	1.6	0	0	0
	1000	3.13	1.46	5.55	1.5	1.45	3.19	1.2	1.14	2.3	8.17	6.67	1.6	0	0	0
	5000	2.2	1.37	6.13	1.37	1.37	4.15	1.81	1.81	4.16	7.9	4.71	1.77	0	0	0
	10000	1.93	1.4	6.3	1.38	1.39	4.43	2.05	2.05	4.95	7.79	4.38	1.93	0	0	0
f_{14}	50	4.61	0.24	2.48	0.2	0.19	0.2	0.19	0.19	0	5.92	5.75	0.4	0.06	4.31	0.1
	100	3.14	0.32	1.9	0.28	0.27	0.29	0.4	0.36	0.15	5.65	5.21	0.38	0	3.07	0.07
	500	2.35	0.05	3.28	0.03	0	0.72	0.72	0.7	0.63	6.41	5.5	0.09	1	1.65	0.37
	1000	1.31	0.2	3.83	0.22	0.2	1.44	1.1	1.08	1.32	6.4	4.74	0	1.01	1.34	0.24
	5000	0.86	0	4.53	0	0	2.42	1.74	1.74	2.72	6.3	3.11	0.46	2.1	0.86	0
	10000	0.53	0.01	4.76	0	0	2.79	1.17	1.17	3.32	6.21	2.73	0.71	2.91	0.53	0.01
f_{15}	50	2.84	0.02	1.1	0.45	0.39	0.15	0.57	0.52	0.12	4.58	4.21	0.39	0.01	1.91	0
	100	3.67	0.25	2.26	0.2	0.19	0	1.01	0.98	0.64	5.8	5.27	0.29	0.47	2.42	0.34
	500	1.44	0	3.83	0.1	0.01	0.83	1.74	1.72	1.55	6.62	5.35	0.47	0.95	0.76	0.36
	1000	0.36	0.01	4.26	0.02	0	1.38	1.79	1.8	1.78	6.66	4.75	0.63	1.3	0.65	0.32
	5000	0.14	0	4.26	0.01	0	1.94	1.77	1.76	2.37	5.79	2.57	0.53	1.26	0.36	0.21
	10000	0.11	0	4	0.01	0	1.96	1.85	1.84	2.67	5.24	1.82	0.43	1.04	0.27	0.21
f_{16}	50	3.09	0	1.43	0.95	0.91	0.7	0.74	0.7	0.39	4.81	4.46	0.31	0.05	2.66	0.02
	100	3.62	0.13	2.29	0.04	0.05	0	1.11	1.12	0.9	5.74	5.2	0.24	0.54	3.39	0.26
	500	1.49	0	3.3	0.64	0.64	0.81	1.69	1.68	1.36	6.04	4.72	0.16	1.16	1.49	0.01
	1000	0.34	0	3.66	0.36	0.4	0.88	1.41	1.42	1.27	6	4.04	0.22	1.19	0.34	0
	5000	0.3	0.01	4.24	0.03	0	1.8	1.54	1.52	2.19	5.73	2.56	0.99	1.93	0.3	0.01
	10000	0.17	0	4.36	0.02	0	2.18	1.65	1.64	2.74	5.54	2.08	1.32	1.89	0.17	0

Table 7: Binary logarithms of relative L_2 Risks w.r.t best method.

