A note on the choice of the number of slices in sliced inverse regression

CLAUDIA BECKER and URSULA GATHER

Corresponding Author: Prof. Dr. Claudia Becker

Address: Faculty of Law, Economics, and Business
School of Economics and Business
Martin-Luther-University Halle-Wittenberg
D-06099 Halle, Germany

Phone: +49 - 345 - 55 23396

Fax: +49 - 345 - 55 27191

E-mail: claudia.becker@wiwi.uni-halle.de

Prof. Dr. Ursula Gather

Address: Department of Statistics
University of Dortmund
D-44221 Dortmund, Germany

Phone: +49 - 231 - 755 3110

Fax: +49 - 231 - 755 5305

E-mail: gather@statistik.uni-dortmund.de

A note on the choice of the number of slices in sliced inverse regression

CLAUDIA BECKER*[1] and URSULA GATHER[2]

Abstract

Sliced inverse regression (SIR) is a clever technique for reducing the dimension of the predictor in regression problems, thus avoiding the curse of dimensionality. There exist many contributions on various aspects of the performance of SIR. Up to now, few attention has been paid to the problem of choosing the number of slices within the SIR procedure appropriately. The aim of this paper is to show that especially the estimation of the reduced dimension can be strongly influenced by the chosen number of slices.

Keywords: Dimension reduction; Estimation of dimension

2000 Mathematics Subject Classification: 62H12

[1]Faculty of Law, Economics, and Business, School of Economics and Business, Martin-Luther-University Halle-Wittenberg, D-06099 Halle, Germany; corresponding author

[2]Department of Statistics, University of Dortmund, D-44221 Dortmund, Germany

1 Introduction

In his fundamental article, [18] Li (1991) proposed the method of sliced inverse regression (SIR) as a means to reduce the dimension of the predictor in a regression setting. The usual regression assumption $Y = g(X_1, \ldots, X_d) + \varepsilon$, for some real-valued random variables $Y, X_1, \ldots, X_d$ and some error ε independent of $\boldsymbol{X} = (X_1, \ldots, X_d)^T$, is replaced by the simpler model

$$Y = f(\boldsymbol{\beta}_1^T \boldsymbol{X}, \ldots, \boldsymbol{\beta}_K^T \boldsymbol{X}, \varepsilon), \tag{1}$$

where $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$ denote the dimension reducing directions, $K \leq d$. If $K < d$, a dimension reduction of the regressor space is achieved, implying that the relevant information on Y only depends on a projection of the original $\boldsymbol{X}$ onto a K-dimensional subspace. While most nonparametric methods for estimating the regression function fail in situations where d is large, due to the so-called 'curse of dimensionality' ([2] Bellman, 1961; [12] Friedman, 1994; [13] Gather and Becker, 2001), they may work well on the reduced space, if K is small enough. Sliced inverse regression provides a method which estimates the space $\mathcal{B}$ spanned by $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$, where we assume that $\mathcal{B}$ equals the central subspace in the sense of [6],[7] Cook (1994, 1996). The main idea is to use the inverse regression curve $E(\boldsymbol{X}|Y)$ which under certain conditions gives information on the space spanned by $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$. Based on a sample $(y_i, \boldsymbol{x}_i^T)^T$, $i = 1, \ldots, n$, a crude estimate of $E(\boldsymbol{X}|Y)$ can be obtained.

Firstly, a so-called slicing of the y-observations is done, then the $\boldsymbol{x}$-data are also splitted into a certain number of disjoint subsets (slices) according to the ranks of corresponding y's. The SIR procedure is described in more detail in Section 2. The performance of SIR has been investigated extensively with respect to various aspects such as its ability of detecting special structures ([8] Cook, 1998; [9] Cook and Weisberg, 1991; Rejoinder in [18] Li, 1991; [19] Li, 1992), its asymptotics ([17] Hsing and Carroll, 1992; [20] Saracco, 1997; [24] Zhu and Fang, 1996; [25] Zhu and Ng, 1995), the determination of the dimension of $\mathcal{B}$ ([3],[4] Bura and Cook, 2001a,b; [10] Cook and Weisberg, 1994, ch. 8; [11] Ferré, 1998; [18] Li, 1991; [21] Schott, 1994; [23] Velilla, 1998), and its robustness ([14],[15] Gather, Hilker, and Becker, 2001, 2002; [22] Sheather and McKean, 1997). Since the dimension K of the reduced space $\mathcal{B}$ in model (1) is usually unknown, estimating K is a crucial part of the procedure. [15] Gather, Hilker, and Becker (2002) show that the procedure proposed by [18] Li (1991) for estimating K can be influenced severely by outliers in the $\boldsymbol{X}$-space.

Less attention has been paid to the question of how many slices should be chosen in the slicing step of the procedure. Most articles consider this as a minor problem, although some authors comment on the choice of the number of slices ([3],[4] Bura and Cook, 2001a,b; [5] Chen and Li, 1998; [10] Cook and Weisberg, 1994; [11] Ferré, 1998; [18] Li, 1991). However, the results

of SIR can be influenced strongly by the number of slices, as is shown by the results of the simulation study described below, where we concentrate on the estimation of the reduced dimension as proposed by [18] Li (1991). This leads to the recommendation to interprete the outcomes of SIR always accompanied by some sort of diagnostics as proposed by [5] Chen and Li (1998), thus checking the correctness of the reduced dimension. On the other hand, this complicates the integration of SIR into an automatic procedure for dimension reduction and estimating the functional relationship. Therefore, a detailed investigation of the effect the number of slices has on estimating the reduced dimension will be helpful in further use of SIR.

This article is organized as follows. In Section 2 we briefly introduce the SIR method together with a procedure to estimate the reduced dimension K, following [18] Li (1991). An example of the variability of estimated values of K is discussed in Section 3. Section 4 contains the results of a simulation study, where the performance of the estimator $\widehat{K}$ is investigated under different choices of the number of slices in the procedure. We finish with some concluding remarks.

2 Sliced inverse regression

As mentioned above, in the regression setting [18] Li (1991) assumes that $Y \in I\!R$ depends on $\boldsymbol{X} \in I\!R^d$ only via K linear combinations $\boldsymbol{\beta}_1^T\boldsymbol{X}, \ldots, \boldsymbol{\beta}_K^T\boldsymbol{X}$, yielding model (1). Here, $\boldsymbol{\beta}_i \in I\!R^d$, $f : I\!R^{K+1} \to I\!R$ and K are unknown. The SIR method yields an estimation of the reduced regressor space $\mathcal{B} = \text{span}[\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K]$, the so-called central subspace. This can be seen as a first step, possibly followed by estimating f, defined on this lower dimensional space (if $K < d$), thus avoiding the curse of dimensionality (see [1] Becker, 2001). Under some conditions on $\boldsymbol{X}$, the centred inverse regression curve $E(\boldsymbol{X}|Y) - \boldsymbol{\mu}$ almost surely falls into $\text{span}[\boldsymbol{\Sigma\beta}_1, \ldots, \boldsymbol{\Sigma\beta}_K]$, where $\boldsymbol{\mu} = E(\boldsymbol{X}), \boldsymbol{\Sigma} = Cov(\boldsymbol{X})$ ([18] Li, 1991). Using the information given by $E(\boldsymbol{X}|Y) - \boldsymbol{\mu}$ leads to estimators of $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$.

Let $(y_i, \boldsymbol{x}_i^T)^T$, $i = 1, \ldots, n$, $\boldsymbol{x}_i \in I\!R^d$, $y_i \in I\!R$ be a sample of $(Y, \boldsymbol{X}^T)^T$. SIR proceeds in the following five steps:

1. Standardize the $\boldsymbol{x}_i$ by $\boldsymbol{z}_i = \widehat{\boldsymbol{\Sigma}}^{-1/2}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})$, $i = 1, \ldots, n$, where $\widehat{\boldsymbol{\Sigma}} = \sum_{i=1}^n (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T/n$, $\overline{\boldsymbol{x}} = \sum_{i=1}^n \boldsymbol{x}_i/n$.

2. Split $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ into H slices S_h, $h = 1, \ldots, H$, according to the size of the corresponding values of $y_1, \ldots, y_n$; let n_h be the number of observations in slice S_h. In the most common version of the procedure, the sample size n is distributed evenly onto the slices: $n_h \approx n/H$.

3. Calculate the slice means: $\widehat{\boldsymbol{m}}_h = \sum_{S_h} \boldsymbol{z}_i / n_h$, $h = 1, \ldots, H$.

4. Carry out a (weighted) principal components analysis for the slice means: $\widehat{\boldsymbol{SIR}} = \sum_{h=1}^{H} n_h \widehat{\boldsymbol{m}}_h \, \widehat{\boldsymbol{m}}_h^T / n$ yielding eigenvalues $\widehat{\lambda}_1 \geq \ldots \geq \widehat{\lambda}_d$, and normalized eigenvectors $\widehat{\boldsymbol{\eta}}_1, \ldots, \widehat{\boldsymbol{\eta}}_d$, respectively.

5. Estimate the dimension reducing directions $\boldsymbol{\beta}_k$ by $\widehat{\boldsymbol{\beta}}_k = \widehat{\boldsymbol{\Sigma}}^{-1/2} \widehat{\boldsymbol{\eta}}_k$, $k = 1, \ldots, K$.

The eigenvectors of $\widehat{\boldsymbol{SIR}}$ corresponding to the K largest eigenvalues yield the estimated directions $\widehat{\boldsymbol{\beta}}_k$. As the dimension K of the reduced regressor space will usually be unknown, it has to be estimated as well. Several authors propose methods for estimating K (e.g. [3], [4] Bura and Cook, 2001a,b; [11] Ferré, 1998; [18] Li, 1991; [21] Schott, 1994; [23] Velilla, 1998). We will use Li's original method here. The dimension K is estimated by successively testing $H_0^k : K = k$ vs. $H_1^k : K > k$, starting with $k = 0$. The number k for which H_0^k is not rejected for the first time gives the estimated dimension $\widehat{K} = k$. The test statistics used in these tests are given by $t_k := n(d-k)\overline{\lambda}_{(d-k)}$, where $\overline{\lambda}_{(d-k)} = \sum_{i=k+1}^{d} \widehat{\lambda}_i / (d-k)$ denotes the mean of the $(d-k)$ smallest eigenvalues of $\widehat{\boldsymbol{SIR}}$. For normally distributed $\boldsymbol{X}$, we have $\overline{\lambda}_{(d-K)} \sim \chi^2_{(d-K)(H-K-1)}$ asymptotically ([18] Li, 1991, p. 321). Thus, H_0^k is rejected if t_k exceeds a certain quantile of $\chi^2_{(d-k)(H-k-1)}$.

Recommendations with respect to the choice of H are barely given. [18] Li

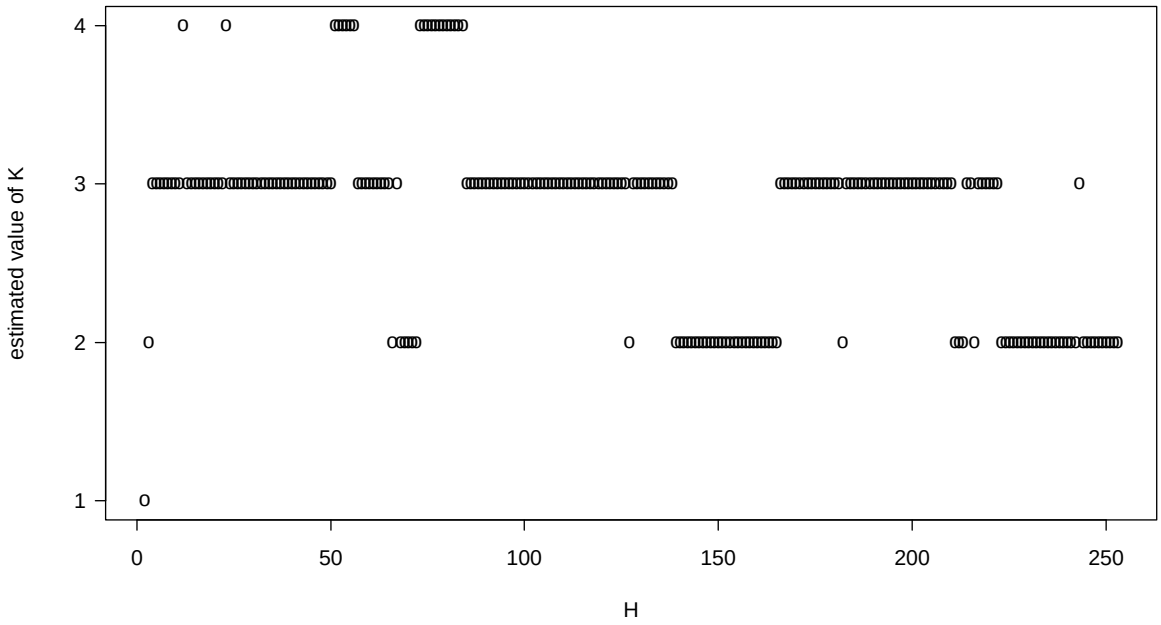
(1991, p. 320) states that "the choice of the number of slices may affect the asymptotic variances of the output estimate. However, the difference is not significant for practical sample sizes in our simulation study." The sample sizes in his simulations are $n = 100$ and $n = 400$. [5] Chen and Li (1998, p. 292) find that "the SIR estimates [...] are not sensitive to the number of slices used." Both comments refer to the estimation of the dimension reducing directions and are not addressing the problem of determining the reduced dimension itself. This topic is treated by [11] Ferré (1998) who assumes that H is chosen "greater than d, to avoid artificial reduction of the dimensionality" (p. 134). He also gives an example where for $d = 10$, $n = 2\,000$, and $H = 10$ his method yields an estimated dimension which is too small, whereas the result improves for $H = 50$. He concludes that "this might be explained by the fact that taking too few slices leads to oversmoothing" (p. 138). This stands in contrast to the comment by [5] Chen and Li (1998, p. 298) who state that "although in theory we can use as many as $H = n/2$ slices [...], practically we find no obvious advantage in using large H." For a modified version of [18] Li's (1991) procedure, [3] Bura and Cook (2001a, p. 1002) find "[...] that the number of slices for the weighted chi-squared test should not be more than 5%-7% of the sample size to keep test levels from being much larger than the nominal level." The same authors see the problem of choosing H appropriately: "Most importantly, SIR can be ambiguous

about the estimate of the dimension as the latter depends sometimes crucially on the choice of the number of slices. As a result, all methods that depend on a tuning constant related to the choice of the number of slices suffer from the same ambiguity in estimation[...]" ([4] Bura and Cook, 2001b, p. 396). In none of these contributions we find a detailed investigation of this problem nor general recommendations. We consider an example.

3 The Boston housing data

Example 1 *The Boston housing data ([16] Harrison and Rubinfeld, 1978) consists of* $d = 13$ *regressor variables concerning housing conditions, and socio-economic and environmental factors in the census tracts of Boston standard metropolitan statistical areas. The dependent variable is the median value of owner-occupied homes. The data set contains* $n = 506$ *observations. Applying SIR to these observations, concentrating only on the estimated value of the reduced dimension* K*, leads to the following results. If we choose* $H = 55$*, we get* $\widehat{K} = 4$*, but for* $H = 57$ *we find* $\widehat{K} = 3$*, whereas taking* 66 *slices yields* $\widehat{K} = 2$*. Figure 1 summarizes the behavior of* $\widehat{K}$ *depending on the choice of* H *for all possibilities of* $H = 2, \ldots, n/2$*. Although for the majority of possible choices of* H *the estimation procedure yields* $\widehat{K} = 3$*, there are also a lot of cases with* $\widehat{K} = 2$*. [5] Chen and Li (1998) report*