	n	B	R	CV	B′	R′	CV′	B″	R″	CV″	AIC	BIC	TS	BR	B*	R*
f_1	50	1(91)	1(98)	4(18)	1(98)	1(98)	3(38)	1(100)	1(100)	1(41)	11(13)	3(19)	1(98)	1(88)	1(82)	1(87)
	100	1(95)	1(99)	10(15)	1(99)	1(99)	3(26)	1(100)	1(99)	1(28)	21(10)	3(18)	1(99)	1(89)	1(85)	1(89)
	500	1(96)	1(99)	36(8)	1(100)	1(99)	11(18)	1(100)	1(100)	6(17)	67(8)	3(15)	1(99)	1(89)	1(86)	1(89)
	1000	1(98)	1(99)	51(7)	1(100)	1(100)	15(13)	1(100)	1(100)	9(13)	77(7)	3(16)	1(99)	1(89)	1(87)	1(88)
	5000	1(99)	1(100)	69(7)	1(100)	1(100)	22(9)	1(100)	1(100)	34(7)	96(12)	1(29)	1(98)	1(90)	1(89)	1(90)
	10000	1(100)	1(100)	72(7)	1(100)	1(100)	26(8)	1(100)	1(100)	59(6)	97(16)	1(35)	1(98)	1(84)	1(84)	1(84)
f_2	50	3(70)	3(75)	6(26)	3(74)	3(76)	4(64)	3(58)	3(60)	4(45)	12(15)	5(24)	7(13)	3(32)	3(36)	3(40)
	100	3(54)	3(51)	11(16)	3(59)	3(52)	5(53)	3(59)	3(53)	6(41)	24(10)	6(20)	8(22)	5(21)	5(20)	4(20)
	500	7(53)	7(54)	42(9)	7(54)	7(54)	11(33)	7(50)	7(49)	12(27)	71(7)	12(16)	22(12)	15(11)	15(11)	15(11)
	1000	9(47)	9(49)	54(7)	9(49)	9(49)	15(25)	9(46)	9(47)	17(23)	86(8)	14(15)	33(10)	20(11)	20(11)	20(11)
	5000	16(33)	15(34)	80(10)	16(34)	15(34)	30(14)	16(38)	16(37)	39(11)	98(24)	22(16)	71(10)	48(5)	48(5)	48(5)
	10000	20(34)	20(32)	85(14)	20(35)	20(32)	38(14)	20(32)	20(34)	59(8)	98(30)	27(16)	95(8)	61(7)	61(7)	61(7)
f_3	50	1(41)	3(35)	5(21)	1(43)	1(35)	3(51)	1(46)	1(43)	3(45)	11(14)	4(21)	1(48)	3(43)	3(42)	3(41)
	100	3(74)	3(81)	9(15)	3(78)	3(81)	5(48)	3(71)	3(73)	5(39)	21(10)	5(19)	4(24)	4(33)	4(32)	4(30)
	500	5(61)	5(58)	38(8)	5(63)	5(58)	11(23)	5(60)	5(58)	10(21)	63(7)	9(18)	16(14)	10(17)	10(17)	10(17)
	1000	7(56)	7(60)	55(8)	7(59)	7(60)	16(17)	7(55)	7(56)	14(18)	84(8)	11(17)	25(12)	14(14)	14(14)	14(14)
	5000	12(38)	12(39)	75(9)	12(39)	12(39)	31(13)	12(36)	12(37)	38(10)	98(25)	17(19)	51(8)	26(9)	26(9)	26(9)
	10000	15(39)	16(35)	84(12)	15(40)	16(35)	38(14)	15(39)	16(33)	60(7)	98(29)	21(17)	75(8)	37(6)	37(6)	37(6)
f_4	50	3(56)	3(58)	6(21)	2(73)	2(68)	3(81)	2(65)	2(59)	3(66)	12(14)	5(23)	6(25)	5(23)	3(51)	3(48)
	100	3(66)	3(78)	10(16)	3(89)	3(89)	4(66)	3(82)	3(82)	4(45)	23(10)	6(21)	10(20)	6(18)	3(48)	3(49)
	500	5(36)	5(53)	38(9)	5(66)	5(69)	7(32)	5(58)	5(64)	9(27)	66(7)	10(19)	20(15)	13(10)	6(26)	5(22)
	1000	8(51)	7(41)	57(9)	6(47)	6(53)	10(25)	7(45)	6(52)	14(18)	85(9)	14(16)	28(11)	24(7)	8(47)	8(26)
	5000	13(36)	13(36)	78(10)	12(42)	11(35)	21(15)	12(45)	12(39)	28(10)	97(21)	19(15)	59(8)	58(2)	13(36)	13(36)
	10000	17(37)	15(35)	84(11)	15(38)	14(42)	26(14)	15(43)	14(43)	41(8)	97(21)	22(16)	81(8)	189(2)	17(37)	15(35)
f_5	50	3(67)	3(76)	7(27)	2(84)	2(84)	2(84)	3(81)	3(69)	3(50)	11(14)	5(19)	7(22)	12(28)	3(67)	3(76)
	100	5(36)	5(38)	11(15)	2(92)	2(93)	2(80)	3(73)	3(76)	4(46)	22(11)	8(19)	12(16)	21(19)	5(36)	5(38)
	500	9(41)	9(46)	40(8)	6(46)	6(43)	7(58)	6(34)	7(35)	10(44)	63(7)	14(16)	29(12)	22(11)	9(41)	9(46)
	1000	11(38)	11(34)	55(9)	9(43)	9(44)	12(33)	10(40)	10(42)	15(30)	85(8)	16(19)	42(12)	59(5)	11(38)	11(34)
	5000	20(31)	19(32)	79(10)	19(36)	19(34)	28(24)	20(28)	19(31)	33(18)	98(28)	26(18)	86(8)	104(4)	20(31)	19(32)
	10000	24(27)	24(32)	87(13)	24(29)	24(32)	39(18)	25(28)	24(34)	49(14)	98(30)	31(16)	122(7)	138(4)	24(27)	24(30)
f_6	50	1(57)	1(50)	5(22)	1(60)	1(51)	3(38)	1(52)	1(50)	3(42)	11(14)	5(20)	1(71)	1(30)	1(28)	2(32)
	100	3(46)	3(52)	9(17)	3(45)	3(48)	5(33)	2(56)	2(53)	4(38)	18(10)	6(16)	4(24)	3(23)	3(27)	3(36)
	500	4(40)	5(38)	39(9)	4(40)	5(38)	12(27)	4(47)	4(41)	10(23)	63(6)	11(14)	13(14)	11(15)	11(15)	11(15)
	1000	6(52)	6(51)	55(9)	6(55)	6(51)	17(18)	6(54)	6(54)	14(17)	82(7)	11(16)	21(10)	15(13)	15(13)	15(13)
	5000	11(41)	11(38)	77(11)	11(42)	11(38)	30(12)	11(47)	11(42)	35(9)	98(24)	18(15)	55(8)	32(9)	32(9)	32(9)
	10000	14(32)	14(36)	80(13)	14(33)	14(36)	38(12)	14(34)	14(38)	61(7)	98(28)	22(16)	70(9)	44(7)	44(7)	44(7)
f_7	50	2(37)	2(39)	7(26)	2(40)	2(41)	3(49)	2(45)	2(43)	3(46)	12(16)	5(20)	1(42)	3(28)	3(27)	3(30)
	100	3(79)	3(87)	12(18)	3(86)	3(87)	5(35)	3(77)	3(79)	5(34)	22(11)	8(14)	4(25)	4(28)	3(24)	3(42)
	500	4(29)	5(34)	42(10)	4(33)	5(36)	14(36)	4(42)	5(49)	16(21)	69(7)	16(17)	36(10)	22(12)	22(12)	22(12)
	1000	13(19)	14(18)	59(10)	13(20)	14(17)	19(19)	5(36)	5(32)	22(16)	84(9)	18(15)	47(8)	24(9)	24(9)	24(9)
	5000	19(25)	19(28)	84(12)	19(26)	19(29)	40(17)	19(33)	19(37)	52(10)	99(29)	28(14)	100(7)	71(6)	71(6)	71(6)
	10000	27(19)	26(19)	86(13)	27(19)	26(20)	48(13)	25(20)	25(19)	74(8)	99(34)	35(16)	134(7)	108(4)	108(4)	108(4)
f_8	50	1(51)	1(46)	7(22)	1(63)	1(53)	4(50)	1(85)	1(76)	4(41)	14(16)	8(20)	1(53)	1(20)	1(19)	1(19)
	100	3(25)	4(25)	14(18)	3(34)	3(29)	6(48)	1(45)	3(40)	7(34)	25(9)	9(17)	9(10)	9(27)	9(27)	9(27)
	500	11(20)	12(23)	46(11)	10(34)	10(35)	16(28)	9(26)	10(21)	19(24)	69(8)	18(16)	32(8)	31(13)	31(13)	31(12)
	1000	16(21)	15(22)	63(8)	14(34)	14(29)	23(20)	14(23)	14(27)	26(16)	88(9)	23(16)	51(8)	32(11)	32(11)	32(11)
	5000	29(20)	29(24)	84(13)	28(22)	28(24)	47(16)	27(21)	28(22)	63(9)	99(30)	35(12)	118(7)	84(5)	84(5)	84(5)
	10000	36(20)	38(20)	91(15)	36(21)	38(18)	59(14)	35(22)	36(21)	89(6)	99(36)	47(14)	165(5)	180(3)	180(3)	180(3)