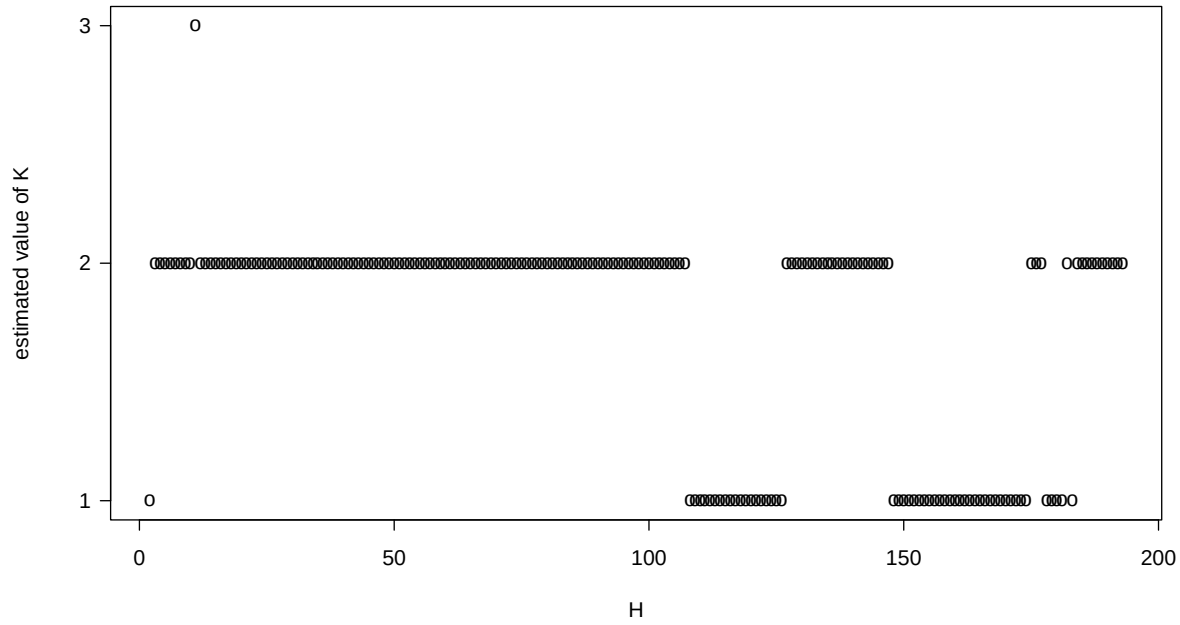
Figure 1: Estimated reduced dimension for the Boston housing data, depending on the choice of H



the application of SIR on this data, mentioning that there is a group of observations that seems to influence the estimation of the dimension reducing directions. Hence, they analyze the data without this group of observations. Following their suggestions, we leave the corresponding observations out and recalculate $\widehat{K}$ *for the various choices of* H*. The results are shown in figure 2. Although we would estimate the reduced dimension to be* 2 *most time, the result is not convincingly unique.*

The example shows that estimating the reduced dimension K may be sensitive to the number of slices chosen in the SIR procedure. In the following section, we discuss the results of a simulation study where the influence of

Figure 2: Estimated reduced dimension for the modified Boston housing data, depending on the choice of H



the choice of H on the estimation of K is investigated in more detail.

4 Studying the influence of H

We investigate the influence of the choice of H on estimating K as follows. For certain choices of n and d, 1 000 data sets are generated according to each of the models specified in table 1. Models A and B were originally used by [18] Li (1991) to illustrate the performance of SIR, model C is taken from [11] Ferré (1998). We use each model in an exact version (A1-C1) as well as in a version including an error term (A2-C2). The results are presented here for the special choice of $d = 5$ and $n = 100, 500$. Using other combinations

Table 1: Simulation models with $\boldsymbol{X} \sim \boldsymbol{N}(\boldsymbol{0}, \mathcal{I})$; $\varepsilon \sim N(0, 0.01)$

Model		K
A1	$Y = X_1$	1
A2	$Y = X_1 + \varepsilon$	1
B1	$Y = X_1/(0.5 + (1.5 + X_2)^2)$	2
B2	$Y = X_1/(0.5 + (1.5 + X_2)^2) + \varepsilon$	2
C1	$Y = X_1 + 0.5X_2 + (X_2 + X_3 + X_4 + X_5)^2$	2
C2	$Y = X_1 + 0.5X_2 + (X_2 + X_3 + X_4 + X_5)^2 + \varepsilon$	2

of n and d essentially leads to the same results. For a given model, we apply the SIR method to each data set for different values of $H \in \{2, \ldots, n/2\}$ and estimate the dimension K, carrying out each test on a 5% level. The results are summarized in figures 3-8 for a sample size of $n = 100$, showing stacked barplots of the counts of $\widehat{K} = 0, \ldots, 4$ for each choice of H; bars belonging to the true K are shaded in black.

First of all, we can see that the ability of SIR to find the true dimension varies strongly with the structure of the functional relationship. In model A, the results are quite satisfying: we get $\widehat{K} = 1$ for almost all of the simulated data sets. For model B, the procedure still estimates the true value of K in most cases. In contrast to this, the results are less convincing for model C,

Figure 3: Results for model A1 for $n = 100$: Counts of $\widehat{K} = 0, \ldots, 4$ in 1 000 simulated samples for $H = 2, \ldots, n/2$; true $K = 1$

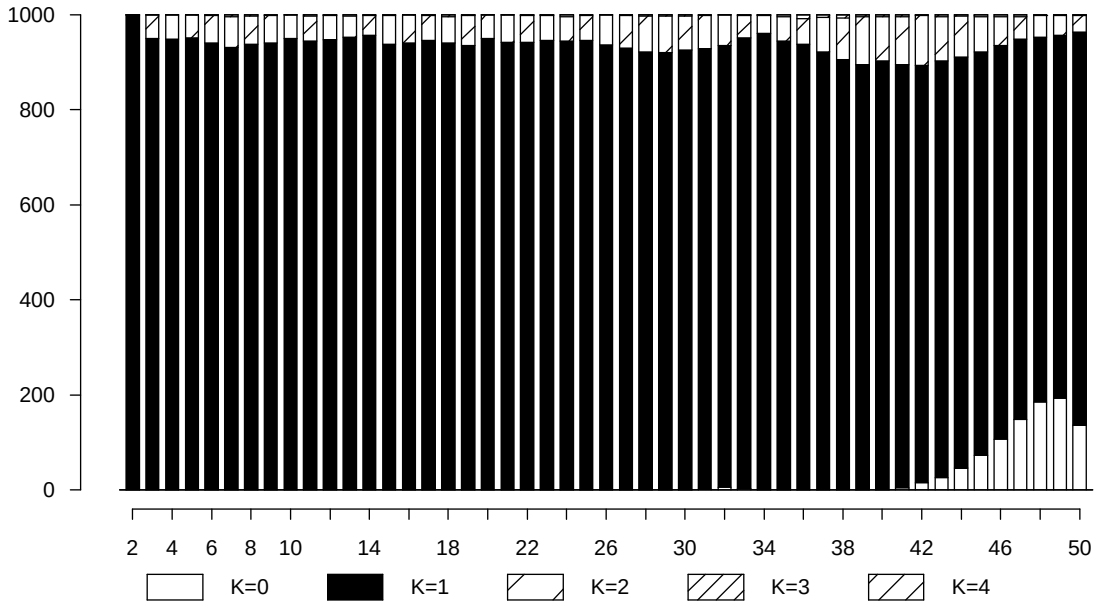


Figure 4: Results for model A2 for $n = 100$: Counts of $\widehat{K} = 0, \ldots, 4$ in 1 000 simulated samples for $H = 2, \ldots, n/2$; true $K = 1$

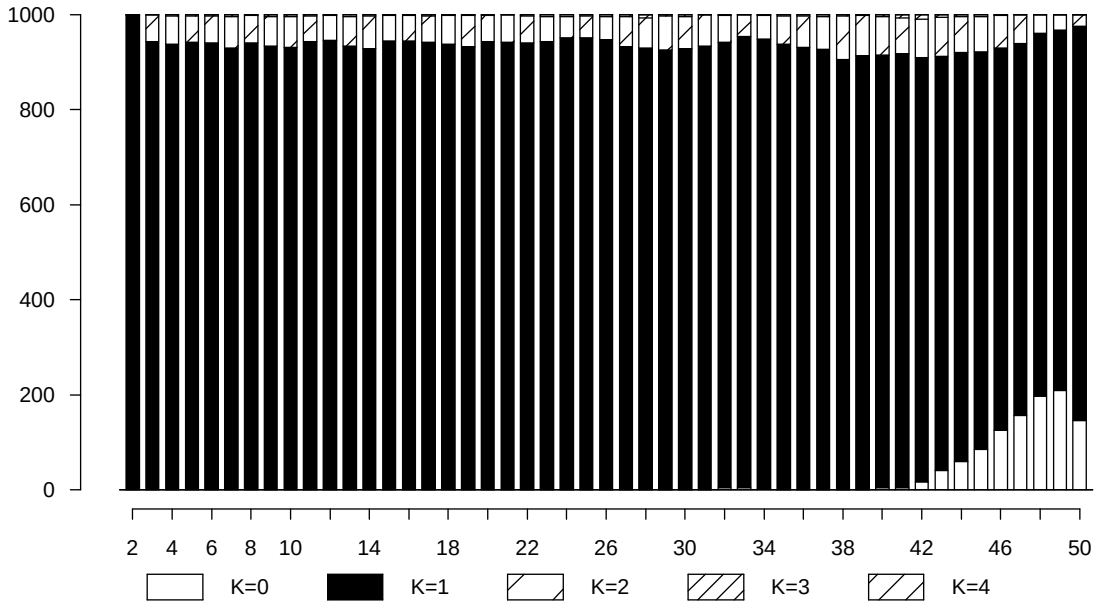


Figure 5: Results for model B1 for $n = 100$: Counts of $\widehat{K} = 0, \ldots, 4$ in 1 000 simulated samples for $H = 2, \ldots, n/2$; true $K = 2$

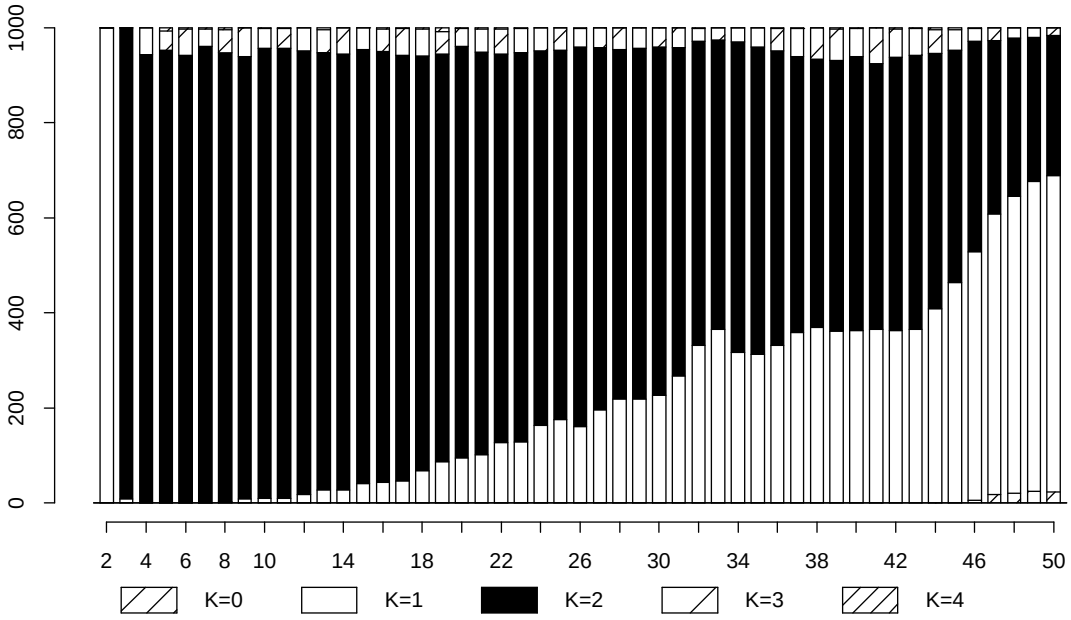


Figure 6: Results for model B2 for $n = 100$: Counts of $\widehat{K} = 0, \ldots, 4$ in 1 000 simulated samples for $H = 2, \ldots, n/2$; true $K = 2$

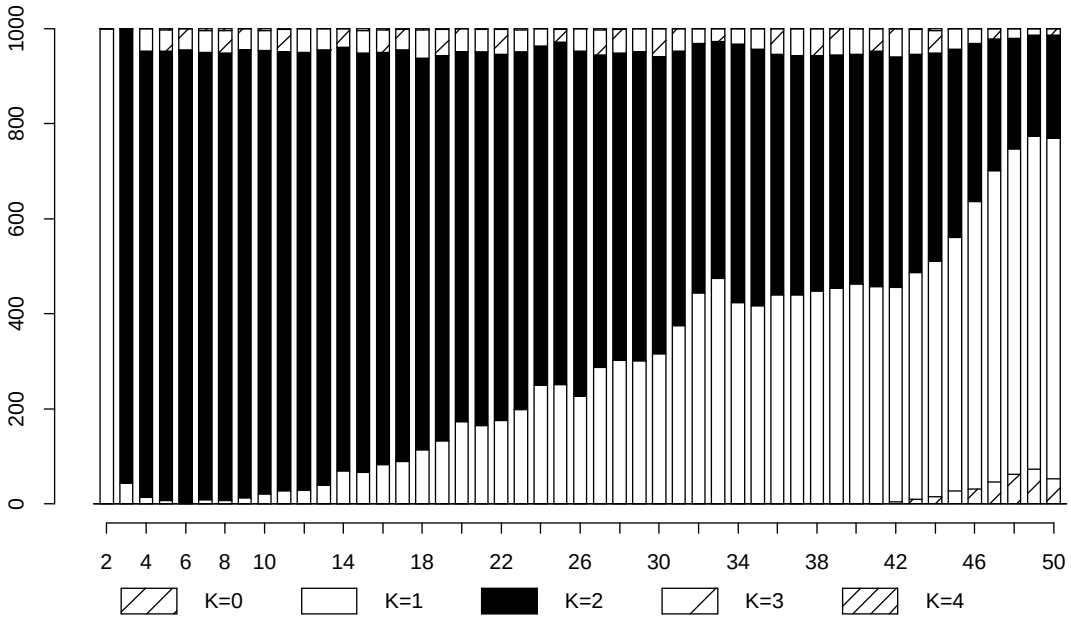


Figure 7: Results for model C1 for $n = 100$: Counts of $\widehat{K} = 0, \ldots, 4$ in 1 000 simulated samples for $H = 2, \ldots, n/2$; true $K = 2$

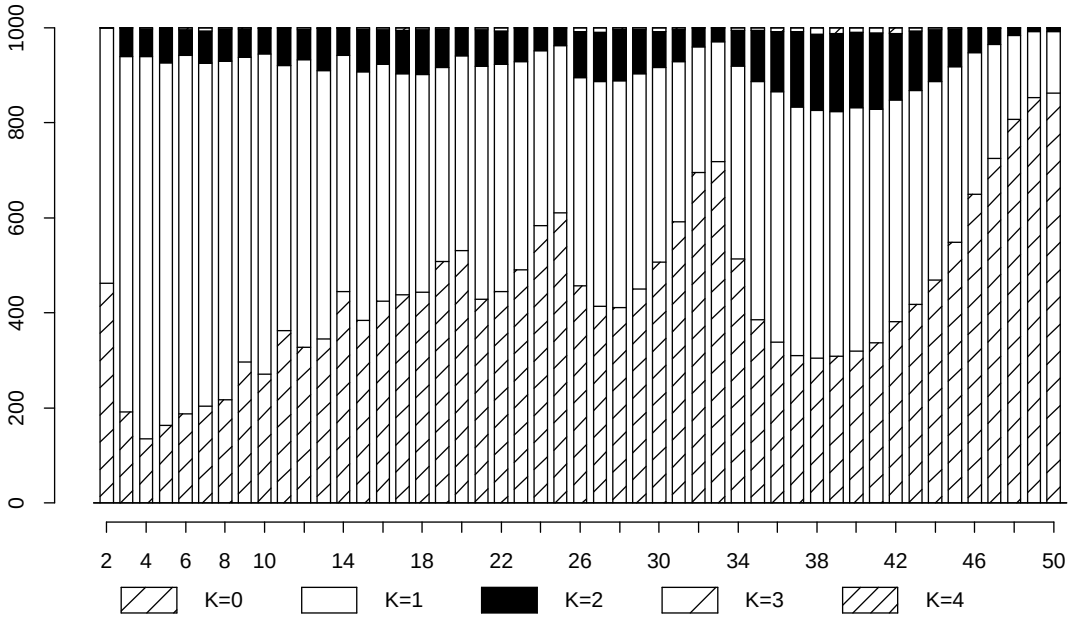
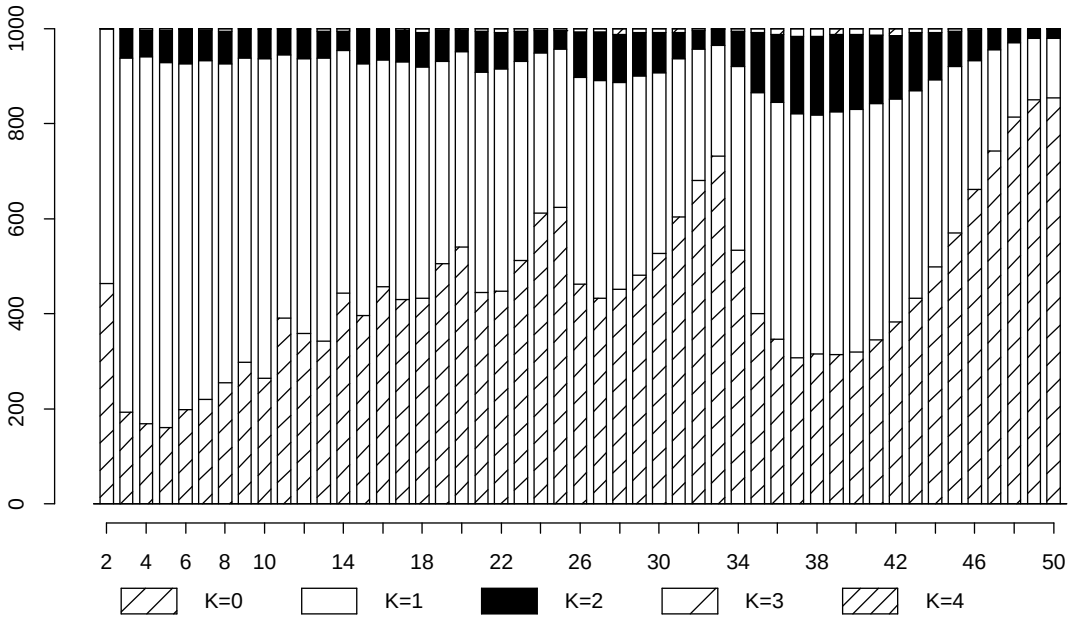