f_9	50	1(68)	1(61)	6(24)	1(79)	1(69)	5(38)	1(82)	1(79)	3(46)	11(15)	5(22)	1(61)	1(42)	1(38)	1(39)
	100	5(64)	5(75)	11(18)	5(69)	5(77)	6(52)	1(37)	3(34)	6(32)	22(11)	8(18)	8(17)	11(18)	5(18)	5(31)
	500	7(52)	7(56)	41(8)	7(65)	7(59)	12(24)	7(46)	7(42)	12(25)	68(8)	11(16)	19(14)	11(31)	11(28)	7(28)
	1000	9(52)	9(56)	58(9)	9(57)	9(57)	16(18)	9(35)	9(37)	16(18)	85(9)	14(16)	28(14)	11(28)	9(31)	9(37)
	5000	14(38)	14(43)	79(11)	14(38)	14(43)	34(12)	14(29)	14(28)	41(11)	98(28)	20(16)	54(9)	33(15)	14(24)	14(30)
	10000	17(41)	17(38)	82(12)	17(42)	17(38)	40(12)	18(35)	18(33)	64(7)	99(34)	24(16)	76(8)	33(12)	17(27)	17(26)
f_{10}	50	5(79)	4(51)	7(23)	2(74)	2(86)	2(100)	5(85)	5(96)	5(84)	14(15)	7(20)	7(24)	11(100)	5(79)	11(100)
	100	5(86)	5(100)	13(16)	2(99)	2(100)	2(100)	5(100)	5(100)	5(84)	24(11)	8(18)	9(20)	19(59)	5(86)	5(100)
	500	5(96)	5(100)	42(9)	3(100)	3(100)	3(100)	5(100)	5(100)	5(85)	67(7)	9(16)	11(13)	80(100)	5(96)	5(100)
	1000	5(98)	5(100)	53(8)	3(100)	3(100)	3(61)	5(99)	5(100)	5(47)	77(7)	7(16)	14(14)	141(86)	5(98)	5(100)
	5000	5(100)	5(100)	77(8)	3(100)	3(100)	5(29)	7(100)	7(100)	7(35)	94(12)	5(25)	17(14)	402(100)	5(100)	5(100)
	10000	5(100)	5(100)	79(9)	3(100)	3(100)	5(20)	7(100)	7(100)	9(23)	89(12)	5(36)	17(12)	402(100)	5(100)	5(100)
f_{11}	50	1(82)	1(95)	9(19)	1(97)	1(96)	3(45)	1(99)	1(99)	1(39)	14(15)	4(17)	1(81)	1(87)	1(73)	1(85)
	100	1(87)	1(94)	16(16)	1(99)	1(99)	5(32)	1(100)	1(100)	2(27)	26(10)	8(12)	1(47)	1(86)	1(75)	1(82)
	500	1(31)	17(14)	49(11)	1(39)	1(18)	20(39)	1(91)	1(86)	23(22)	74(9)	23(17)	62(8)	39(12)	39(12)	39(12)
	1000	21(29)	21(32)	66(11)	20(36)	20(33)	31(20)	18(17)	20(25)	35(17)	94(12)	27(15)	82(7)	49(8)	49(8)	49(8)
	5000	40(20)	42(18)	90(17)	40(20)	42(20)	65(17)	38(16)	40(21)	79(10)	99(35)	52(13)	174(6)	99(7)	99(7)	99(7)
	10000	52(13)	53(16)	94(19)	53(14)	53(15)	78(15)	49(13)	52(15)	111(7)	99(39)	63(13)	241(6)	139(7)	139(7)	139(7)
f_{12}	50	1(53)	1(71)	5(19)	1(91)	1(90)	3(62)	1(93)	1(93)	3(50)	11(14)	4(22)	1(93)	1(64)	1(39)	1(54)
	100	2(38)	1(44)	10(17)	1(63)	1(59)	5(28)	1(78)	1(77)	3(44)	21(11)	6(18)	1(70)	1(39)	2(28)	1(29)
	500	3(67)	3(78)	41(9)	3(90)	3(88)	12(21)	3(91)	3(90)	8(22)	65(7)	8(17)	7(21)	7(19)	3(32)	3(38)
	1000	4(37)	4(36)	55(7)	3(39)	3(37)	17(14)	3(50)	3(46)	13(15)	83(8)	10(16)	12(19)	9(14)	5(19)	4(17)
	5000	7(60)	7(62)	72(10)	7(64)	7(64)	27(10)	7(58)	7(59)	35(9)	97(22)	12(18)	26(11)	23(7)	7(17)	7(19)
	10000	9(46)	9(47)	82(10)	9(48)	9(48)	31(11)	9(46)	9(46)	69(7)	99(26)	14(16)	37(11)	27(5)	9(10)	9(12)