Figure 8: Results for model C2 for $n = 100$: Counts of $\widehat{K} = 0, \ldots, 4$ in 1 000 simulated samples for $H = 2, \ldots, n/2$; true $K = 2$

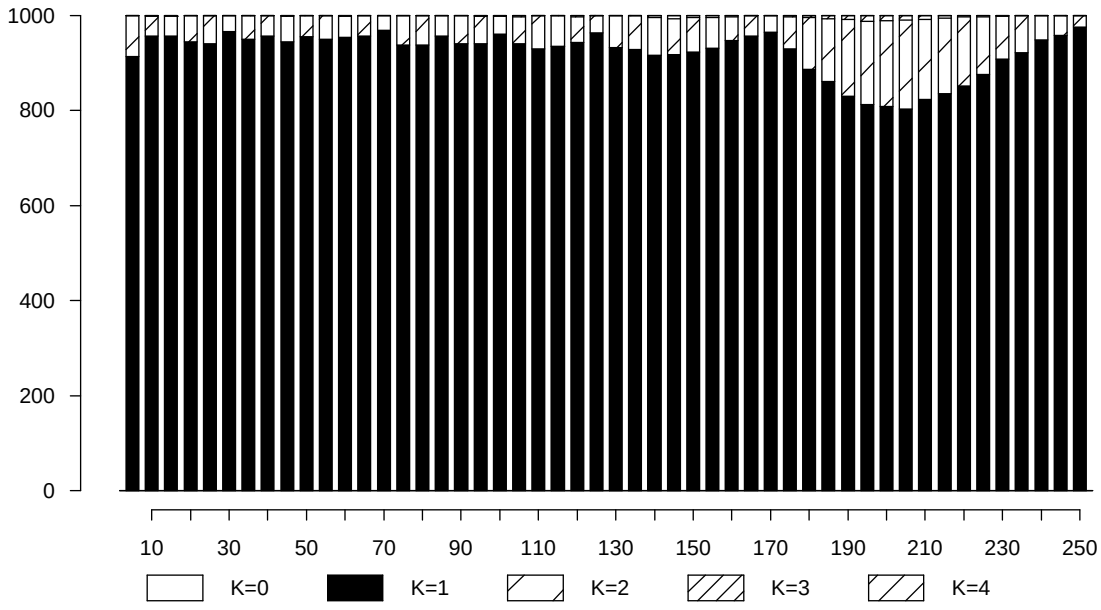


where we get $\widehat{K} = 2$ only for less than 20% of the samples for each choice of H. As can be expected, for the data generated according to models A2-C2, the performance of $\widehat{K}$ becomes slightly worse than for data from the exact models A1-C1, but the results do not differ much. We thus restrict to the exact models in further investigations.

It can be seen immediately from the figures that, even if the estimated dimension reducing directions may not be affected much by the number of slices ([18] Li, 1991; [5] Chen and Li, 1998), the situation is different for the estimated dimension. We see that, even in model A, where the estimator of K behaves best, taking a very large number of slices increases the risk of estimating the wrong dimension. In both models A and B we find the tendency to underestimate K for large values of H. Taking only two slices is not appropriate either, looking at the results of models B and C.

For a sample size which is larger compared to the dimension, the estimation of K generally improves and becomes less dependent on the choice of H. Figures 9-11 give the stacked barplots for models A1-C1 for $n = 500$ and choices of $H = 5(5)250$. We see that for A1 and B1 and large H we still have a greater risk to estimate K wrongly. In contrast to the case of $n = 100$, the procedure now tends to overestimate K for a large number of slices. The same general effect can be seen with C1, where also for large values of H the dimension is estimated to be larger. Contrary to models

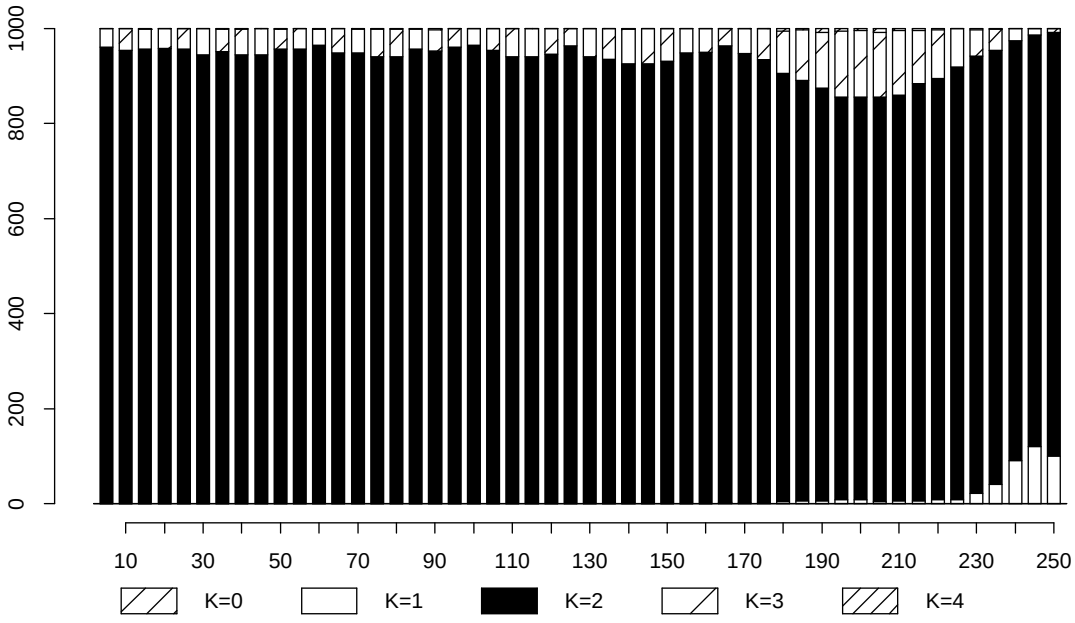
Figure 9: Results of model A1 for $n = 500$: Counts of $\widehat{K} = 0, \ldots, 4$ in 1 000 simulated samples for $H = 5(5)n/2$; true $K = 1$



A1 and B1, in this case the effect seems to result in a better estimation of K. But it is well known that SIR is not able to detect the quadratic part of the structure in model C1, hence the visible improvement here is merely due to the coincidence of a tendency to overestimate K with the non-ability of detecting one essential dimension.

To get further insight into the performance of the tests about K, we look at the distributions of the corresponding test statistics, depending on the choice of the number H of slices. Figures 12 and 13 show how these distributions vary with H for model B1. Recall that the true value is $K = 2$ here, hence when testing for $K = 0$ and for $K = 1$ the distributions of the

Figure 10: Results of model B1 for $n = 500$: Counts of $\widehat{K} = 0, \ldots, 4$ in 1 000 simulated samples for $H = 5(5)n/2$; true $K = 2$



test statistics should lie well separated and to the right of the respective χ^2 distributions. This is indeed the case for the test for $K = 0$ with smaller values of H (see figure 12), but the larger H is chosen, the larger becomes the overlap between the two distributions. The effect is much stronger for the test concerning $K = 1$ (see figure 13). Here, from $H = 30$ on, we see a substantial overlap between the two distributions. The consequence is a clear loss in power of the test. The effects for models A1 and C1 are similar. Figure 14 shows the situation for model C1 when checking $K = 0$.

To illustrate further the dependence of $\widehat{K}$ on the number H of slices, we look at the performance of $\widehat{K}$ for selected data sets from the simulations

Figure 11: Results of model C1 for $n = 500$: Counts of $\widehat{K} = 0, \ldots, 4$ in 1 000 simulated samples for $H = 5(5)n/2$; true $K = 2$

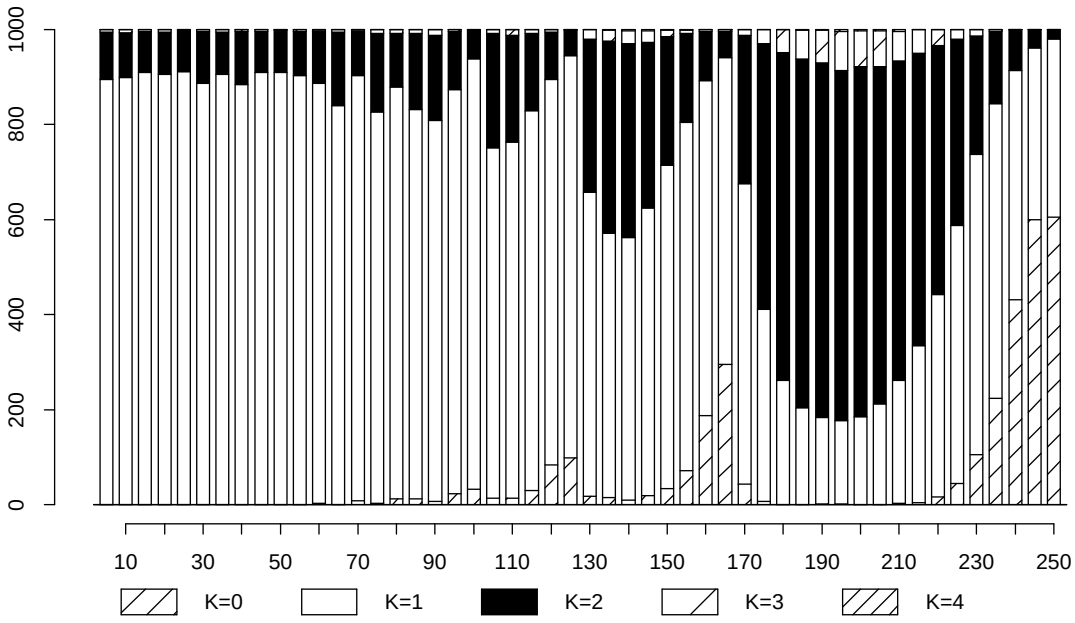


Figure 12: Results of model B1 for $n = 100$, $H = 10(10)50$: Histogram of simulated distribution of t_0 when testing for $K = 0$ compared to densities of χ^2; df = 45, 95, 145, 195, 245, critical values = 61.6562, 118.7516, 174.1010, 228.5799, 282.5115, respectively

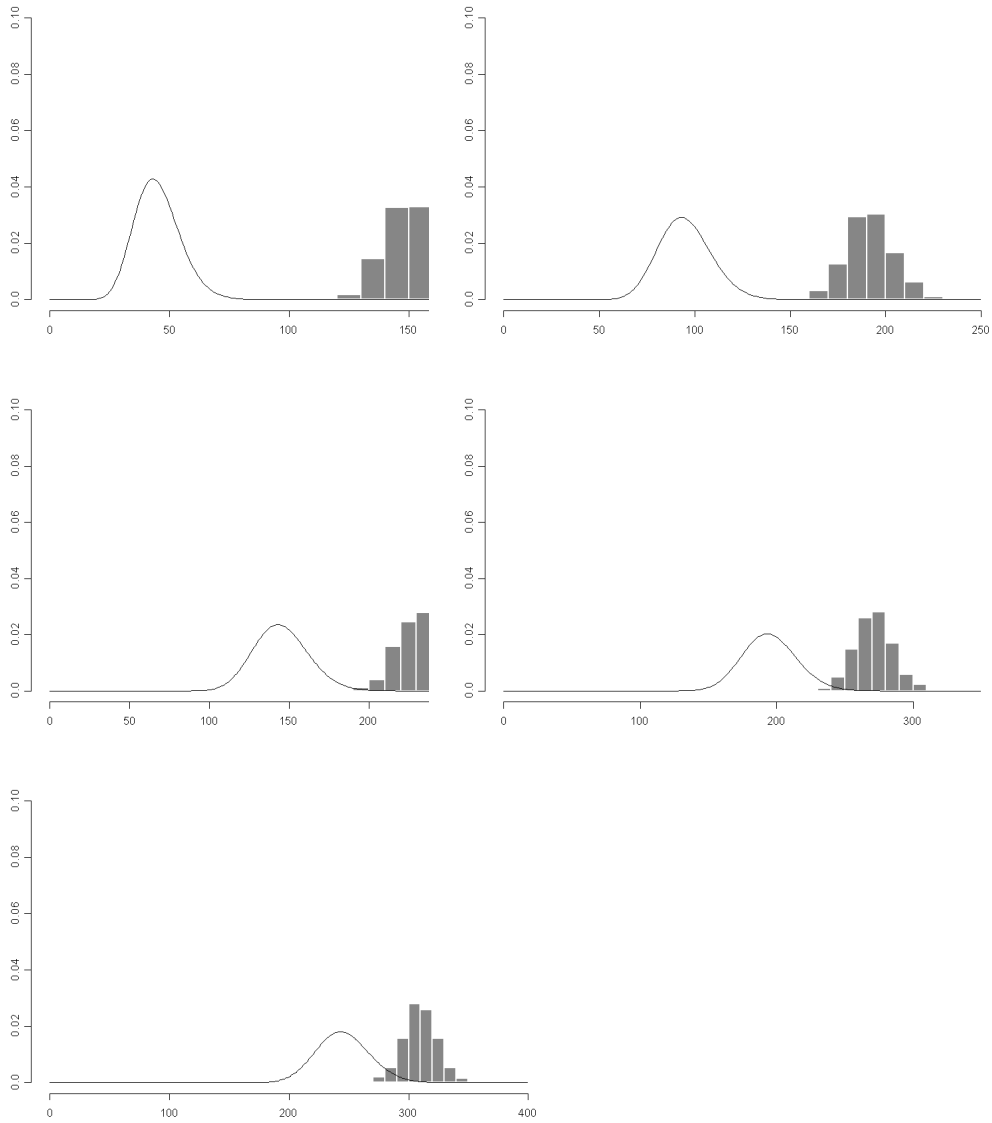


Figure 13: Results of model B1 for $n = 100$, $H = 10(10)50$: Histogram of simulated distribution of t_1 when testing for $K = 1$ compared to densities of χ^2; df = 32, 72, 112, 152, 192, critical values = 46.1943, 92.8083, 137.7015, 181.7702, 225.3288, respectively