f_{13}	50	1(78)	1(79)	5(21)	1(83)	1(79)	3(43)	1(93)	1(92)	3(45)	11(15)	5(19)	1(89)	1(44)	1(42)	1(44)
	100	1(60)	1(52)	10(16)	1(63)	1(53)	5(50)	1(77)	1(71)	5(33)	19(10)	7(18)	1(63)	5(50)	5(50)	5(50)
	500	3(46)	3(39)	40(8)	3(48)	3(39)	12(19)	3(63)	3(57)	8(21)	66(8)	9(15)	7(15)	5(99)	5(99)	5(99)
	1000	5(85)	5(93)	53(8)	5(91)	5(94)	18(17)	5(89)	5(92)	14(15)	83(9)	9(16)	9(15)	5(100)	5(100)	5(100)
	5000	5(98)	5(99)	72(8)	5(100)	5(100)	28(10)	5(100)	5(100)	37(8)	96(20)	7(16)	16(11)	5(100)	5(100)	5(100)
	10000	5(97)	5(99)	78(10)	5(100)	5(99)	30(10)	5(100)	5(100)	64(6)	99(21)	5(21)	19(12)	5(100)	5(100)	5(100)
f_{14}	50	1(56)	1(57)	6(22)	1(64)	1(59)	4(36)	1(83)	1(83)	2(45)	11(13)	5(20)	1(80)	3(39)	3(38)	3(37)
	100	2(55)	2(62)	11(17)	2(69)	2(71)	5(43)	1(56)	2(52)	4(29)	21(10)	7(19)	1(43)	3(70)	3(68)	3(66)
	500	4(68)	4(68)	41(8)	4(70)	4(67)	12(18)	4(66)	4(65)	9(22)	68(7)	8(16)	9(18)	3(29)	4(50)	4(53)
	1000	5(57)	5(63)	54(8)	5(60)	5(63)	17(14)	5(53)	5(57)	12(13)	80(6)	8(15)	12(16)	23(56)	5(54)	5(61)
	5000	5(98)	5(100)	70(8)	5(100)	5(100)	27(11)	5(100)	5(99)	36(8)	97(21)	7(19)	17(12)	23(99)	5(98)	5(100)
	10000	5(98)	5(99)	78(9)	5(100)	5(100)	31(11)	5(100)	5(99)	60(6)	97(21)	5(22)	20(11)	23(85)	5(98)	5(99)
f_{15}	50	4(44)	4(51)	6(24)	3(37)	3(42)	4(54)	1(56)	1(49)	4(51)	11(13)	6(23)	1(24)	6(18)	4(25)	4(27)
	100	4(61)	4(66)	12(16)	4(73)	4(75)	6(40)	4(51)	4(54)	6(37)	24(10)	8(17)	8(15)	9(18)	4(18)	4(26)
	500	7(41)	7(43)	42(8)	7(43)	7(44)	12(24)	7(38)	7(39)	12(21)	69(8)	11(15)	19(12)	10(55)	10(46)	10(41)
	1000	8(75)	8(82)	56(8)	8(80)	8(83)	16(15)	8(53)	8(54)	17(15)	81(7)	11(16)	22(11)	10(67)	10(48)	8(46)
	5000	8(81)	8(80)	76(10)	8(84)	8(81)	29(11)	9(49)	9(46)	45(8)	97(22)	12(16)	36(10)	10(92)	10(58)	10(53)
	10000	8(54)	8(49)	81(10)	8(54)	8(49)	34(11)	9(35)	9(34)	69(7)	98(23)	12(17)	40(9)	10(96)	10(66)	10(64)
f_{16}	50	4(72)	4(78)	6(26)	3(41)	3(42)	4(62)	3(42)	3(44)	4(52)	12(17)	7(21)	7(23)	7(27)	4(40)	4(39)
	100	4(77)	4(82)	11(17)	4(90)	4(84)	5(48)	4(87)	4(85)	6(36)	21(10)	8(18)	8(18)	9(35)	4(51)	4(56)
	500	6(40)	6(47)	43(8)	5(70)	5(71)	12(18)	5(39)	5(37)	13(19)	68(6)	11(16)	15(14)	22(17)	6(40)	6(47)
	1000	7(58)	7(49)	54(9)	7(54)	7(45)	16(17)	7(41)	8(40)	17(15)	84(8)	12(16)	21(11)	49(19)	7(58)	7(49)
	5000	8(44)	8(36)	75(9)	8(46)	8(37)	30(11)	9(38)	9(34)	39(8)	96(20)	12(19)	35(7)	100(77)	8(44)	8(36)
	10000	10(78)	10(86)	79(11)	10(80)	10(86)	33(10)	10(65)	10(70)	71(6)	98(27)	12(20)	41(8)	100(98)	10(78)	10(86)

Table 8: Modes of the number of bins chosen. The numbers in parentheses give the frequency of the mode in percent.