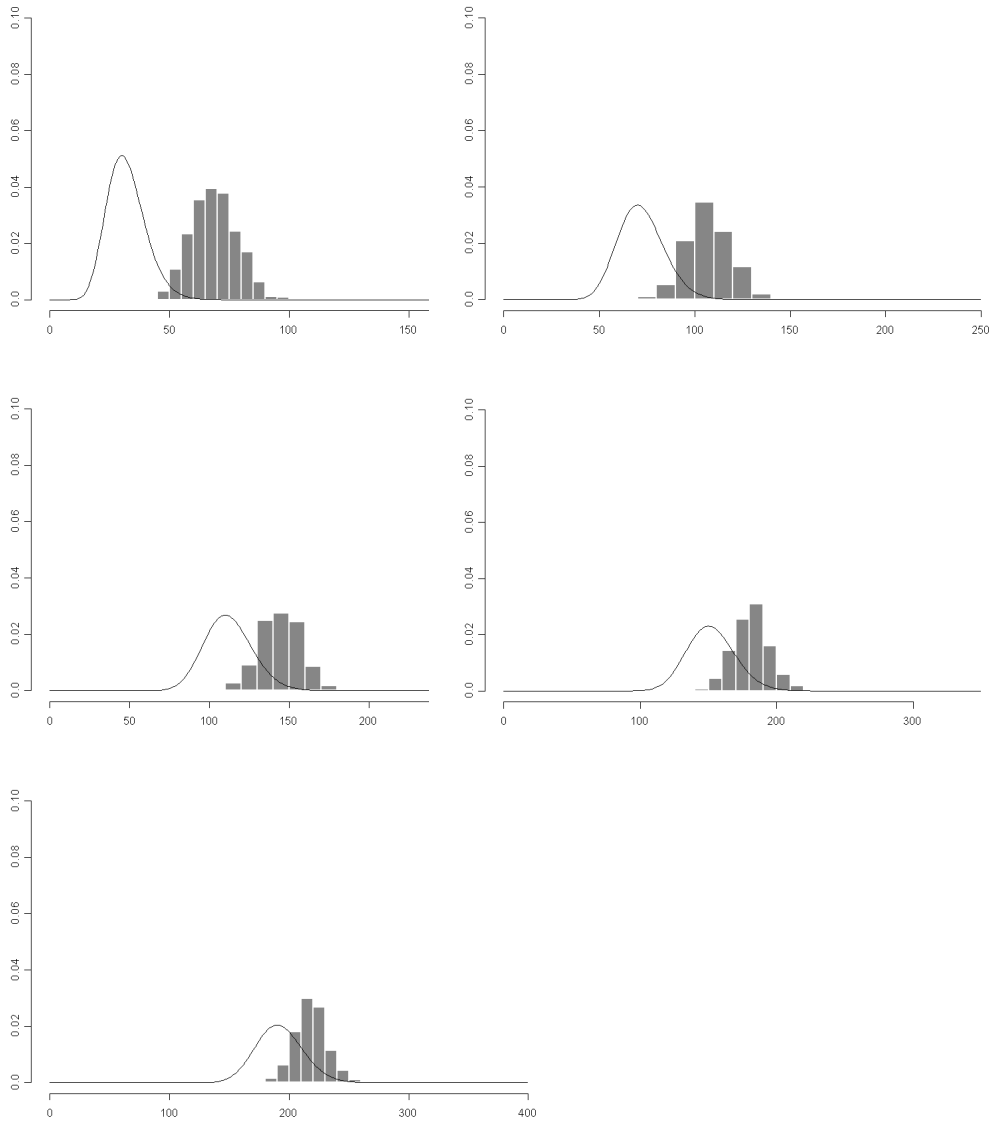


Figure 14: Results of model C1 for $n = 100$, $H = 10(10)50$: Histogram of simulated distribution of t_0 when testing for $K = 0$ compared to densities of χ^2; df = 45, 95, 145, 195, 245, critical values = 61.6562, 118.7516, 174.1010, 228.5799, 282.5115, respectively

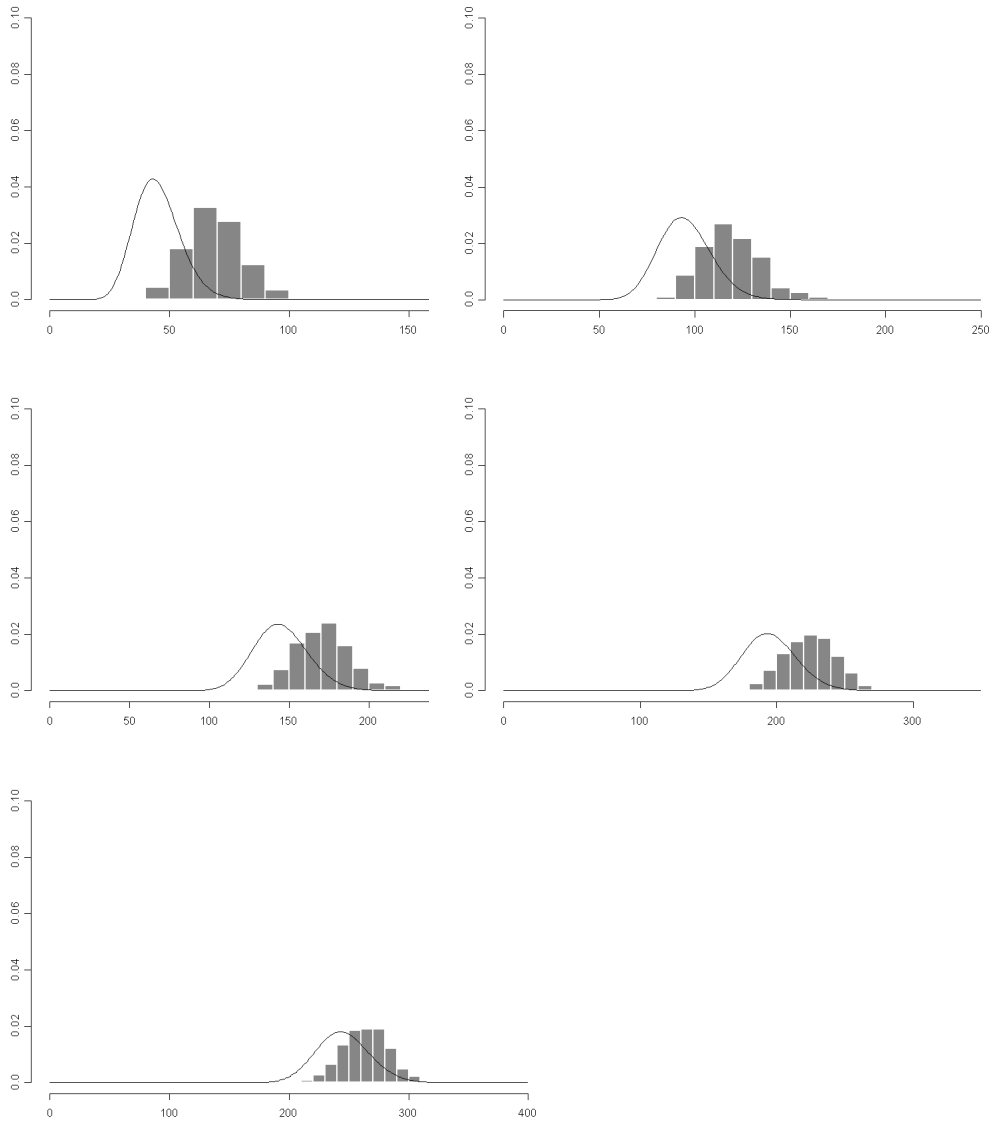
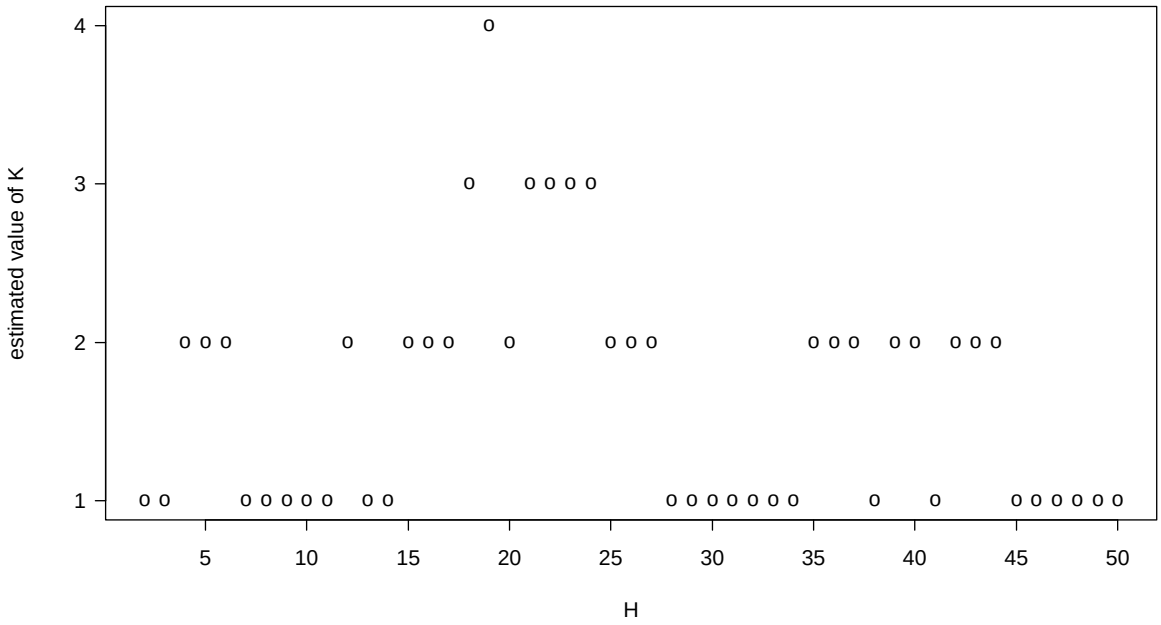


Figure 15: Dataset from model A1, $n = 100$: estimated K depending on the choice of H; true $K = 1$



(figures 15-16). The figures show how $\widehat{K}$ varies for one data set when choosing different values of H. Figure 15 shows an example for a data set generated by model A1. The estimated values $\widehat{K}$ vary between 1 and 4. Figure 16 shows the situation of an 'extreme' data set from model B1 where K is estimated wrongly for more than 40 of the possible 49 choices of H ($n = 100$). Table 2 summarizes the information on this aspect of performance for $n = 100$. For each of the exact models we count in how many of the 1 000 simulated data sets K is estimated wrongly for more than 40, 30, 20, and 10 of the possible choices of H, respectively. We observe again the variability of the results due to the different structures of the models. Apart from this,

Figure 16: Dataset from model B1, $n = 100$: estimated K depending on the choice of H; true $K = 2$

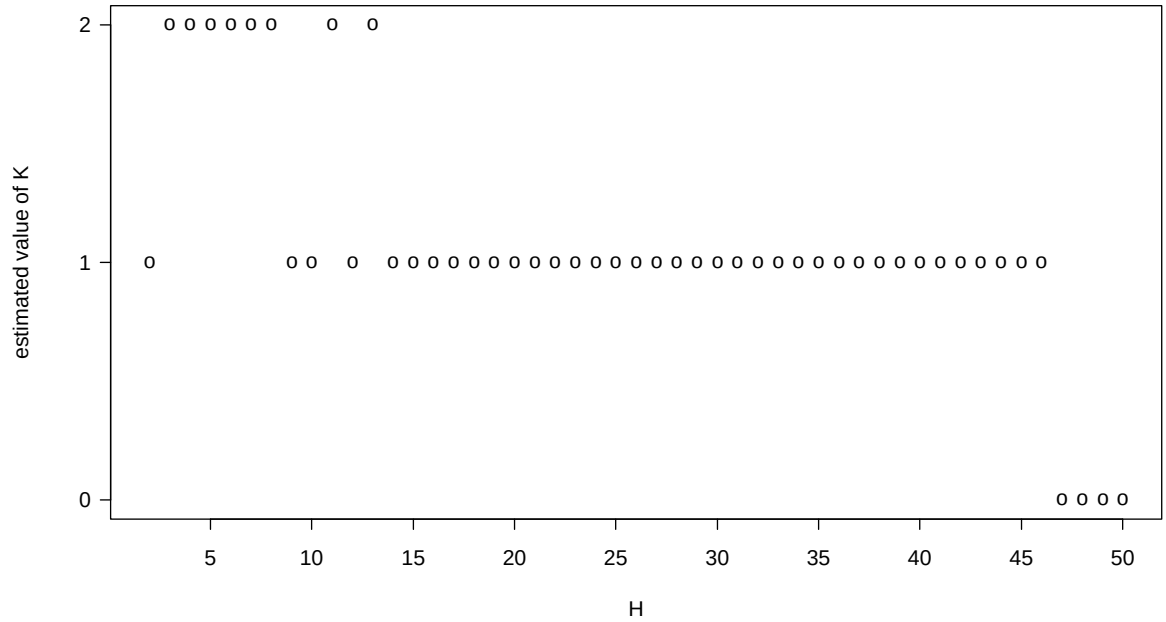


Table 2: Performance of $\widehat{K}$ for $n = 100$ and $H = 2, \ldots, n/2$

	$\widehat{K} \neq K$ for more than ... choices of H			
	40	30	20	10
Model A1 ($K = 1$)	0	7	28	98
Model B1 ($K = 2$)	2	48	221	592
Model C1 ($K = 2$)	835	952	990	1 000

we see that for all models a quite considerable part of the data sets leads to K being estimated wrongly for more than 10 of the possible 49 choices of H. Even in model A1, where $\widehat{K}$ behaves best, this is the case for almost 10% of the data sets.

5 Concluding remarks

In investigations on sliced inverse regression and its various modifications, the choice of the number of slices has not been paid much attention to yet. Our simulation results show that the choice of this 'tuning parameter' may influence the outcome of SIR quite strongly, when considering the estimated dimension of the reduced regressor space. Hence, the choice of H should be carefully done. First impressions from the simulation results lead to the conclusion that the performance of SIR with respect to the estimated dimension is best if H is chosen to be not too large. More precisely, values of $H \approx 0.1n$ seems to be a reasonable choice. Altogether, the development of a possible best choice of H seems worth further investigations.

Acknowledgments

This work was partially supported by the German Science Foundation (Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 475, Reduction of Complexity in Multivariate Data Structures).

References

[1] Becker, C., 2001, Dimension adjustment methods. *Proceedings of the 53rd Session of the Internat. Statist. Institute*. Published on CD-ROM.

[2] Bellman, R.E., 1961, *Adaptive control processes* (Princeton University Press).

[3] Bura, E., and Cook, R.D., 2001a, Extending sliced inverse regression: the weighted chi-squared test. *J. Amer. Statist. Assoc.*, **96**, 996–1003.

[4] Bura, E., and Cook, R.D., 2001b, Estimating the structural dimension of regressions via parametric inverse regression. *J. Royal Statist. Soc. B*, **63**, 393–410.

[5] Chen, C.-H., and Li, K.-C., 1998, Can SIR be as popular as multiple linear regression? *Statist. Sinica*, **8**, 289–316.

[6] Cook, R.D., 1994, On the interpretation of regression plots. *J. Amer. Statist. Assoc.*, **89**, 177–189.

[7] Cook, R.D., 1996, Graphics for regressions with a binary response. *J. Amer. Statist. Assoc.*, **91**, 983–992.

[8] Cook, R.D., 1998, Principal Hessian directions revisited. *J. Amer. Statist. Assoc.*, **93**, 84–100.

[9] Cook, R.D., and Weisberg, S., 1991, Comment on 'Sliced inverse regression for dimension reduction' by K.-C. Li. *J. Amer. Statist. Assoc.*, **86**, 328–332.

[10] Cook, R.D., and Weisberg, S., 1994, *An introduction to regression graphics* (New York: Wiley).

[11] Ferré, L., 1998, Determining the dimension in sliced inverse regression and related methods. *J. Amer. Statist. Assoc.*, **93**, 132–140.

[12] Friedman, J.H., 1994, An overview of predictive learning and function approximation. In: V. Cherkassky, J.H. Friedman, and H. Wechsler (Eds.), *From statistics to neural networks. Theory and pattern recognition applications*, 1–61 (Berlin: Springer).

[13] Gather, U., and Becker, C., 2001, The curse of dimensionality - a challenge for mathematical statistics. *Jahresberichte der deutschen Mathematiker Vereinigung*, **103**, 19–36.

[14] Gather, U., Hilker, T., and Becker, C., 2001, A robustified version of sliced inverse regression. In: L.T. Fernholz, S. Morgenthaler, and W. Stahel (Eds.),

Statistics in genetics and in the environmental sciences, 147–157 (Basel: Birkhäuser).

[15] Gather, U., Hilker, T., and Becker, C., 2002, A note on outlier sensitivity of sliced inverse regression. *Statistics*, **13**, 271-281.

[16] Harrison, D., and Rubinfeld, D.L., 1978, Hedonic housing prices and the demand for clean air. *J. Environmental Economics Management*, **5**, 81–102.

[17] Hsing, T., and Carroll, R.J., 1992, An asymptotic theory for sliced inverse regression. *Ann. Statist.*, **20**, 1040–1061.

[18] Li, K.-C., 1991, Sliced inverse regression for dimension reduction (with discussion). *J. Amer. Statist. Assoc.*, **86**, 316–342.

[19] Li, K.-C., 1992, On principal Hessian directions for data visualization and dimension reduction: another application of Stein's lemma. *J. Amer. Statist. Assoc.*, **87**, 1025–1039.

[20] Saracco, J., 1997, An asymptotic theory for sliced inverse regression. *Comm. Statist. – Theory Methods*, **26**, 2141–2171.

[21] Schott, J.R., 1994, Determining the dimensionality in sliced inverse regression. *J. Amer. Statist. Assoc.*, **89**, 141–148.

[22] Sheather, S.J., and McKean, J.W., 1997, A comparison of procedures based on inverse regression, L_1-*Statistical Procedures and Related Topics*. IMS Lecture Notes, **31**, 271–278.

[23] Velilla, S., 1998, Assessing the number of linear components in a general regression problem. *J. Amer. Statist. Assoc.*, **93**, 1088–1098.

[24] Zhu, L.X., and Fang, K.T., 1996, Asymptotics for kernel estimate of sliced inverse regression. *Ann. Statist.*, **24**, 1053–1068.

[25] Zhu, L.X., and Ng, K.W., 1995, Asymptotics of sliced inverse regression. *Statist. Sinica*, **5**, 727–736.